



Universidad del Azuay
Facultad de Ciencias de la Administración
Escuela de Ingeniería de Sistemas

**“Sistema de Soporte a la toma de Decisiones en el
área de Tráfico Telefónico” de la Empresa
“E.T.A.P.A.”**

**Tesis previa a la obtención del Título
de INGENIERO EN SISTEMAS**

Autoras:

Diana María Astudillo Cobos

Ximena Elizabeth Arguello Cabrera

Director:

Ing. Pablo Pintado Zumba

Codirector:

Ing. Marcos Orellana Cordero

Cuenca – Ecuador

2006

DEDICATORIA

Quiero dedicar esta tesis

A DIOS porque Él ha representado mi fuerza y mi guía. Si Él no lo hubiera permitido, no hubiera podido estudiar ni tener la posibilidad de obtener mi título. Gracias por estar a mi lado siempre y no olvidarte de mí nunca.

A mi familia quienes me han dado todo lo que tengo.

A mi papá por confiar en mí y por darme siempre su apoyo.

A mi mamá por estar al lado mío y el de mis hermanos y por dedicar su vida a educarnos y cuidarnos.

A mis amigos:

A Silvia por ayudarme a seguir con mis estudios y por todas las veces que me escucho cuando más lo necesite.

A Joys, Marco, Juanito, Fer, Cristian, Juan Diego, Juan Pablo, Braulio, Jhofre, Arturo, Cristina, Sonia y José Luís por ser mis amigos y compartir conmigo todos los momentos buenos y malos de la U.

A mi amiga Diana de quien he aprendido muchas cosas sobre todo el amor a DIOS, la responsabilidad y la constancia.

Y a JC, porque me ayudó comprender lo que es el respeto y la sinceridad.

GRACIAS a todos porque después de DIOS y mi familia ustedes fueron y son lo más importante en mi vida.

Ximena.

DEDICATORIA

Dedico esta tesis a Dios en honor al inmenso amor y gratitud que siento, por que Él es el pilar fundamental de mi vida, me ha otorgado unos padres ejemplares y hermanos comprensibles que me han dado toda comodidad, tranquilidad y apoyo incondicional para alcanzar mis metas y poder cumplirlas de la mejor manera. Doy gracias a mis padres porque el resultado de esta tesis representa un esfuerzo mutuo.

Diana.

AGRADECIMIENTO

Queremos agradecer a todas las personas que hicieron posible la realización de esta tesis. A nuestros profesores que nos guiaron durante todos estos años de estudio, especialmente a nuestro Director de tesis Ing. Pablo Pintado y a nuestro Codirector Ing. Marcos Orellana.

Al personal de la empresa E.T.A.P.A, al Ing. Xavier Ordóñez (Director Administrativo), al Ing. Esteban Garzón (Jefe del Dpto. de Desarrollo) y al Ing. Fernando Uyaguari (Líder del proyecto *Data Warehouse*) quienes permitieron e hicieron posible la realización de esta tesis.

Al Ing. Pablo Pérez y al Ing. Carlos Plaza por brindarnos su ayuda, apoyo constante y amistad durante todo este tiempo de trabajo.

A nuestra familia porque sin su apoyo incondicional no hubiéramos podido continuar. Gracias por estar a nuestro lado siempre.

Y Gracias a DIOS porque en los buenos y malos momentos ha estado con nosotros y a representado nuestra fuerza. Sin él nada es posible.

Las Autoras.

ÍNDICE GENERAL

DEDICATORIA	ii
AGRADECIMIENTO	iv
ÍNDICE DE GRÁFICOS	xi
ÍNDICE DE TABLAS	xii
ÍNDICE DE ANEXOS.....	xii
RESUMEN.....	xiii
ABSTRACT	xiii
INTRODUCCIÓN	1
CAPÍTULO 1	
MODELOS DE PROCESOS.....	3
1.1. Introducción	3
1.2. Definición <i>UML</i>	3
1.2.1. Objetivos de <i>UML</i>	4
1.2.2. Modelos o diagramas	4
1.2.2.1. Tipos de diagramas	4
1.2.2.1.1. Modelo estático (estructural)	4
1.2.2.1.2. Modelo dinámico (comportamiento)	5
1.3. Proceso de desarrollo de <i>software</i>	6
1.4. Proceso unificado racional (<i>RUP</i>).....	6
1.4.1. Características <i>RUP</i>	7
1.4.1.1. El proceso unificado está dirigido por casos de uso	7
1.4.1.2. El proceso unificado está centrado en la arquitectura.....	8
1.4.1.3. El proceso unificado es iterativo e incremental	8
1.5. Modelo <i>MSF</i>	10
1.5.1. Principios fundamentales de <i>MSF</i>	11
1.5.1.1. El trabajo hacia una visión compartida	11
1.5.1.2. Mantenerse ágil (Esperar cambios).....	11
1.5.1.3. Enfocarse a entregar valor al negocio	12
1.5.1.4. Promover una comunicación abierta.....	12
1.5.2. Conceptos claves para el modelo proceso <i>MSF</i>	12
1.5.2.1. Clientes	12
1.5.2.2. Inversionistas	13
1.5.2.3. ¿Qué es una solución?.....	13
1.5.2.3.1. Elementos de una solución	13
1.5.2.4. Puntos de referencias	14
1.5.2.5. Alcance	15
1.5.2.6. Administrar cambios	16
1.5.2.6.1. El triángulo del cambio.....	16
1.5.2.6.2. La matriz de cambio del proyecto	17
1.5.3. Características del modelo de proceso	18
1.5.3.1. Características de un modelo basado en hitos.....	18
1.5.3.2. Características de un modelo iterativo	19
1.5.3.2.1. Liberación de Versiones	19
1.5.3.2.2. Crear documentos vivos	19
1.5.3.2.3. Definición de puntos de referencia (línea base), bloqueo.....	20
1.5.3.2.4. Construcciones diarias	20
1.5.3.2.5. Administración de configuración	20

1.5.3.2.6. Guía para liberaciones en versión.....	20
1.5.3.3. Modelo integrado de desarrollo e implementación.....	21
1.5.4. Notas para usar el modelo de proceso integrado.....	21
1.5.5. Fases del modelo de proceso e hitos.....	22
1.5.5.1. Fase de visionamiento.....	22
1.5.5.1.1. Entregables.....	23
1.5.5.1.2. Hitos secundarios sugeridos.....	23
1.5.5.2. Fase de planeación.....	24
1.5.5.2.1. Planes del proyecto aprobados.....	25
1.5.5.2.2. Entregables.....	26
1.5.5.3. Fase de desarrollo.....	26
1.5.5.3.1. Hito del alcance completo.....	26
1.5.5.3.2. Entregables.....	27
1.5.5.4. Fase de estabilización.....	27
1.5.5.4.1. Hito de liberación disponible.....	27
1.5.5.4.2. Entregables.....	28
1.5.5.5. Fase de implementación.....	28
1.5.5.5.1. Implementación del hito completo.....	28
1.5.5.5.2. Entregables.....	29
1.6 Conclusión.....	29

CAPÍTULO 2

TECNOLOGÍA DE INFORMACIÓN.....	30
2.1. Introducción.....	30
2.2. Inteligencia de negocios (<i>BI</i>).....	30
2.3. Funcionalidad de los sistemas analíticos.....	31
2.4. Niveles de decisiones de negocio.....	32
2.5. Problemas que resuelve la inteligencia de negocios.....	32
2.6. Componentes <i>BI</i>	33
2.7. Ventajas de una solución <i>BI</i>	33
2.8. Sistemas de información.....	34
2.8.1. Sistema de información ejecutiva (<i>EIS</i>).....	35
2.8.1.1. Lo que debe contener un <i>EIS</i>	36
2.8.1.1.1. Interfaz gráfica fácil de usar y ver.....	36
2.8.1.1.2. Alarmas o semáforos.....	37
2.8.1.1.3. Integración de información proveniente de los cubos.....	37
2.8.2. Sistema de soporte de decisiones (<i>DSS</i>).....	37
2.8.2.1. Definición.....	38
2.8.2.2. Características.....	38
2.8.2.3. Ciclo de vida.....	39
2.9. <i>Management Relation Customer CRM</i>	40
2.9.1. Definición.....	41
2.10. Conclusión.....	41

CAPÍTULO 3

<i>DATA WAREHOUSE</i>	43
3.1. Introducción.....	43
3.2. Definición.....	43
3.3. Características.....	43
3.3.1. Orientado a tema.....	44
3.3.2. No volátil.....	45
3.3.3. Integrada.....	46
3.3.4. Tiempo variante.....	47
3.4. Estructura del <i>Data Warehouse</i>	47

3.4.1. Detalle de datos actuales	48
3.4.2. Detalle de datos antiguos.....	48
3.4.3. Datos ligeramente resumidos.	48
3.4.4. Datos completamente resumidos.....	48
3.4.5. Metadata	49
3.5. Elementos constituyentes de una arquitectura <i>Data Warehouse</i>	49
3.5.1. Identificación de fuentes	49
3.5.2. Nivel de organización de datos	50
3.5.3. Nivel de acceso a la información	50
3.6. Operaciones dentro del ambiente <i>Data Warehouse</i>	51
3.6.1. <i>ETL</i> (Extracción, Transformación y Carga)	51
3.6.1.1. Proceso <i>ETL</i>	52
3.6.1.2. Operaciones diarias.....	52
3.6.1.3. Evolución del <i>Data Warehouse</i>	52
3.6.1.4. Extracción en <i>Data Warehouses</i>	52
3.6.1.4.1. Consideraciones para la extracción	53
3.6.1.5. Transporte de datos	53
3.6.1.6. Transformación y carga de los datos.....	54
3.6.2. Elaborar la Metadata	55
3.6.3. Dar acceso de usuario final	55
3.7. <i>Data Marts</i>	56
3.7.1. Distribución de la <i>Data</i> del <i>Data Warehouse</i> a los <i>Data Marts</i>	58
3.8. Conclusión.....	58
CAPÍTULO 4	
ANÁLISIS DIMENSIONAL.....	60
4.1. Introducción	60
4.2. Análisis Dimensional / Multidimensional	61
4.3. Cubo multidimensional	62
4.4. Modelos de datos dimensionales	63
4.4.1. Esquema Estrella	64
4.4.1.1. Ventajas	64
4.4.1.2. Consideraciones	65
4.4.1.3. Dimensiones.....	65
4.4.1.3.1. Estructura de una dimensión.....	66
4.4.1.3.2. Dimensiones privadas y compartidas	67
4.4.1.3.3. Jerarquías de dimensión.....	67
4.4.1.3.4. Niveles y miembros	68
4.4.1.4. Tablas de hecho.....	69
4.4.1.4.1. Presentación de medidas.....	70
4.4.1.4.2. Agregaciones	70
4.4.1.4.3. Mejorar los tiempos de respuesta (diseño físico/lógico)	71
4.4.2. Esquema copo de nieve	72
4.5. Análisis para la extracción	73
4.6. Análisis para la integración	75
4.6.1. Formato de integración.....	75
4.6.2. Semántica de integración	77
4.7. Conclusión.....	77
CAPÍTULO 5	
TRILOGÍA - MODELO CONCEPTUAL/ LÓGICO /FÍSICO	79
5.1. Introducción	79
5.2. Modelo Conceptual, Lógico y Físico	80
5.3. Modelo Conceptual	80

5.3.1. Definición.....	84
5.3.2. Requerimientos.....	84
5.3.2.1. Tratamiento de las circunstancias y la retrospección.....	86
5.3.3. <i>Dot Modeling</i>	87
5.3.3.1. Componentes de un Modelo Punto de Comportamiento	90
5.3.3.2. Modelo punto y el Modelo Conceptual General.....	92
5.3.4. Talleres de trabajo del modelado punto	93
5.3.4.1. Metodología del Modelado Punto.....	93
5.3.4.1.1. Taller de trabajo de información estratégica	94
5.3.4.1.1.1. Proceso del taller de trabajo de información estratégica	95
5.3.4.1.1.1.1. Primer paso: Introducción al Taller	95
5.3.4.1.1.1.2. Segundo Paso: Primera sesión	97
5.3.4.1.1.2. Taller de trabajo cubierto	104
5.3.4.1.2. Taller de trabajo del Componente de Análisis.....	104
5.3.4.1.2.1. El proceso del taller de trabajo para el componente de Análisis .	105
5.3.4.1.2.2. Taller de trabajo cubierto	110
5.4. Modelo lógico	110
5.4.1. Introducción al modelo lógico.....	110
5.4.2. El uso de los atributos de existencia.....	111
5.4.3. El uso de la dimensión tiempo	112
5.4.4. Esquema lógico	114
5.4.5. Consideraciones de ejecución	114
5.4.6. Seleccionar una solución	115
5.5. La implementación física	116
5.5.1. Modelo de desarrollo <i>Data Warehouse</i>	116
5.5.1.1. La capa de validación, integración y mapeo (<i>VIM</i>)	119
5.5.1.1.1. Validación de datos	119
5.5.1.1.1.1. Pérdida de datos	119
5.5.1.1.1.2. Data errónea	122
5.5.1.1.1.3. Data fuera de tiempo	122
5.5.1.1.1.4. Data inconsistente	122
5.5.1.1.2. Integración.....	123
5.5.1.1.3. Mapeo	123
5.6. Conclusión.....	123

CAPÍTULO 6

PROCESAMIENTO ANALÍTICO EN LÍNEA	125
6.1. Introducción	125
6.2. Bases de datos <i>OLAP</i>	125
6.3. Tipos de almacenamiento de datos.....	126
6.3.1. <i>MOLAP</i>	126
6.3.1.1. Arquitectura <i>MOLAP</i>	127
6.3.1.2. Ventajas	127
6.3.1.3. Desventajas	127
6.3.2. <i>ROLAP</i>	128
6.3.2.1. Arquitectura	128
6.3.2.2. Ventajas	129
6.3.2.3. Desventajas	129
6.3.3. <i>HOLAP</i>	129
6.4. Seleccionar el mejor modo de almacenamiento	130
6.4.1. Tipo de almacenamiento	130
6.4.2. Tiempos de proceso para cada tipo de almacenamiento	130
6.4.3. Requerimientos de espacio en disco para cada tipo de almacenamiento	131
6.4.4. Requerimientos de disco para cubos <i>MOLAP</i> comparado con el Esquema estrella	132

6.5. Operaciones <i>OLAP</i>	133
6.6. Conclusión.....	135

CAPÍTULO 7

ENTREGABLES <i>MSF</i>	136
7.1. Fase I: Documento visión/alcance.....	136
7.1.1. Reuniones de levantamiento (visión global fase I).....	136
7.1.2. Introducción	137
7.1.3. Oportunidad del negocio.	138
7.1.3.1. Descripción de la oportunidad	138
7.1.4. Planteamiento del problema	139
7.1.5. Visión del proyecto	139
7.1.6. Concepto de la solución.	139
7.1.7. Metas, objetivos, asunciones, y limitaciones	140
7.1.7.1. Metas, objetivos, asunciones.....	140
7.1.7.2. Limitaciones.....	141
7.1.8. Análisis de uso	141
7.1.9. Perfiles de usuario	142
7.1.10. Requerimientos.....	142
7.1.10.1. Requerimientos de negocio.....	142
7.1.11. Consideraciones importantes.....	144
7.1.12. Procesos funcionales: análisis tráfico telefónico.....	145
7.1.13. Procesos de conmutación	145
7.1.14. Variables necesarias	148
7.1.14.1. Telefonía	148
7.1.14.2. Conmutación.....	148
7.1.15. Reportes.....	149
7.1.15.1. Telefonía:	149
7.1.15.2. Conmutación.....	149
7.1.16. Alcance del proyecto.....	150
7.1.17. Lista de funciones / características	150
7.1.18. Criterio de aceptación.....	151
7.1.19. Criterio operacional.....	151
7.1.20. Matriz de riesgos del proyecto	151
7.2. Fase II: Especificación y diseño.....	153
7.2.1. Diseño de la base dimensional (Modelo Lógico, Físico)	153
7.2.1.1. Introducción	153
7.2.1.2. Diseño lógico	154
7.2.1.2.1. Tráfico telefónico	154
7.2.1.2.1.1. Tráfico entrante	155
7.2.1.2.1.2. Tráfico saliente.....	156
7.2.1.2.1.2.1. Tráfico saliente con operadora (destinos)	156
7.2.1.2.1.2.2. Tráfico local	157
7.2.1.2.1.2.3. Tráfico local detallado (abonados).....	158
7.2.1.2.2. Control de calidad de circulación del tráfico	160
7.2.1.2.2.1. Auxiliares (Clase A).....	160
7.2.1.2.2.2. Tráfico erlangs (Clase C)	161
7.2.1.2.2.3. Haces (Clase FB).....	162
7.2.1.2.2.4. URA (Clase R)	163
7.2.1.2.2.5. Tipos de llamadas (Clase T).....	163
7.2.1.3. Estructura de la base multidimensional	165
7.2.1.3.1. Tráfico telefónico	166
7.2.1.3.1.1. Cubo físico	166
7.2.1.3.1.2. Cubos virtuales.....	166
7.2.1.3.1.3. Dimensiones.....	166

7.2.1.3.1.4. Medidas	167
7.2.1.3.2. Control de calidad circulación tráfico (conmutación)	168
7.2.1.3.2.1. Cubos.....	168
7.2.1.3.2.2. Cubos virtuales.....	168
7.2.1.3.2.3. Dimensiones	169
7.2.1.3.2.4. Medidas	170
7.2.2. Documento de identificación de fuentes	171
7.2.2.1. Introducción	171
7.2.2.2. Identificación de fuentes de la información seleccionada.....	171
7.2.2.2.1. Tráfico telefónico	171
7.2.2.2.2. Control de calidad circulación conmutación	173
7.2.2.3. Consideraciones importantes:	174
7.2.2.4. Puntos a Tomar en Cuenta:	175
7.3. Fase III Desarrollo.....	176
7.3.1. Documento técnico.....	176
7.3.1.1. Proceso general tráfico telefónico.....	176
7.3.1.1.1. Extracción y transformación.....	177
7.3.1.1.1.1. Extracción y transformación tráfico telefónico	178
7.3.1.1.1.1.1. Extracción diccionarios.....	178
7.3.1.1.1.1.2. Extracción tráfico entrante	179
7.3.1.1.1.1.3. Transformación tráfico entrante.....	179
7.3.1.1.1.1.4. Extracción tráfico saliente.....	181
7.3.1.1.1.1.5. Transformación tráfico saliente	182
7.3.1.1.1.2. Extracción conmutación	185
7.3.1.1.1.2.1. Diccionarios control de calidad de circulación tráfico telefónico (conmutación)	185
7.3.1.1.1.2.2. Extracción y transformación auxiliares.....	186
7.3.1.1.1.2.3. Extracción y transformación tráfico erlangs	186
7.3.1.1.1.2.4. Extracción y transformación tráfico haces.....	187
7.3.1.1.1.2.5. Extracción y transformación URA.....	188
7.3.1.1.1.2.6. Extracción y transformación tipos de llamadas	188
7.3.1.1.2. Población de datos	193
7.3.1.1.2.1. Población hechos tráfico telefónico	193
7.3.1.1.2.2. Población hechos conmutación	194
7.3.1.1.2.3. Población dimensiones tráfico telefónico.....	194
7.3.1.1.2.4. Población dimensiones conmutación	195
7.3.1.1.3. Control de integridad	196
7.3.1.1.3.1. Integración conmutación	197
7.3.1.1.3.2. Integración tráfico telefónico	197
7.3.1.1.4. Secuencia de ejecución	197
7.3.1.1.4.1. Proceso de conmutación.....	198
7.3.1.1.4.2. Proceso tráfico telefónico.....	201
7.3.1.1.5. Cubos	203
7.3.1.1.5.1. Dimensiones	203
7.3.1.1.5.1.1. Jerarquías asociadas a una dimensión.....	203
7.3.1.1.5.2. Medidas	204
7.3.1.1.5.3. Modelo de cubos	204
7.3.1.1.5.4. Diseño de almacenamiento de cubos.....	204
7.3.1.1.5.5. Generación de cubos y dimensiones	204
7.3.1.2. Proceso para administrar la base de datos.....	205
7.3.1.2.1. Proceso para crear respaldo y truncar log.....	205
7.3.1.2.2. Restaurar base de datos.....	206
7.3.1.3. Visualización.....	206
CONCLUSIONES	207

RECOMENDACIONES	209
GLOSARIO	246
BIBLIOGRAFÍA	254

ÍNDICE DE GRÁFICOS

Gráfico 1.1. Ciclo de vida del Proceso Unificado.....	9
Gráfico 1.2. Modelo <i>MSF</i>	11
Gráfico 1.3. Elementos de una solución.....	14
Gráfico 1.4. Triángulo de cambio	16
Gráfico 1.5. Matriz de cambios.....	17
Gráfico 1.6. Liberación de versiones	19
Gráfico 2.1. Esquema Conceptual Inteligencia de Negocios.....	34
Gráfico 2.2. Esquema Conceptual Sistema de Información Ejecutivo	37
Gráfico 2.3. Esquema Conceptual Sistema de Soporte de Decisiones	40
Gráfico 3.1. Características de Data Warehouse Orientado al Tema.....	44
Gráfico 3.2. Actualización de datos (Característica no volátil)	45
Gráfico 3.3. Integración de datos	46
Gráfico 3.4. Tiempo Variante	47
Gráfico 3.5. Niveles de esquematización que podrían encontrarse en un <i>DWH</i>	47
Gráfico 3.6. Elementos de la Arquitectura <i>Data Warehouse</i>	50
Gráfico 3.7. Esquema Conceptual <i>Data Warehouse</i>	56
Gráfico 3.8. Distribución de los Datos a los <i>Data Mart</i>	58
Gráfico 4.1. Cubo tridimensional.....	62
Gráfico 4.2. Modelo Dimensional “Consumo de Tráfico Saliente “	63
Gráfico 4.3. Esquema Estrella.....	64
Gráfico 4.4. Relación entre los niveles y miembros en una dimensión	67
Gráfico 4.5. Esquema Copo de Nieve.....	73
Gráfico 5.1. Modelo general para un cliente con circunstancias cambiantes	82
Gráfico 5.2. Ejemplo de un reporte de dos dimensiones.....	88
Gráfico 5.3. Reporte tráfico saliente con tres dimensiones.....	89
Gráfico 5.4. Modelo Multidimensional Punto Sencillo	89
Gráfico 5.5. Modelo Punto para el Área de Tráfico Telefónico (E.T.A.P.A).....	93
Gráfico 5.6 Hoja de trabajo de hecho	107
Gráfico 5.7 Ejemplo de jerarquía	107
Gráfico 6.1. <i>MOLAP</i>	128
Gráfico 6.2. <i>ROLAP</i>	129
Gráfico 6.3. Tiempo de procesamiento para cubos.....	131
Gráfico 6.4. Espacio en Disco requerido para cubos	131
Gráfico 6.5. Consulta <i>OLAP</i>	133
Gráfico 7.1 Proceso funcional Tráfico Telefónico.....	145
Gráfico 7.2 Proceso de Conmutación	147
Gráfico 7.3. Esquema Estrella Tráfico Entrante	155
Gráfico 7.4. Esquema Estrella Tráfico Saliente	157
Gráfico 7.5. Esquema Estrella Tráfico local	158
Gráfico 7.6. Esquema Estrella Tráfico local detallado	158
Gráfico 7.7. Esquema Estrella Tráfico Telefónico.....	159

Gráfico 7.8. Esquema Estrella Auxiliares (Conmutación).....	160
Gráfico 7.9. Esquema Estrella Tráfico Erlangs (Conmutación)	161
Gráfico 7.10. Esquema Estrella Haces (Conmutación).....	162
Gráfico 7.11. Esquema Estrella Unidad de Conexión de Abonado (Conmutación) 163	
Gráfico 7.12. Esquema Estrella Tipos de llamadas (Conmutación)	164
Gráfico 7.13. Esquema Estrella Control de Calidad Circulación Tráfico (Conmutación)	165

ÍNDICE DE TABLAS

Tabla 1.1. Explicación de Solución y Producto	13
Tabla 1.2. Rol del Equipo en la Fase de Visionamiento	23
Tabla 1.3. Rol del Equipo en la Fase de Planeación	26
Tabla 4.1. Relaciones entre los niveles y miembros de la dimensión regular	68
Tabla 5.1. Comparación de almacenamiento de espacio	132

ÍNDICE DE ANEXOS

ANEXO I METADATA	210
1.1. Tráfico Telefónico.....	210
1.1.1 Tráfico Entrante.....	210
1.1.2. Tráfico Saliente	215
1.1.2.1. Tráfico local	218
1.1.2.2. Tráfico local detallado	219
1.2. Control de conmutación	220
1.2.1. Tráfico Auxiliares	220
1.2.2. Tráfico Erlangs.....	222
1.2.3. Tráfico Haces	223
1.2.4. Tráfico URA.....	225
1.2.5. Tipos de Llamadas en la central	227
ANEXO II DICCIONARIO DE DATOS	230
2.1. Descripción de las Bases de datos utilizadas.....	230
2.1.1. Base de datos TANQUE	231
2.1.1.1. Tablas para Tráfico Telefónico	231
2.1.1.2. Tablas para el control de calidad Conmutación	236
2.1.2. Base de datos DWH	245

RESUMEN

Este documento es una recopilación de manuales, libros y documentos electrónicos que presentan ideas fundamentales para desarrollar un *DWH* como parte de una solución a problemas empresariales.

Los niveles de granularidad al que se desea llegar para analizar el comportamiento de una empresa se basan en objetivos que permiten cumplir su visión y misión. Los requerimientos de la gente de negocios llevarán al desarrollo de una base de datos única, consistente e integrada (*Data Warehouse*).

Se ha utilizado la metodología *MSF* para controlar y conocer las tareas para que conviertan requerimientos en *software*, así como herramientas para el análisis y la población de datos.

ABSTRACT

This document is a summary of manuals, books and electronic documents that present fundamental ideas to develop a *DWH* like a part of a solution for management problems.

The granularity levels at which you want to reach for analyzing the behaviour of an enterprise are based on objectives that allow to reach their vision and mission. Business People's requirements will drive to the development of an unique, consistent and integrated database (*Data Warehouse*).

The *MSF* methodology has been used to control and to know the tasks to transform requirements into software, as well as tools for the analysis and the load of data.

INTRODUCCIÓN

Las empresas o gente de negocios basan sus decisiones en el análisis de la información, por ello es importante que dicha información sea real, esté consolidada y se presente a tiempo. La falta de integridad de la información y diferentes estándares lleva muchas veces a contar con información errónea, lo que influye en las decisiones que se toman a diario.

Para que una empresa pueda lograr sus objetivos y sea más competitiva tiene que buscar maneras para conseguir información confiable, uniendo a personas que conozcan el negocio y la tecnología, para obtener el mejor provecho. Gracias a esta necesidad ha evolucionado la tecnología de información, enfocada específicamente en la inteligencia de negocios, para la cual existen herramientas que permiten generar conocimiento a partir de datos de diversas fuentes tomando sólo las necesarias para construir indicadores que nos muestre el comportamiento de la empresa y además la facilidad de analizar la información a través de los diferentes niveles de detalle.

En vista de la importancia que tienen estos factores dentro de una empresa, se considera necesaria la implementación de un sistema que solvete estos problemas, y que a menudo se dan en los sistemas transaccionales en línea (*OLTP*).

Se hace necesaria la construcción de una base de datos consolidada, integrada e histórica, que nos permita obtener información válida y a tiempo, para la toma de decisiones, lo que conlleva al conocimiento de: requerimientos de información (Modelo Conceptual), indicadores, herramientas a utilizarse, bases de datos de las cuales se extraerá la información (Fuentes de Extracción), disponibilidad de recursos tanto humanos como técnicos dentro de la empresa, establecer prioridades en cuanto al almacenamiento en disco y ejecución de consultas.

Habiendo conocido cada uno de estos elementos podremos establecer la metodología y herramientas para el análisis, extracción, transporte y carga de datos al *Data Warehouse*, así como la forma en la que se presentará la información al usuario final.

Actualmente en el mercado se puede encontrar gran cantidad de herramientas para dichas tareas. Para elegir la correcta, quien se encargue del desarrollo e implementación del sistema, deberá basarse en las necesidades de la empresa y los recursos que esta posee.

CAPÍTULO 1

MODELOS DE PROCESOS

1.1. Introducción

Este capítulo está orientado a la obtención de un modelo de proceso de desarrollo de *software* que se adapte a las necesidades del proyecto, es decir que nos ayude a determinar las actividades necesarias para transformar los requisitos del usuario en un sistema *software*, por lo cual es necesario plantear los conceptos básicos de: *Microsoft Solution Framework (MSF)* y el Proceso Unificado de *software (RUP)* para lo cual se ha utilizado el libro “El proceso unificado de desarrollo de *software*” de los autores Jacobson Ivar, Booch Grady, Rumbaugh James y la información proporcionada por *Microsoft* desde su sitio *web* www.microsoft.com.

Los modelos de procesos sirven como una guía para planificar las actividades. Pueden basarse en el lenguaje estándar *UML*, que es independiente de cualquier lenguaje de programación y metodología que se use para el desarrollo del *software*.

1.2. Definición *UML*

Es un lenguaje estándar de la industria que sirve para visualizar, especificar, construir y documentar los artefactos (entregables) que se generan en el proceso de desarrollo de un sistema *software* orientados a objetos y/o basados en componentes, así como para el modelado de negocios y otros sistemas no *software*. Representa una colección de las mejores prácticas de ingeniería que han mostrado éxito en el modelado de sistemas grandes y complejos.

UML usa mayormente notaciones gráficas para expresar el diseño de proyectos de *software*, este ayuda a los equipos del proyecto a comunicarse. *UML* no es una metodología o proceso, ni tampoco un lenguaje de programación

1.2.1. Objetivos de *UML*

Los principales objetivos en el diseño de *UML* son:

- Dar a los usuarios un lenguaje de modelado visual expresivo, listo para usar, para que puedan desarrollar e intercambiar modelos significativos.
- Proveer mecanismos extensivos y especializados para ampliar los conceptos centrales.
- Ser independiente de los lenguajes de programación particulares y de los procesos de desarrollo.
- Proveer una base formal para el entendimiento del lenguaje modelado.
- Soportar los conceptos de desarrollo de alto nivel tales como colaboraciones, marco de trabajo, patrones y componentes.

1.2.2. Modelos o diagramas

Es una simplificación de la realidad, proporciona “los planos” de un sistema y puede ser detallado, en función de los elementos que sean relevantes en cada momento, es decir sirve para capturar “lo esencial”.

1.2.2.1. Tipos de diagramas

El sistema puede describirse desde los distintos puntos de vista:

1.2.2.1.1. Modelo estático (estructural)

Un **diagrama de clases** presenta un conjunto de clases, interfaces, colaboraciones y las relaciones entre ellas.

Un **diagrama de objetos** representa un conjunto de objetos y sus relaciones. Se utilizan para describir estructuras de datos, instantáneas de los elementos encontrados en los diagramas de clases. Cubre los mismos aspectos que los diagramas de clases pero desde una perspectiva de casos reales o prototípicos.

Un **diagrama de componentes** muestra un conjunto de componentes y sus relaciones, se utiliza para describir la vista de implementación estática de un sistema, se relaciona con los diagramas de clases, ya que un componente normalmente se corresponde con una o más clases, interfaces o colaboraciones.

Un **diagrama de despliegue** muestra un conjunto de nodos y sus relaciones. Los diagramas de despliegue se relacionan con los diagramas de componentes, ya que un nodo normalmente incluye uno o más componentes.

1.2.2.1.2. Modelo dinámico (comportamiento)

Los **diagramas de casos de uso** organizan los comportamientos del sistema es decir modela la funcionalidad del sistema de acuerdo a la percepción de los usuarios externos (actores). Tiene como propósito enumerar actores mostrando qué actores participan y casos de uso.

Los **Diagramas de interacción** son un nombre colectivo que se da a los diagramas de secuencia y de colaboración. Ambos diagramas son isomorfos, es decir, se puede convertir de uno a otro sin pérdida de información.

- Un diagrama de secuencia es un diagrama de interacción que resalta la ordenación temporal de los mensajes, presenta un conjunto de objetos y los mensajes enviados y recibidos por ellos. Los objetos suelen ser instancias con nombre o anónimas de clases, pero también pueden representar instancias de otros elementos, tales como colaboraciones, componentes y nodos.
- Un diagrama de colaboración es un diagrama de interacción que resalta la organización estructural de los objetos que envían y reciben mensajes. Un diagrama de colaboración muestra un conjunto de objetos, enlaces entre estos objetos y mensajes enviados y recibidos por ellos.

Un **diagrama de estado** representa una máquina de estados, constituida por estados, transiciones, eventos y actividades. Son especialmente importantes para modelar el comportamiento de una interfaz, una clase o una colaboración. Los diagramas de estados resaltan el comportamiento dirigido por eventos de un objeto, lo que es especialmente útil al modelar sistemas reactivos.

Un diagrama de actividades es un tipo especial de diagrama de estados que muestra el flujo secuencial de actividades en un sistema. Son especialmente importantes para modelar la función del sistema, así como para resaltar el flujo de control entre objetos.

1.3. Proceso de desarrollo de *software*

Un proceso define quién está haciendo qué, cuando, y cómo alcanzar un determinado objetivo (construir un producto *software* o mejorar uno existente), proporcionando normas para el desarrollo eficiente. Este debe servir de guía para todos los participantes (clientes, usuarios, desarrolladores y directores ejecutivos).

Un proceso de desarrollo de *software* debe ser capaz de evolucionar durante muchos años, a las realidades que permitan las tecnologías (lenguaje de programación, sistemas operativos, estructuras de red, etc.), herramientas, personas y patrones de organización, estos deben estar equilibrados.

1.4. Proceso unificado racional (*RUP*)

Debido a los avances de la tecnología y las facilidades de obtener información los usuarios exigen un *software* mucho más complejo y sofisticado que se ajuste a sus necesidades, sin embargo conseguirlo es difícil, mientras no cambiemos los métodos para desarrollar el *software*.

La dificultad se encuentra en los desarrolladores al momento de coordinar el trabajo de un gran proyecto. Se necesitará de un método común, un proceso que:

- Proporcione una guía para ordenar las actividades de un equipo.
- Dirija las tareas de cada desarrollador por separado y del equipo como un todo.
- Especificar los entregables que deben desarrollarse.
- Obtener criterios para el control, métricas de productos y actividades del proyecto.

El proceso unificado es un conjunto de actividades necesarias para transformar los requisitos en un *software*, además debe servir como un marco de trabajo genérico especializado para cualquier sistema de *software*, áreas de aplicación, niveles de aptitud y tamaño.

El proceso unificado utiliza el Lenguaje Unificado de Modelado (*UML*) como parte esencial, pero lo que lo hace único es que está dirigido por casos de uso, centrado en la arquitectura, iterativo e incremental.

1.4.1. Características RUP

Los tres factores claves del proceso Unificado son:

1. Centrado en la arquitectura ya que proporciona la estructura sobre la cual guiar las iteraciones.
2. Dirigidos por caso de uso, ya que definen los objetivos y dirige el trabajo de cada iteración.
3. La iteración controla el riesgo de no sacar el producto al mercado y además reconoce lo que se ignora a menudo: que no se puede obtener todos los requerimientos al inicio.

La eliminación de uno de ellos reducirá drásticamente el valor del Proceso Unificado.

1.4.1.1. El proceso unificado está dirigido por casos de uso

Un sistema de *software* está siempre orientado a las necesidades de usuarios, que no sólo son personas, sino otros sistemas, los cuales van a interactuar con el *software* en desarrollo.

Para poder obtener la funcionalidad del sistema se debe contestar a la pregunta ¿Qué debe hacer el sistema para cada usuario?.

1.4.1.2. El proceso unificado está centrado en la arquitectura

La arquitectura *software* nace de las necesidades de la empresa como las perciben los usuarios y los inversionistas y se reflejan en los casos de uso, además está influenciada en la plataforma en la que tiene que funcionar el *software* (*hardware*), los bloques reutilizables que se dispone, etc. La arquitectura es una vista del diseño completo, esto ayuda a centrarse en los objetivos como pueden ser la comprensibilidad, capacidad de adaptación al cambio y reutilización.

La arquitectura (forma) y los casos de uso (función) están íntimamente relacionados porque la arquitectura debe permitir el desarrollo de los casos de usos y estos deben encajar en la arquitectura, por lo cual, ambos deben evolucionar en paralelo.

1.4.1.3. El proceso unificado es iterativo e incremental

Es necesario dividir en varias partes un proyecto de *software*, a las cuales se las conoce como iteración, que es el resultado de un incremento. Las iteraciones hacen referencia a pasos en el flujo de trabajo, y los incrementos al crecimiento del producto. Para obtener una efectividad, las iteraciones deben estar controladas.

La iteración tiene dos factores el primero trata de los casos de uso el cual amplía el producto desarrollado, en segundo lugar, la iteración trata de los riesgos más importantes.

Las iteraciones sucesivas se construyen sobre los artefactos de desarrollo tal como quedan al final de la última iteración.

En cada iteración se identifican y especifican los casos de uso relevantes, se crea un diseño utilizando la arquitectura seleccionada como guía, se implementa el diseño mediante componentes y se verifican que estos satisfagan los casos de uso. Si una iteración cumple con los objetivos, el desarrollo continúa con la siguiente iteración, caso contrario se debe revisar y probar con un nuevo enfoque.

Beneficios de un proceso iterativo controlado:

- Reduce el coste, por ejemplo si se tiene que repetir la iteración se pierde el esfuerzo empleado en este, y no el valor del producto entero.
- Reduce el riesgo de no entregar a tiempo el proyecto.
- Acelera el ritmo del esfuerzo de desarrollo debido a que trabajan de manera más eficiente.
- Permite la facilidad de cambios en vista. Los requisitos no siempre se conoce con exactitud al principio del proyecto.

1.4.2. Vida del proceso unificado

El Proceso Unificado se repite a lo largo de una serie de ciclos que constituyen la vida de un sistema, cada ciclo constituye una versión del producto para los clientes.

Cada ciclo consta de cuatro fases: inicio, elaboración, construcción y transición.

Cada fase se subdivide en iteraciones.

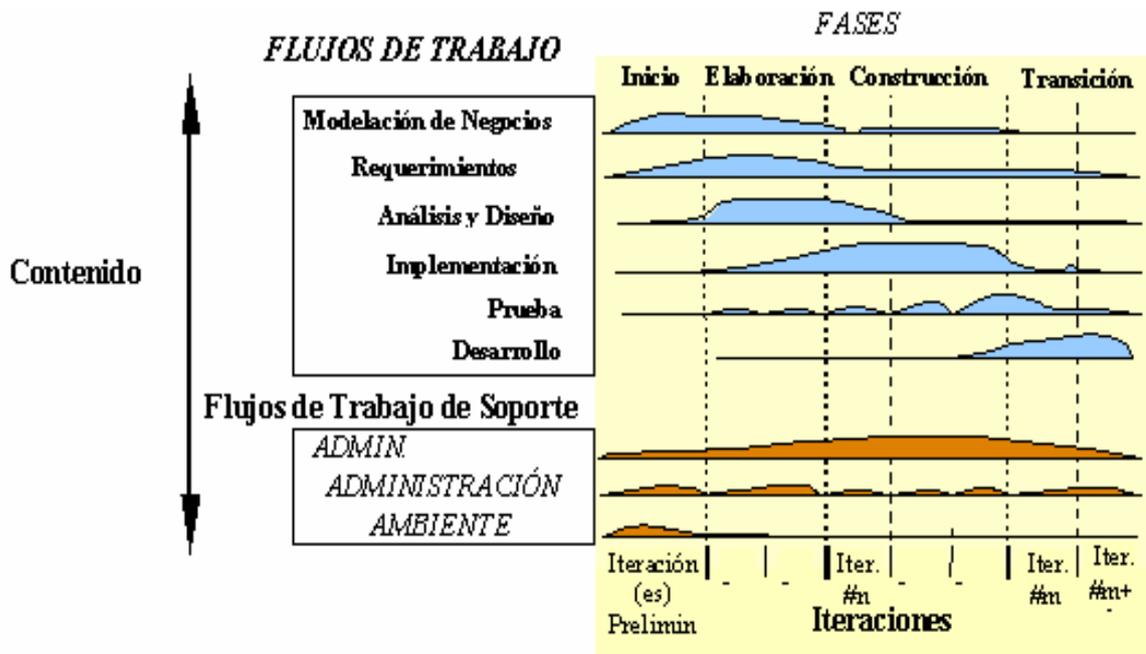


Gráfico 1.1. Ciclo de vida del Proceso Unificado¹
Jacobson Ivar, "Proceso Unificado de Desarrollo de Software"

1.5. Modelo *MSF*

Microsoft Solution Framework (MSF) es una serie flexible e interrelacionada de conceptos, modelos y prácticas de uso que controlan la planificación, el desarrollo y la gestión de proyectos tecnológicos. *MSF* se centra en los modelos de proceso y de equipo dejando en un segundo plano las elecciones tecnológicas. Originalmente creado en 1994 para resolver los problemas a los que se enfrentaban las empresas al realizar sus proyectos, se ha convertido en un modelo práctico que facilita el éxito de los proyectos tecnológicos.

El proceso *MSF* mostrado en el gráfico 1.2, combina los mejores principios de los modelos espiral y cascada.

Modelo en Cascada: usa hitos como puntos de transición y evolución, además cada conjunto de tareas debe ser completado antes de que la siguiente fase pueda empezar. El modelo cascada funciona mejor en proyectos donde es factible delinear claramente un conjunto fijo de requerimientos cambiantes del proyecto desde un inicio. Los puntos fijos de transición entre fases facilitan el seguimiento del cronograma y la asignación de responsabilidades y obligaciones.

Modelo Espiral: este modelo se enfoca en la continua necesidad de refinar los requerimientos y presupuesto para el proyecto. Es muy eficaz cuando es usado para el desarrollo rápido de aplicaciones en un proyecto muy pequeño.

Este modelo estimula una gran sinergia entre el equipo del desarrollo y el cliente porque el cliente provee una retroalimentación y aprobación para todas las etapas del proyecto. Sin embargo, debido a que el modelo no incorpora puntos de control claros, el proceso de desarrollo puede llegar a ser caótico. Por estas razones el modelo *MSF* se deriva de los beneficios de predicción de la planeación basada en hitos del modelo en cascada, así como de los beneficios de retroalimentación y creatividad del modelo espiral.

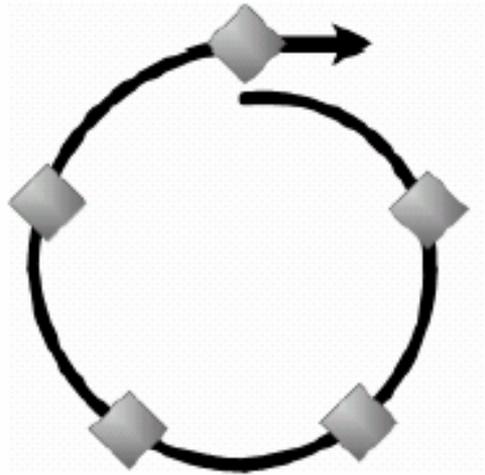


Gráfico 1.2. Modelo MSF ²
Microsoft MSF WhitePaper

1.5.1. Principios fundamentales de MSF

El modelo de proceso está directamente asociado con los siguientes cuatro principios:

1.5.1.1. El trabajo hacia una visión compartida

Lo fundamental para el éxito de cualquier actividad en conjunto es que los miembros del equipo y el cliente tengan una visión compartida, esto es, un claro entendimiento de cuales son las metas y objetivos para la solución. Los miembros del equipo y clientes traen consigo presunciones de lo que la actividad va a hacer para la organización.

Una visión compartida causa que esas presunciones estén claras y asegura que todos los participantes estén trabajando para lograr la misma meta. Clarificar y lograr un compromiso de una visión compartida es tan importante que el modelo proceso *MSF* designa una fase y un hito principal para este propósito.

1.5.1.2. Mantenerse ágil (Esperar cambios)

Las disciplinas tradicionales para la administración de un proyecto asumen que los requerimientos pueden estar claramente articulados desde el inicio y que no cambiarán significativamente durante el ciclo de vida del proyecto. *MSF* hace la suposición fundamental que el cambio continuo debería ser esperado y manejado.

1.5.1.3. Enfocarse a entregar valor al negocio

Tanto las soluciones de éxito así como los objetivos en las organizaciones deben satisfacer algunas necesidades básicas y entregar valor al cliente. Una solución no provee un valor hasta que ésta es desarrollada en la producción. Por esta razón, el ciclo de vida del modelo de proceso *MSF* incluye ambos: el desarrollo e implementación de una solución.

1.5.1.4. Promover una comunicación abierta

El modelo de proceso *MSF* es un modelo abierto y honesto para comunicaciones entre el equipo y los principales inversionistas. Un flujo libre de información no sólo reduce las oportunidades de malos entendidos y esfuerzo gastado; sino que también asegura que todos los miembros del equipo puedan contribuir a la reducción de incertidumbres que rodean el proyecto. Por estas razones, el modelo de proceso *MSF* provee puntos de revisión. Los entregables documentados conservan el progreso del proyecto visible y bien comunicado dentro del equipo, inversionistas y el cliente.

1.5.2. Conceptos claves para el modelo proceso *MSF*

Para entender el modelo de proceso, es importante entender la forma como *MSF* ha definido los siguientes conceptos y términos:

1.5.2.1. Clientes

En las soluciones de negocio, el cliente es la persona u organización que delega el proyecto, provee fondos, y quien espera lograr el valor de negocios de la solución. Los usuarios son las personas que interactúan con la solución en su trabajo. Por ejemplo: los usuarios son los empleados, mientras que el cliente es un miembro de la administración encargado de establecer el nuevo sistema.

El involucramiento del cliente en proyectos *IT* es esencial para el éxito. Se debe permitir a éste poder formar y modificar los requerimientos y establecer puntos de control para revisar el progreso. Estas actividades requieren tiempo y compromiso.

1.5.2.2. Inversionistas

Los inversionistas son particulares que tienen un interés en involucrarse en el resultado del proyecto. Sus metas no siempre son idénticas a aquellas del cliente. Las responsabilidades del rol de administración de un producto incluyen la identificación de los inversionistas clave del proyecto, tomando en cuenta sus necesidades y administrando las relaciones entre inversionistas.

1.5.2.3. ¿Qué es una solución?

En el uso diario, una solución es simplemente una estrategia o método para resolver un problema. En *MSF* tiene un significado muy específico. Es la entrega coordinada de elementos requeridos (tales como tecnologías, documentación, entrenamiento y soporte) para responder exitosamente a un problema único de negocios del cliente. A pesar de que *MSF* es usado para desarrollar productos comerciales para un mercado masivo se enfoca principalmente en la entrega de soluciones entalladas para un cliente específico. Una solución puede incluir uno o más productos de *software*, pero la diferencia entre productos y soluciones puede ser conservada de forma clara. Las diferencias son resumidas en la siguiente tabla:

Productos	Solución <i>MSF</i>
Diseñados para las necesidades de un mercado masivo	Diseñado entallado para encajar en necesidades individuales del cliente
Entregado como productos aplicados o “bits” (para descarga, CD-ROM, etc.)	Entregado como un proyecto.

Tabla 1.1. Explicación de Solución y Producto.³
Microsoft MSF WhitePaper

1.5.2.3.1. Elementos de una solución

Los proyectos pueden variar en complejidad y en esfuerzo necesarios para su desarrollo. Algunos elementos mostrados en el gráfico 1.3. pueden no ser necesarios en una implementación relativamente simple. Sin embargo, proyectos más complejos, es probable que requieran todos los elementos.

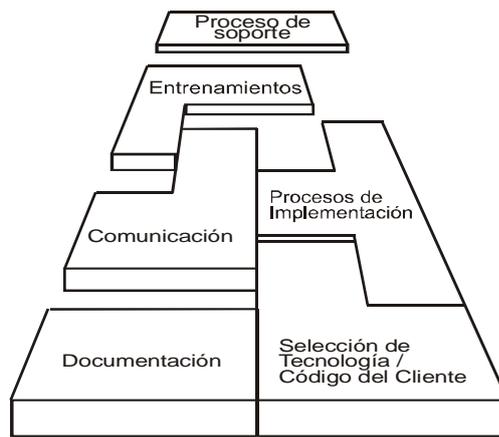


Gráfico 1.3. Elementos de una solución ⁴
Microsoft MSF WhitePaper

- **Tecnologías seleccionadas / código del cliente:** pueden ser nuevas, actualizadas, modificadas, o incluir componentes añadidos. Las tecnologías pueden incluir *hardware*, *software*, periféricos, o componentes de red. El código del cliente es código desarrollado para un proyecto específico.
- **Entrenamiento:** para quienes estarán usando o dando soporte a la solución que se implementa.
- **Documentación:** se refiere a toda la información requerida para instalar, mantener, soportar y usar la solución.
- **Procesos de soporte:** incluyen los procedimientos necesarios para realizar respaldos, restauraciones, recuperación de desastres, resolución de problemas y funciones de ayuda en el escritorio.
- **Comunicaciones externas:** involucran mantener a los inversionistas externos informados del progreso de la implementación y las formas en las cuales la solución los afectará.
- **Procesos de implementación:** incluyen procesos de instalación ó desinstalación para la implementación de *hardware* y *software*, herramientas de implementación automatizadas y procedimientos para *rollback* emergentes.

1.5.2.4. Puntos de referencias

Un punto de referencia es una medida o un estado conocido por el cual algo es medido. Algunos ejemplos de entregables son: código fuente, configuraciones de

servidor, cronogramas, especificaciones, manuales de usuario y presupuestos. Sin puntos de referencia es imposible administrar la variación.

1.5.2.5. Alcance

El alcance es la suma de entregables y servicios a ser provistos en el proyecto. El alcance define lo que se debe hacer para mantener e integrar la visión compartida, asignada contra la realidad y refleja lo que el cliente considera esencial para el éxito de la liberación. Como una parte de la definición de alcance, la funcionalidad menos urgente es transferida a proyectos futuros. Los beneficios de definir el alcance son:

- Dividir una visión a largo plazo en partes realizables.
- Definir las características que estarán en cada liberación.
- Permitir la flexibilidad para los cambios.
- Proveer una línea de fondo para los cambios.
- Enfocar al equipo a identificar que trabajo debe realizarse.
- Facilita dividir tareas grandes y confusas, en tareas pequeñas y entendibles.
- Identificar el trabajo específico del proyecto que no está claramente asociado con alguna característica específica, tal como preparar reportes de estado.
- Facilita subdividir el trabajo entre compañeros en el equipo.
- Aclara esas partes de la solución de las que el equipo es responsable, así como las partes de las cuales no lo es.
- Asegura que todas las partes de la solución tienen un responsable para construirla o mantenerla. Para grandes soluciones especialmente, las características son parte de la solución, pero no parte de los entregables.

Las características del alcance de una solución deben ser definidas y administradas así como el alcance del trabajo y servicios a ser provistos por el equipo del proyecto. El término “alcance” tiene dos aspectos: el alcance de la solución y el alcance del proyecto. Existe una correlación entre estos dos, pero no son lo mismo. Entender esta distinción ayuda al equipo a administrar el cronograma y el costo de sus proyectos. El alcance de la solución describe las características de la solución y los entregables.

Una característica es un aspecto notable de una aplicación o parte de *hardware*. Por ejemplo, la habilidad de revisar antes de imprimir es una característica de una aplicación *Word*, el manual de usuario, archivos de ayuda en línea, guías de operaciones y entrenamiento también son características de toda la solución.

1.5.2.6. Administrar cambios

La administración del alcance es crítica para el éxito del proyecto. Algunos proyectos *IT* fracasan porque son completados tarde o están sobre el presupuesto debido a un alcance de administración pobre. La administración incluye aclarar el alcance temprano y un buen rastreo del proyecto y control de cambio. Hacer efectivos los cambios es clave para el éxito, debido a la inherente incertidumbre y riesgo involucrados con los proyectos *IT*.

1.5.2.6.1. El triángulo del cambio

Existe una relación bien conocida entre las variables del proyecto: Recursos (gente y dinero), Cronograma (tiempo), y Características (alcance). Estas variables existen en un triángulo de relación como se muestra en el gráfico 1.4.

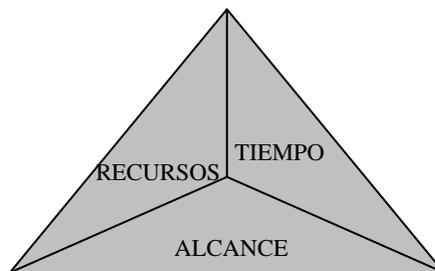


Gráfico 1.4. Triángulo de cambio⁵
Microsoft MSF WhitePaper

Después de haber establecido el triángulo, cualquier cambio de uno de sus lados requiere una corrección de uno o de los otros dos lados para mantener el balance del proyecto.

Éste incluye potencialmente, el mismo lado en el cual el cambio ocurrió primero. La clave para implementar una solución que coincida con las necesidades del cliente cuando ellos la necesitan es encontrar el correcto balance entre recursos, fechas de

implementación y características. Los clientes son a veces renuentes a recortar las características favoritas. El triángulo de cambio ayuda a explicar las restricciones y presenta opciones de cambio. Las características tienen un nivel fijo de calidad que se presume es no negociable. Puede ver la calidad como una cuarta dimensión la cual transformaría el triángulo en un tetraedro (o una pirámide de tres lados). Aunque bajar la barra de calidad resulta en una reducción simultánea de recursos, cronogramas, y características, es obviamente una receta para el fracaso.

1.5.2.6.2. La matriz de cambio del proyecto

Otra poderosa herramienta para administrar los cambios es la matriz de cambios del proyecto, mostrada en el gráfico 1.5. Éste es un acuerdo entre el equipo y el cliente, hecho de forma temprana en el proyecto, considerando las prioridades por defecto al tomar decisiones de cambio. Si es necesario puede haber excepciones para las prioridades por defecto. Pero el principal beneficio de establecer prioridades por defecto es ayudar a hacer cambios menos contenciosos.

	FIJO	SELECCIONADO	AJUSTABLE
RECURSOS	✓		
CRONOGRAMA		✓	
ALCANCE			✓

Gráfico 1.5. Matriz de cambios⁶
Microsoft MSF WhitePaper

El gráfico 1.5. muestra la típica matriz de cambios usada por los equipos de producto *Microsoft*. Esta matriz ayuda a identificar las restricciones del proyecto que son esencialmente no variables (representada por la columna Fijo), las restricciones que son prioridades deseables (representadas por la columna Seleccionado), y las restricciones que pueden ser ajustadas (representada por la columna Ajustable) para adaptarse a aquellas restricciones que son Fijas y Seleccionadas. El equipo y el cliente deben revisar todas las restricciones del proyecto cuidadosamente y ser preparadas para hacer elecciones difíciles.

Para entender como la matriz de intercambio trabaja, las variables de recursos, cronograma y características pueden ser insertados en los espacios de la siguiente oración:

Dado lo _____ establecidos, seleccionaremos un _____ y ajustaremos _____ si es necesario.

Algunas posibles sentencias lógicas:

- Dados los Recursos establecidos, seleccionaremos un cronograma y ajustaremos el conjunto de características de ser necesario.
- Dado un conjunto de características establecidas, seleccionaremos un nivel de recursos y ajustaremos al cronograma como sea necesario.
- Dado un cronograma fijo, seleccionaremos un nivel de recursos y lo ajustaremos a un conjunto de características como sea necesario.

1.5.3. Características del modelo de proceso

Las tres distintas características del proceso *MSF* son:

- Un modelo basado en hitos.
- Un modelo iterativo.
- Un modelo integrado para construir e implementar soluciones.

1.5.3.1. Características de un modelo basado en hitos

Los hitos son usados para planear y monitorear el progreso del proyecto. *MSF* distingue entre dos tipos de hitos: hitos principales y secundarios.

Los hitos principales sirven para la transición de una fase a otra y para la transición de responsabilidad a través de roles. Son lo suficientemente genéricos para cualquier tipo de proyecto *IT*. El logro del éxito de un hito principal representa el acuerdo del equipo y el cliente sobre el alcance, para continuar con el proyecto.

Los hitos secundarios sirven como indicadores de progreso temprano y segmentan los esfuerzos de trabajos grandes en piezas trabajables. Varían dependiendo del tipo de proyecto. *MSF* sugiere un conjunto de hitos secundarios, pero los equipos pueden definirlos.

1.5.3.2. Características de un modelo iterativo

La práctica de un desarrollo iterativo es un tema recurrente en *MSF*. Los entregables que se desarrollan son: código, documentos, diseños, planes, etc.

1.5.3.2.1. Liberación de Versiones

MSF recomienda que las soluciones sean desarrolladas construyendo, probando e implementando la funcionalidad central. Luego los conjuntos de características son agregados. Esto es conocido como una estrategia de liberación de versión. Algunos proyectos pequeños pueden sólo necesitar una versión. El gráfico 1.6 muestra como la funcionalidad se desarrolla sobre algunas versiones.

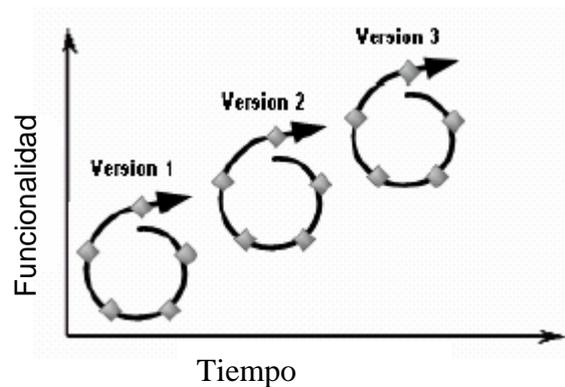


Gráfico 1.6. Liberación de versiones⁷
Microsoft MSF WhitePaper

1.5.3.2.2. Crear documentos vivos

Para evitar el espiral fuera de control, el desarrollo iterativo requiere documentación que se adapte a los cambios del proyecto. Los documentos del proyecto *MSF* son desarrollados iterativamente así como el código.

1.5.3.2.3. Definición de puntos de referencia (línea base), bloqueo

Creando y definiendo puntos de referencia de los documentos del proyecto de forma temprana en el proceso los miembros del equipo están autorizados a empezar el desarrollo del trabajo sin las demoras que puedan incurrir en una excesiva planeación.

1.5.3.2.4. Construcciones diarias

MSF defiende la preparación frecuente de construcciones de todos los componentes de la solución para pruebas y revisión. Este modelo es recomendado para desarrollar código así como “estructuras” de componentes *hardware* y *software*. Este modelo permite que la estabilidad de la solución total sea bien entendida, con una prueba de datos amplia, antes de que la solución sea liberada a la producción.

1.5.3.2.5. Administración de configuración

Es el rastreo formalizado y control del estado de varios elementos del proyecto. Estos elementos incluyen una versión de control para código, documentación, manuales de usuario, archivos de ayuda, cronogramas y planes. También incluye el rastreo de configuraciones de *hardware*, redes y *software* de una solución. La administración de configuración es el rastreo del estado de los entregables del proyecto y documentos.

1.5.3.2.6. Guía para liberaciones en versión.

Las liberaciones en versión mejoran las relaciones del equipo con el cliente y asegura que las mejores ideas sean reflejadas en la solución. Las guías para facilitar la adopción de liberaciones en versión son:

- Crear un plan multi_liberación: Decidir que construir ahora y que aplazar. Diseñar un cronograma.
- Entregar primero la funcionalidad principal: brindar un valor inmediato para tener una base sobre la construcción y retroalimentación que ayude a manejar el desarrollo en las iteraciones subsecuentes.

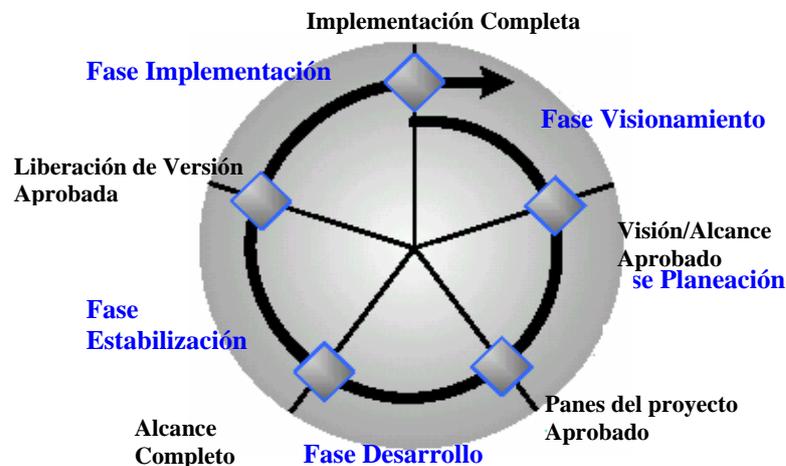
- Priorizar usando un cronograma de Manejo de riesgos: Identificar las características más arriesgadas y su impacto.
- Recorrer las iteraciones rápidamente: Mejorar el producto continuamente con la participación del cliente, a través del uso de las versiones.
- Establecer el control de cambio: Se debe considerar el control de cambio de las características y funcionalidad de la solución. Dichos cambios se harán sólo con autorización formal previo al análisis de viabilidad, impacto del costo y del cronograma.
- Detenerse en la creación de versiones nuevas cuando no agregan más valor.

1.5.3.3. Modelo integrado de desarrollo e implementación

Como se estableció anteriormente, una solución no provee un valor hasta que es completamente implementado en la vida de la producción. Es por esta razón que el modelo de proceso *MSF* sigue la trayectoria de una solución hasta el punto en el cual ésta comienza a entregar un valor (cuando la implementación se ha completado).

1.5.4. Notas para usar el modelo de proceso integrado

Mientras el modelo de proceso en el gráfico 1.7 muestra fases de igual tamaño, esto no implica que cada fase tome cantidades similares de tiempo ya que puede variar dependiendo del proyecto. Se puede pensar que las actividades asociadas con una fase son sólo hechas durante esa fase, sin embargo algunas tareas de cada fase pueden estar presentes en otras.



Gráficos 1.7. Hitos y fases del modelo de proceso *MSF*⁸ Microsoft *MSF WhitePaper*

Para que no exista confusión las fases deben estar caracterizadas por metas y entregables.

1.5.5. Fases del modelo de proceso e hitos

MSF consta de 5 fases, cada una de las cuales culmina en un hito externamente visible como se muestra en el gráfico 1.7.

1.5.5.1. Fase de visionamiento

La fase de visionamiento dirige uno de los requerimientos fundamentales para el éxito del proyecto. El equipo debe tener una clara visión de lo que quiere para cumplir con el cliente y estar en capacidad de establecerlo en términos que motivarán al equipo entero y al cliente. A través de la creación de una visión de alto nivel de metas de proyecto y restricciones puede servir como una forma temprana de planear; establecer la plataforma para el proceso de planeación más formal que tomará lugar durante la fase de planeación del proyecto. Las actividades primarias logradas durante el visionamiento son la formación del equipo principal y la preparación de un documento de visión/alcance. La delineación de la visión del proyecto y la identificación del alcance del proyecto son actividades distintas y son requeridas para un proyecto exitoso. La visión es una vista no limitada de lo que una solución puede ser. El alcance identifica la parte de la visión que puede ser lograda dentro de las restricciones del proyecto. La administración de riesgos es un proceso que continúa a lo largo del proyecto. Durante la fase de visionamiento, el equipo prepara un documento de riesgo, presenta los riesgos más altos de acuerdo con el documento de visión/alcance. Los requerimientos de negocios deben ser identificados y analizados (estos son definidos más rigurosamente durante la fase de planeación). El rol del equipo primario manejando la fase de visionamiento es el rol de la administración del producto.

El hito de visión/alcance aprobado culmina la fase, en este punto, el equipo del proyecto y el cliente se han puesto de acuerdo en lo referente a la dirección del proyecto, así como las características que incluirá y no incluirá en la solución, y un cronograma general para la entrega.

1.5.5.1.1. Entregables

Los entregables para la fase del visionamiento son:

- Documento de visión/alcance.
- Documento de valoración de riesgos.
- Documento de estructura del proyecto.
- El enfoque del Equipo durante la fase de visionamiento.

La tabla siguiente describe el enfoque y las áreas de responsabilidad de cada rol del equipo durante la fase de visionamiento.

ROL	ENFOQUE
Administración del Producto	Metas globales, identificar las necesidades del cliente, requerimientos: documento de visión/alcance.
Administración del Programa	Metas de diseño: concepto de solución; estructura del proyecto.
Desarrollo	Prototipos, opciones de desarrollo y tecnología; análisis de factibilidad.
Experiencia de Usuario	Necesidades de ejecución de usuario e implicaciones
Prueba	Estrategias de prueba, criterio de aceptación de prueba; implicaciones.
Administración en la Liberación	Implicaciones de implementaciones; administración de operaciones y soporte; criterio de aceptación operacional.

Tabla 1.2. Rol del Equipo en la Fase de Visionamiento⁹
Microsoft MSF WhitePaper

1.5.5.1.2. Hitos secundarios sugeridos

El **documento de estructura del proyecto** incluye información de como el equipo está organizado y quien juega que roles y ha especificado responsabilidades, también aclara la cadena de responsabilidades para el cliente y puntos designados de contacto que el equipo del proyecto tiene con el cliente. Esto puede variar dependiendo de las circunstancias del proyecto.

Visión/alcance redactados: la primera redacción del documento visión/alcance ha sido completado y es circulado entre el equipo, el cliente y los inversionistas para la

revisión. Durante el ciclo de revisión. El documento sufre iteraciones de retroalimentación, discusión y cambio.

1.5.5.2. Fase de planeación.

Durante esta fase el equipo prepara la especificación funcional, trabaja en el proceso del diseño y prepara planes de trabajo, estima costos, y cronogramas para los diferentes entregables. Temprano en esta fase, el equipo analiza y documenta requerimientos que caen en cuatro categorías extensas: requerimientos de negocios, usuarios, operacionales y de sistema (aquellos de la misma solución).

Mientras el equipo se mueve al diseño de la solución y crea las especificaciones funcionales, es importante mantener la habilidad de identificar entre los requerimientos y las características. Mantener la habilidad de identificar sirve como una forma de revisar la exactitud del diseño y verificar si el diseño conoce de las metas y requerimientos de la solución.

El proceso de diseño da al equipo una forma sistemática de trabajar desde el concepto abstracto a los detalles técnicos específicos. Esto comienza con un análisis sistemático de perfiles de usuario el cual describe varios tipos de usuarios y sus funciones de trabajo.

Gran parte de esto es a menudo realizado en la fase de visionamiento. Estos son divididos en una serie de escenarios de uso, donde un tipo particular de usuario está intentando completar un tipo de actividad, tal como la administración de claves de usuario para un sistema administrador. Finalmente cada escenario de uso es dividido en una secuencia específica de tareas, conocidas como casos de uso, el cual el usuario ejecuta para completar esa actividad.

Existen tres niveles en el proceso de diseño: el diseño conceptual, el diseño lógico, el diseño físico. Cada nivel es completado y definido en una línea base en una secuencia alternada.

Los resultados del proceso de diseño son documentados en las especificaciones funcionales que describen en detalle como cada característica se mira y comporta, también describe la arquitectura y el diseño para todas las características. La especificación funcional sirve para múltiples propósitos, tales como:

- Instrucciones a los desarrolladores de que construir.
- Bases para el trabajo estimado.
- Acuerdo con el cliente en lo que exactamente se construirá.
- Punto de sincronización para el equipo completo.

Una vez que la especificación funcional es definida como línea base, la planeación detallada puede iniciar.

En la culminación de la fase de planeación los clientes y los miembros del equipo han llegado a un acuerdo en detalle de lo que se va a entregar y cuando.

1.5.5.2.1. Planes del proyecto aprobados.

En el hito de los planes del proyecto aprobado, el equipo y los inversionistas claves del proyecto están de acuerdo en que los hitos secundarios han sido conocidos, que las fechas son realistas, que los roles y responsabilidades del proyecto están bien definidos, y que los mecanismos están establecidos para las áreas de dirección del riesgo del proyecto. Las especificaciones funcionales y el cronograma del proyecto proveen las bases para tomar decisiones de futuros intercambios. Después de que el equipo apruebe las especificaciones, planes y cronogramas, los documentos llegan a ser la línea base del proyecto. La línea base toma en cuenta las diferentes decisiones que son obtenidas por un consenso aplicando las tres variables de planeación del proyecto: recursos, cronograma, y características. Después la línea base es terminada y aprobada, el equipo pasa a la fase del desarrollo. Al definir el equipo una línea base, ésta es ubicada bajo el control de cambio. Esto no significa que todas las decisiones alcanzadas en la fase de planeación son las finales. Pero significa que mientras el trabajo progresa en la fase del desarrollo, el equipo revisaría y aprobaría cualquiera de los cambios sugeridos en la línea base.

1.5.5.2.2. Entregables

Los siguientes entregables son producidos durante la fase de planeación:

- Especificación funcional.
- Plan de administración de riesgo.

La siguiente tabla describe las áreas de enfoque y responsabilidad de cada rol del equipo durante la planeación.

Rol	Enfoque
Administración del producto	Diseño conceptual; análisis de requerimientos de negocios; plan de comunicaciones.
Administración del programa	Diseño conceptual y lógico; especificación funcional; estimados de desarrollo.
Desarrollo	Casos de usos de escenario, requerimientos de usuario, requerimientos de localización/accesibilidad; documentación de usuario, documentación de usuario, entrenamiento.
Experiencia del usuario	Evaluación de diseño; requerimientos de prueba.
Prueba	Evaluación de diseño; requerimientos de prueba.
Administración de liberación	Evaluación de diseño; requerimientos de operaciones.

Tabla 1.3. Rol del Equipo en la Fase de Planeación¹⁰
Microsoft MSF WhitePaper

1.5.5.3. Fase de desarrollo

En esta fase el equipo logra la mayor parte de la construcción de los componentes de la solución e involucra el desarrollo de código, el trabajo puede continuar en la fase de estabilización en respuesta a la pruebas. La infraestructura es desarrollada durante esta fase y todos los roles están activos en los entregables de construcción y prueba.

1.5.5.3.1. Hito del alcance completo

La fase de desarrollo culmina con el hito del alcance completo. En este hito, todas las características están completas y la solución está lista para la prueba externa y la estabilización.

Este hito es la oportunidad para los clientes y usuarios, personal de soporte, e inversionistas clave del proyecto para evaluar la solución e identificar cualquier problema que haya quedado y que debe ser revisado antes de que la solución sea liberada.

1.5.5.3.2. Entregables

Los entregables de la fase de desarrollo son:

- Código fuente y ejecutables.
- Instalación de scripts y conjuntos de configuración para la implementación.
- Especificación funcional congelada.
- Ejecución de elementos de soporte.
- Especificaciones de prueba y casos de prueba.

1.5.5.4. Fase de estabilización

La fase de estabilización conduce a la prueba en una solución cuyas características están completas, durante esta fase se enfatiza el uso y operación bajo condiciones de ambiente real. El equipo se enfoca en resolver y tratar (priorizando) los errores y preparando la solución para la liberación.

Una vez que una construcción ha sido considerada lo suficientemente estable como para ser un candidato a la liberación, la solución es implementada en un grupo piloto. La fase de estabilización culmina en la liberación de un hito disponible. Una vez revisada y aprobada, la solución está lista para ser completamente implementada en el ambiente vivo de la producción.

1.5.5.4.1. Hito de liberación disponible

El hito de liberación disponible ocurre en el punto en que el equipo tiene direccionado todos los problemas pendientes y han liberado la solución o la han puesto en uso. En este hito, la responsabilidad de continuidad de la administración y

soporte de la solución oficialmente se transfiere del equipo del proyecto a los equipos de operación y soporte.

1.5.5.4.2. Entregables

Los entregables de la fase de estabilización son:

- Liberación de notas.
- Ejecución de los elementos de soporte.
- Resultados de pruebas y herramientas de prueba.
- Código Fuente y ejecutables.
- Documentos del proyecto.
- Revisión de hito.

1.5.5.5. Fase de implementación

Durante esta fase, el equipo implementa la tecnología principal y los componentes, estabiliza la implementación, pasa el proyecto a operaciones y soporte y obtiene la aprobación final del cliente del proyecto. Después de la implementación, el equipo del proyecto conduce una revisión del proyecto y un estudio de satisfacción del cliente. Las actividades de estabilización pueden continuar durante este período mientras los componentes son transferidos desde un ambiente de prueba a un ambiente de producción.

1.5.5.5.1. Implementación del hito completo

El hito de implementación completo culmina con la fase de implementación. En este momento, la solución implementada debería proveer el valor de negocios esperado al cliente y el equipo debería tener terminado los procesos y actividades implementadas para alcanzar esta meta. El cliente debe estar de acuerdo que el equipo ha conocido sus objetivos antes de que este pueda declarar que la solución está en producción y culminar el proyecto. Esto requiere una solución estable, así como un criterio de

éxito establecido claramente. Para que la solución pueda ser considerada estable, las operaciones apropiadas y los sistemas de soporte deben estar establecidos.

1.5.5.5.2. Entregables

- Sistemas de operación e información de soporte.
- Procedimientos y procesos.
- Base de conocimiento, reportes, libros de registro.
- Reportes de culminación de proyecto.
- Versiones finales de todos los documentos del proyecto.
- Datos de satisfacción del cliente/usuario.
- Definición de pasos siguientes.

1.6 Conclusión

Es importante que todo proyecto tenga un control, más aún en proyectos de *software*, que es intangible y no se puede demostrar su funcionalidad si no hasta el final cuando la solución esté terminada. Por esta razón es importante definir la metodología que se va a usar para tener clara las actividades necesarias que llevarán a convertir en *software* los requerimientos de un cliente, tenerlos informados de los avances a través de documentos y poder entregar a tiempo la solución, cumpliendo las expectativas del cliente utilizando los recursos que se hayan establecido.

Para facilitar el proyecto de tesis se ha escogido la metodología *MSF* por que describe una secuencia de alto nivel de actividades para construir e implementar soluciones *IT*, además es una metodología probada en grandes proyectos de *Software* desarrollados por *Microsoft* e incluso el proceso unificado (*RUP*), puede ser usado como un complemento.

REFERENCIAS

¹JACOBSON, Ivar; BOOCH, Grady; RUMBAUGH, James. “El Proceso Unificado de desarrollo de *Software*”. *Pearson Education*. S.A. España, 2000, pag. 11.

^{2,3,4,5,6,7,8,9,10} MICROSOFT. *Microsoft Solutions Framework*. <http://www.microsoft.com/MSF> Junio 2003, [consulta 15 de octubre de 2005]

CAPÍTULO 2

TECNOLOGÍA DE INFORMACIÓN

2.1. Introducción

Debido a que la información es un recurso primordial en cualquier organización, siempre ha existido la necesidad de automatizarla, con el objetivo de almacenar los grandes volúmenes de información. En la actualidad no basta sólo con esto, además la información debe ayudar a la toma de decisiones en base a: informes gráficos y resúmenes que tengan el contenido necesario para tomar buenas decisiones, por esta razón ha evolucionado el concepto de tecnología de información, diseñando herramientas y métodos empleados para recabar, retener, manipular o distribuir información de una forma sencilla y rápida permitiendo a una empresa ser competitiva, al utilizar información que genere conocimiento justo a tiempo.

En este capítulo se detallan conceptos de Inteligencia de Negocios, recopilados de documentos electrónicos en los que se detalla la funcionalidad, características, ventajas, problemas que soluciona y componentes, además los sistemas de información Ejecutiva (*EIS*) y Sistema de Soporte a la toma de decisiones (*DSS*), con sus elementos característicos. Al final del capítulo se describe el concepto de la Administración de relaciones con el cliente (*Customer Relationship Management CRM*). Es importante conocer estrategias para mantener a nuestros clientes ya que estos están involucrados directamente con la rentabilidad del negocio, además son quienes generan día a día la información la cual debe ser analizada para la toma de decisiones.

2.2. Inteligencia de negocios (*BI*)

La Inteligencia de Negocios o *Business Intelligence (BI)* se puede definir como el proceso que permite analizar la información de la empresa y extraer un conocimiento de dicha información para permitir a las empresas actuar rápida y eficientemente para alcanzar sus metas. *BI* convierte las bases de datos operacionales en almacenes de datos inteligentes, es decir las transacciones cotidianas en información estratégica.

Posibilita la planeación y la toma de decisiones corporativas en tiempo real, combinando las mejores políticas organizacionales con la información, además de proporcionar información correcta, en el momento y lugar correcto, permitiendo incrementar la efectividad de una empresa.

La inteligencia de negocios tiene mucho éxito ya que le da una ventaja a las empresas sobre sus competidores al juntar a las personas y a la tecnología para resolver problemas. El vínculo entre la inteligencia de Negocios y la tecnología es fundamental, porque es la zona donde se traducen los datos en información útil para la organización. Es un puente que transforma, depura e integra datos.

Los sistemas analíticos o de toma de decisiones recuperan los datos de las operaciones diarias de la organización, generados por los sistemas transaccionales (*OLTP*) y los procesan con el objeto de tenerlos disponibles para quienes toman decisiones.

2.3. Funcionalidad de los sistemas analíticos

Las principales funciones que los sistemas analíticos incluyen son:

- Consultas *ad-hoc*: Permiten a los expertos tomadores de decisiones no depender del área de sistemas para realizar una consulta o cruce de información. Para efectuar este tipo de consultas es imprescindible que el área de sistemas genere previamente una descripción en el sistema, a través de la definición de metadatos y de los conceptos de negocio. Con ayuda de los metadatos, el informático hace “la traducción” entre los campos de la base de datos para formar los conceptos de negocios, que maneja quienes toman decisiones.
- Sistema de semáforos: visualiza el nivel de los indicadores claves del negocio para alertar sobre los problemas que puedan ocurrir en los procesos.
- Cruces de información.
- Sumarización.
- Comparación de conceptos de negocio en diferentes períodos.

- Ordenamientos y agrupamientos de información por región, tráfico, central entre otros conceptos.
- Generación de gráficos para visualizar de forma sencilla la información.

2.4. Niveles de decisiones de negocio

La toma de decisiones puede darse en tres niveles:

1. La alta dirección de las empresas, que puede analizar y monitorear tendencias, patrones, metas y objetivos de la organización. Sus decisiones son de tipo estratégico.
2. Los administradores y analistas, responsables de decisiones de tipo táctico, quienes lo hacen de manera analítica y también demandan información considerable.
3. Los empleados, que toman decisiones rutinarias y cuentan con datos limitados a su dominio. Sus decisiones son de tipo operacional.

Las herramientas dinámicas para toma de decisiones (como representaciones gráficas), son adecuadas para los ejecutivos y los administradores. Por su parte, las herramientas estáticas (herramientas de reportes u hojas de cálculo con un formato fijo cuya información se actualiza frecuentemente) son las más convenientes para los empleados.

2.5. Problemas que resuelve la inteligencia de negocios

La solución *BI* es aplicada a empresas donde se dan los siguientes problemas:

- Frustración por no poder encontrar información que está dentro de la empresa, que existe y es necesaria para el análisis.
- Quienes trabajan con la información se pasan más tiempo recolectándola que analizándola.
- Pérdida de tiempo tratando de dar formato a reportes en Excel.
- Pérdida de oportunidades de negocio por recibir información retrasada.

- Falta de una guía que indique lo que sucedió cuando administradores anteriores implementaron una determinada estrategia.
- Desconocimiento de que hacer con toda la información que se tiene.
- No poder conocer los productos más rentables durante un período determinado.
- No poder saber si quienes laboran en la empresa están alcanzando los objetivos.
- No conocer si se mantiene una comunicación estrecha entre diversas áreas de su empresa hacia una estrategia común.

2.6. Componentes *BI*

Multidimensionalidad: La información multidimensional se puede encontrar en hojas de cálculo, bases de datos, etc. Una herramienta de *BI* debe de ser capaz de reunir información dispersa en toda la empresa e incluso en diferentes fuentes para así proporcionar a los departamentos la accesibilidad, poder y flexibilidad que necesitan para analizar la información.

Data Mining: Las empresas suelen generar grandes cantidades de información sobre sus procesos. Pero el éxito de los negocios depende por lo general de la habilidad para ver nuevas tendencias o cambios en las tendencias. Las aplicaciones de *Data Mining* pueden identificar tendencias y comportamientos, no sólo para extraer información, sino también para descubrir las relaciones en bases de datos que pueden identificar comportamientos que no son muy evidentes.

Data Warehouse: Es la respuesta de la tecnología de información a la descentralización en la toma de decisiones. Coloca información de todas las áreas funcionales de la organización en manos de quien toma las decisiones. También proporciona herramientas para búsqueda y análisis.

2.7. Ventajas de una solución *BI*

- Poder relacionar toda la información de negocio disponible.
- Permite cumplir objetivos.

- Reducir costos y competir más eficazmente en el mercado, debido a la disponibilidad de las soluciones en menor tiempo y su aplicación con poco esfuerzo de desarrollo.
- Ampliar la capacidad de análisis y obtención de conocimientos ocultos en un entorno de negocio complejo, resaltando el descubrimiento de nuevas oportunidades de relación con el cliente (comercial, de marketing) y mejora en la gestión del negocio (detección de desviaciones ó estándares definidos).
- Aplicar nuevas tecnologías de la información para acceder, presentar y distribuir la información de forma ágil y eficaz.

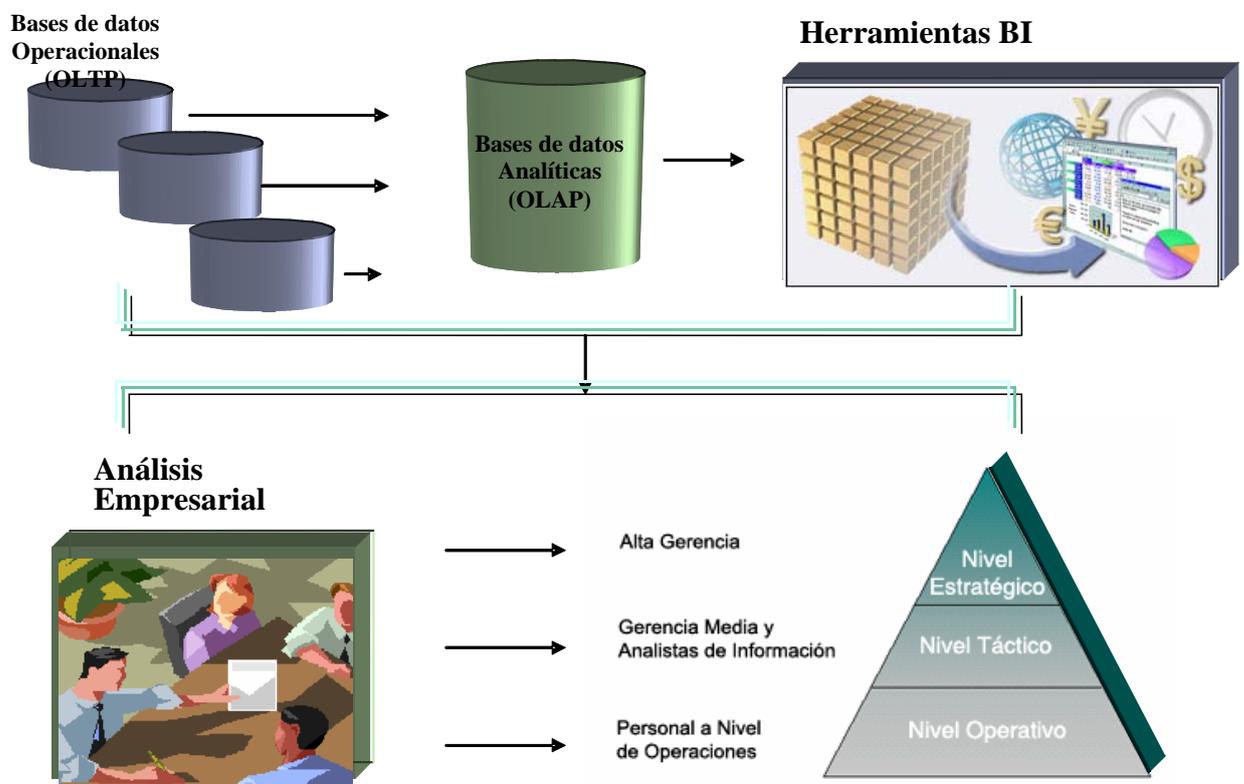


Gráfico 2.1. Esquema Conceptual Inteligencia de Negocios

2.8. Sistemas de información

Los sistemas de información deben diseñarse para responder dinámicamente lo que se necesita saber en una empresa y tener acceso a todos los conocimientos de la misma como respuesta a los cambios locales y rápidos del entorno de trabajo. Quienes manejan la información han de ser capaces de anticiparse y responder a los cambios lo antes posible a través del uso de los sistemas de información, ya que éstos deben suministrar la información a tiempo.

Algunos de estos sistemas deben permitir también tomar decisiones efectivas y en tiempo real frente a dichos cambios. Uno de los sistemas de información es el llamado Sistema de Soporte de Decisión (*DSS: Decision Support Systems*) que permite a los directivos desarrollar sus propios modelos de decisión, bases de datos, etc. Los sistemas de soporte de decisión aportan flexibilidad y se adaptan a las necesidades cambiantes. Los sistemas de soporte de decisión que están orientados hacia los directivos se suelen llamar Sistemas de Información para Ejecutivos (*EIS: Executive Information Systems*).

2.8.1. Sistema de información ejecutiva (*EIS*)

EIS se refiere a cualquier sistema de *software* que muestre información ejecutiva de las diferentes áreas del negocio en un sólo sistema, facilitando el monitoreo de la empresa. Es una técnica de más alto nivel dentro de las herramientas de *BI* y tiene como objetivo primordial proveer toda la información necesaria a la gente que toma decisiones, de modo fácil y prácticamente sin necesidad de interactuar con el mismo sistema. En términos formales, un *EIS* es un sistema de información que permite a los ejecutivos acceso rápido y efectivo a información compartida, crítica para el negocio, utilizando interfaces gráficas.

Las interfaces que son utilizadas en estos sistemas deben ser más sofisticadas que los sistemas transaccionales y deben incluir, en el menor número de páginas posible, la mayor cantidad de información que el usuario necesita para monitorear su empresa. La función principal de un *EIS* es el monitoreo de la empresa y dicho monitoreo implica una visualización de lo que se quiere controlar, por tal motivo, no debe existir mayor esfuerzo por parte del usuario que simplemente ve las consultas que cotidianamente realiza.

La información que un *EIS* presenta es producto de indicadores empresariales que cotidianamente son evaluados para analizar su comportamiento y lo que permite determinar su cualidad son las propiedades que le son asignadas a cada indicador. En caso de que los usuarios deseen ver más información relacionada y entrar en detalle, estarían entrando a un Sistema de Soporte de Decisiones (*DSS*), el cual, a pesar de que en conjunto con un *EIS*, forman una poderosa solución de *BI*, merece una

explicación independiente por la diferencia de su uso. Debido a la función principal de monitoreo, un *EIS* es considerado como un sistema muy fácil de usar y de desarrollar, pero con funcionalidades analíticas muy limitadas.

La diferencia de los *EIS* con otros sistemas no sólo es la facilidad de uso. Aparte del *front-end*, los *EIS* interpretan y manipulan de forma diferente la información, pues trabajan con formatos de datos no típicos, tales como *Data Warehouse* o *Data Mart*. Prácticamente todos los *EIS* obtienen sus datos de matrices multidimensionales denominadas "cubos" y las herramientas que desarrollan estos sistemas tienen tecnología que permite consultas amplias y complejas de diversas fuentes de datos.

Las partes importantes de un *EIS* son: la interfaz de usuario y la base de datos multidimensional, esto montado en una arquitectura Cliente/Servidor.

El Cliente es precisamente el usuario final, a quien se le debe presentar, de una forma amigable y sencilla, la mayor cantidad de información posible para que le apoye a visualizar sus variables de control. Cada escenario en un *EIS* se adecua a los casos particulares de cada usuario, proveyendo la información necesaria para tomar decisiones.

2.8.1.1. Lo que debe contener un *EIS*

Para ser considerado un *EIS*, un sistema debe reunir una serie de características adicionales a las que por defecto debe contener al formar parte de la Tecnología *BI*, es decir, brindar información y que sirva de apoyo a la toma de decisiones.

2.8.1.1.1. Interfaz gráfica fácil de usar y ver

Una característica de los *EIS* es poseer interfaces gráficas sencillas, que tengan una curva de aprendizaje corta y además, deberán ser vistosas e intuitivas para facilitar la labor de monitoreo del tomador de decisiones. La idea es que se interactúe con el sistema el menor tiempo posible.

2.8.1.1.2. Alarmas o semáforos

Un típico *EIS* cuenta con funciones que le permiten al usuario notar rápidamente los errores y los valores destacables de la información para que el ejecutivo sólo ponga atención donde se han sobrepasado ciertos rangos de tolerancia.

2.8.1.1.3. Integración de información proveniente de los cubos

Un *EIS* debe permitir integrar información de cualquier aplicación y presentársela al usuario final de una forma transparente para él, por eso se crea la base para no moverse entre aplicaciones.

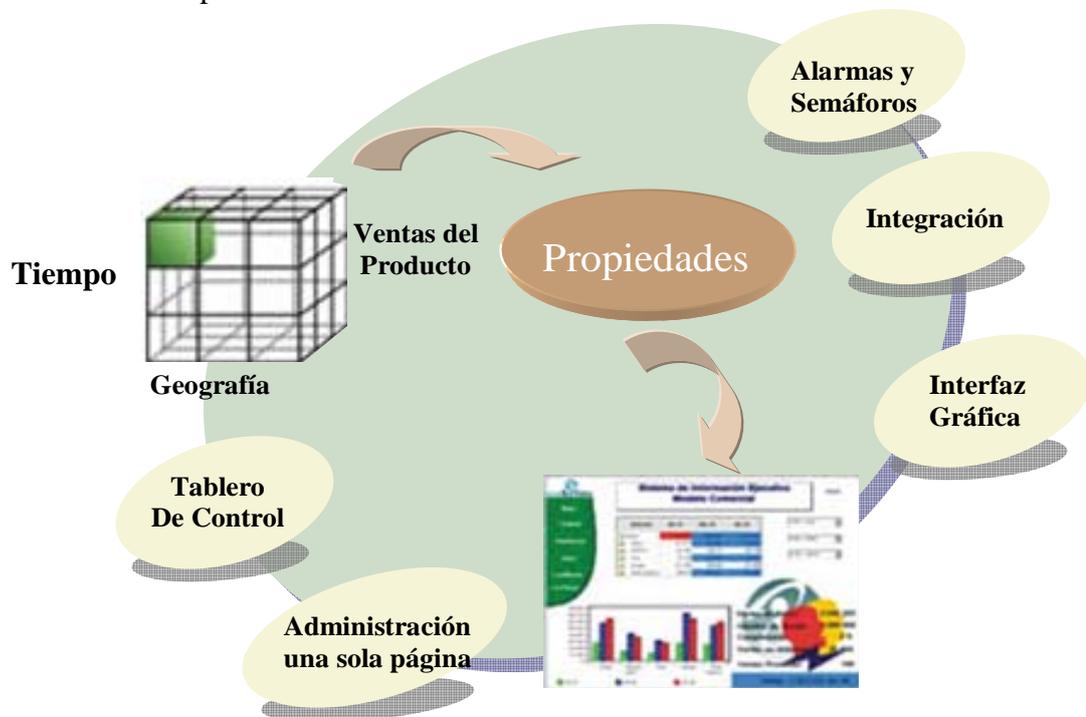


Gráfico 2.2. Esquema Conceptual Sistema de Información Ejecutivo ¹
www.bitam.com (consulta en línea)

2.8.2. Sistema de soporte de decisiones (DSS)

El uso exagerado de los *DSS* con frecuencia lo lleva a rebasar límites de su aplicación y confundirse con términos como *OLAP*, *Data Warehouse* ó *EIS*, lo cierto es que, independientemente, del término que llegue a usurpar, siempre se asocia al soporte a la toma de decisiones y de alguna forma, todos los conceptos señalados tienen en la toma de decisiones el punto de encuentro.

2.8.2.1. Definición

DSS se refiere a cualquier sistema de *software* que permite el análisis de las diferentes variables del negocio para apoyar una decisión. Puede considerarse como un sistema que se basa en un almacén de datos y crea una base de datos multidimensional, permitiéndole al usuario procesar analíticamente la información en línea (*OLAP*), con la habilidad de poder girar y taladrar dentro de los datos lo que permite hacer uso de *OLAP*, *Data Warehouse* y otros conceptos asociados.

Los *DSS* están asociados a las jerarquías dentro de los conceptos de los cubos del *Data Warehouse*, debido a que cuando se desea conocerse el "por qué" de un dato visto en un *EIS*, el *DSS* debe permitir conocer el detalle de la misma y de las partes que lo componen, es donde se revisan las jerarquías que fueron definidas en los cubos de información, para ver por cual de las rutas se profundiza hacia el detalle. Los *DSS* son la contraparte de los sistemas *OLTP* que apoyan la transacción diaria.

2.8.2.2. Características

En un *DSS* la **consistencia** se mide globalmente. A pesar de que no importan las transacciones individuales, son precisamente estas las que, en suma, conforman una gran transacción completa y consistente. De hecho, antes de poder formar una gran transacción, a partir de los sistemas operacionales, debe verificarse la calidad de los datos para garantizar la información correcta y coherente.

Un sistema *DSS* procesa una transacción por día (puede ser por cualquier período de tiempo, lo común es por día), pero esa transacción es producto de miles o millones de registros que han sido procesados en el día. En vez de llamarle transacción, se le llama carga de información de producción. En este caso lo importante es el estado de consistencia del sistema antes de la carga y el estado de consistencia una vez que se ha terminado de efectuar la carga. En cuanto a los usuarios y administradores de un *DSS*, no son los que ingresan cada una de las operaciones en sus PCs, sino los que solicitan una o dos hojas con resúmenes totalizados de miles de operaciones. El usuario de *OLTP* cambia con frecuencia el tipo de información que solicita, de ahí que sus requerimientos no son planeados, sino heurísticos.

Una característica importante de los *DSS* es que la demanda puede ser muy variable, por lo que es normal implementarlo en una máquina distinta de *OLTP* convirtiéndose en un servidor que puede ser accedido por diversos usuarios, pero también puede tener información integrada de múltiples sistemas remotos de *OLTP*. El tiempo es un factor importante de los *DSS* visto como dimensión. La inconsistencia temporal se debe evitar en un *DSS*. Mientras que un *OLTP* es complicado para explicar la historia, un *DSS* lo que hace es tomar fotografías instantáneas de la empresa en un momento determinado de la historia, al sumar esa serie de fotografías se conforman capas que pueden explicar como era la empresa en determinado período de la historia resolviendo los problemas de representación de tiempo que se tiene en los sistemas *OLTP*:

- Primero, el *DSS* permanece quieto durante el día cuando los usuarios están realizando consultas.
- Después, almacenando cuidadosamente la información de cada instantánea del *DSS*, se puede representar cualquier momento en el tiempo correctamente.

Las fotografías instantáneas del sistema *OLTP* se llaman 'extracción de la información de producción', mientras que el envío hacia el *DSS* se llama 'migración de la información de producción'.

2.8.2.3. Ciclo de vida

El ciclo de vida de un *DSS* es una situación muy particular. Mientras que en un tradicional ciclo de vida de un sistema se deben entender los requerimientos, analizar las etapas del diseño y después desarrollar; en un *DSS* el ciclo de vida es inverso, es decir, inicia con la información de los sistemas *OLTP*, una vez que se tiene, es integrada y se prueba para saber cual es real y útil y cual no, luego se escriben los programas o *scripts* para manejar esta información y por último, los programas se analizan, y finalmente se entienden los requerimientos del sistema.

El ciclo de vida de un *DSS* se orienta a la información, mientras que el de *OLTP* se orienta a los requerimientos.

tiempo de vida del valor de los clientes”. Ayuda a que la organización brinde cortesía, honestidad y fidelidad, cualidades que deberían ser garantías para el cliente, sin embargo los negocios están enfocados a obtener ganancia, por lo cual desconocen que el pilar fundamental son los clientes, por esta razón el *CRM* se encarga de:

- Conocer mejor a nuestros clientes.
- Interactuar apropiadamente con nuestros clientes.

Todos los días en los negocios se recolecta grandes cantidades de información acerca de clientes, se recomienda hacer un análisis para conocer las circunstancias y llegar a comprender el comportamiento de los mismos.

2.9.1. Definición

Es una estrategia de negocio que sitúa al cliente como el corazón de la compañía, permitiendo a las empresas identificar, atraer y retener a sus clientes ayudando a incrementar la satisfacción de éstos y a optimizar así la rentabilidad de sus negocios.

CRM como estrategia no sólo implica disponer del *software* adecuado que permita gestionar las relaciones con los clientes, sino que además, supone un cambio en los procesos de la empresa y el involucramiento de todos los empleados de la misma para que esta estrategia tenga éxito.

2.10. Conclusión

La Tecnología de Información facilita, tener un control de la información, que fluye dentro de una empresa en los diferentes niveles, tomando como parte la inteligencia de negocios, herramientas que permiten generar conocimiento proporcionando reportes, pronósticos, tendencias, alarmas, utilizando los sistemas de información ejecutiva y sistemas de soporte de decisiones, con el único objetivo de tomar las mejores decisiones en base al comportamiento real de la empresa y utilizando herramientas que no necesiten de un experto para ser manipuladas, siendo el único requisito que conozca el negocio de la empresa.

Administración de las relaciones con el cliente (*CRM*) permite alinear la estrategia corporativa alrededor de los Clientes (en lugar de los productos y los canales de distribución). Una aplicación orientada a soportar el *CRM* permitirá el intercambio dinámico de información de los Clientes en todos los niveles de la empresa, centralizada en un ambiente que la administre de manera consistente y actual, analizando para entender, anticipar y satisfacer las necesidades del cliente con el fin de generar lealtad. La implementación de una estrategia de *CRM* debe estar respaldada por un grupo multidisciplinario conformado por gente de todos los niveles y áreas de la organización, con el respaldo de la gerencia, para poder direccionar los procesos, ajustar las necesidades e implementar una solución exitosa y acorde con las necesidades de la compañía, estrategias corporativas y su visión.

En el desarrollo esta tesis no se utiliza *CRM* porque el área de negocio en la que se está implementando no está orientada al cliente.

REFERENCIAS

¹ BITAM. Acerca de *Business Intelligence*. <http://www.bitam.com/spanish/TecAnalisis.html> 2004, [consulta 10 de abril de 2006]

CAPÍTULO 3

DATA WAREHOUSE

3.1. Introducción

Las organizaciones siempre han usado los datos desde sus sistemas operacionales para atender sus necesidades de información, enfocándose a tomar decisiones muchas veces en datos inconsistentes que no son confiables, por estas razones se considera como una solución al *Data Warehouse*.

En este capítulo se detalla los conceptos básicos sobre la tecnología *Data Warehouse*, sus componentes *Data Marts* los mismos que forma parte de la Inteligencia de Negocios. Conoceremos sus características, arquitectura y como construir una base de datos única e integrada, utilizando los procesos de extracción, transformación y carga.

3.2. Definición

Un *Data Warehouse* es una alternativa tecnológica y de administración de negocios, que cubre los aspectos del manejo de información para la toma de decisiones, desde su extracción en los sistemas, depuración, transformación, el diseño de estructuras de datos o modelos especiales para el almacenamiento de los mismos hasta la explotación de la información mediante herramientas comerciales de fácil uso. Es un repositorio central que contiene la información más valiosa de la empresa. Los datos que se almacenan en un *DWH* han pasado por un proceso que asegura su consistencia, además, el acceso a dicho repositorio se puede realizar de forma rápida.

3.3. Características

Un *Data Warehouse* es una base de datos que contiene cuatro características: ¹

- Orientado a tema.
- No volátil.

- Integrada.
- De tiempo variante.

3.3.1. Orientado a tema

La información se clasifica en base a aspectos de interés para la empresa. Los datos tomados están en contraste con los clásicos procesos orientados a las aplicaciones. Un *Data Warehouse* se diseña para consultar eficientemente la información relativa a actividades básicas de la organización, y no para exportar los procesos que se realizan en la misma. El Gráfico 3.1. muestra los dos tipos de orientaciones.

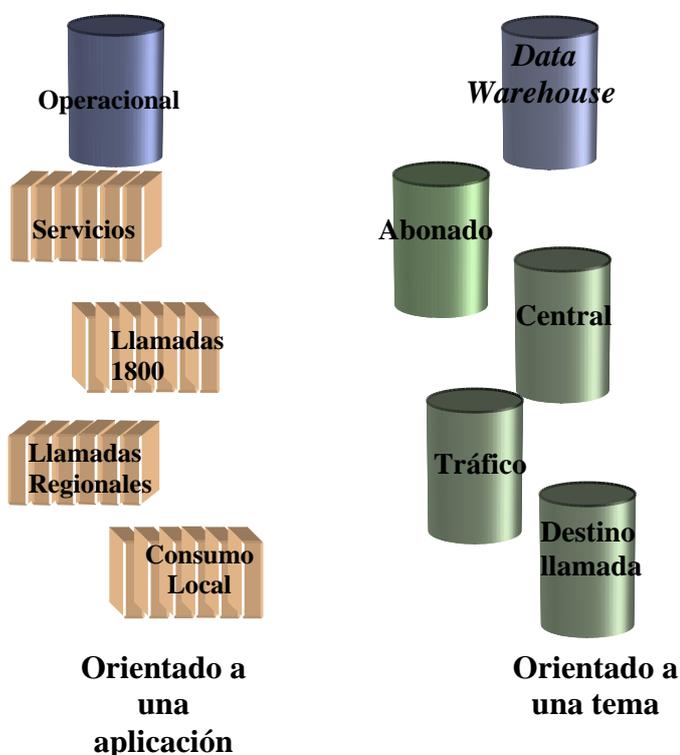


Gráfico 3.1. Características de Data Warehouse Orientado al Tema ²
www.sqlmax.com/dataw1.asp (consulta en línea)

El ambiente operacional se diseña alrededor de las aplicaciones, la base de datos combina elementos en una estructura que acomoda las necesidades de la aplicación. El ambiente *Data Warehousing* se organiza alrededor de sujetos. Las diferencias entre la orientación de procesos de las aplicaciones y la orientación a temas, radican en el contenido de la *Data* a nivel detallado. En el *Data Warehouse* se excluye la información que no será usada por el proceso de sistemas de soporte de decisiones,

mientras que la información de los sistemas orientados a aplicaciones contiene datos para satisfacer de inmediato los requerimientos funcionales y de proceso, que pueden ser usados o no por el analista de soporte de decisiones. Otra diferencia importante está en la interrelación de la información. Los datos operacionales mantienen una relación continua entre dos o más tablas basadas en una regla comercial que está vigente. Los del *Data Warehouse* miden un espectro de tiempo y las relaciones encontradas son muchas.

3.3.2. No volátil

Significa que la *Data* una vez establecida en la bodega, no es usualmente tema de cambio. Cualquiera que está usando la base de datos tiene confianza de que una consulta siempre producirá el mismo resultado no importa cuan a menudo sea ejecutada. Las bases de datos operacionales son extremadamente volátiles en el sentido que están constantemente cambiando es probable que una consulta no produzca el mismo resultado dos veces si está accediendo a tablas que se están actualizando frecuentemente.

En el gráfico 3.2. se muestra que la actualización (insertar, borrar y modificar), se hace regularmente en el ambiente operacional sobre una base de registro por registro pero la manipulación básica de los datos que ocurre en el *Data Warehouse* es mucho más simple. Hay dos únicos tipos de operaciones: la carga inicial de datos y el acceso a los mismos.

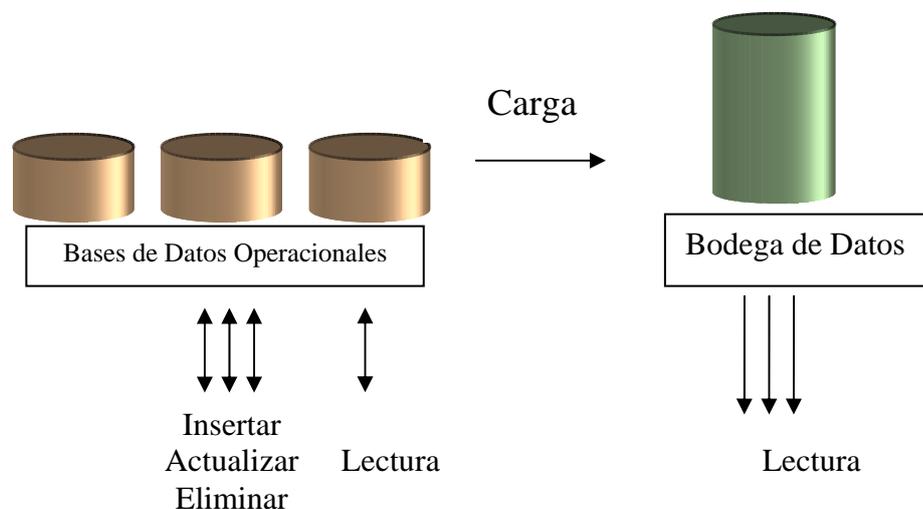


Gráfico 3.2. Actualización de datos (Característica no volátil)

3.3.3. Integrada

La integración de datos se muestra de muchas maneras: en convenciones de nombres consistentes, en la medida uniforme de variables, en la codificación de estructuras consistentes, en atributos físicos de los datos, fuentes múltiples y otros. Se debe asegurar que se codifique en un sólo formato (aa/mm/dd) ó contenido del dato (masculino/femenino ó 1/0), se estandarice tomando en cuenta las medidas (cm, yardas, minutos, segundos) y el uso de un mismo nombre para la descripción del dato.

El contraste de la integración encontrada en el *Data Warehouse* con la carencia de integración del ambiente de aplicaciones, se muestran en el gráfico 3.3. con diferencias bien marcadas.

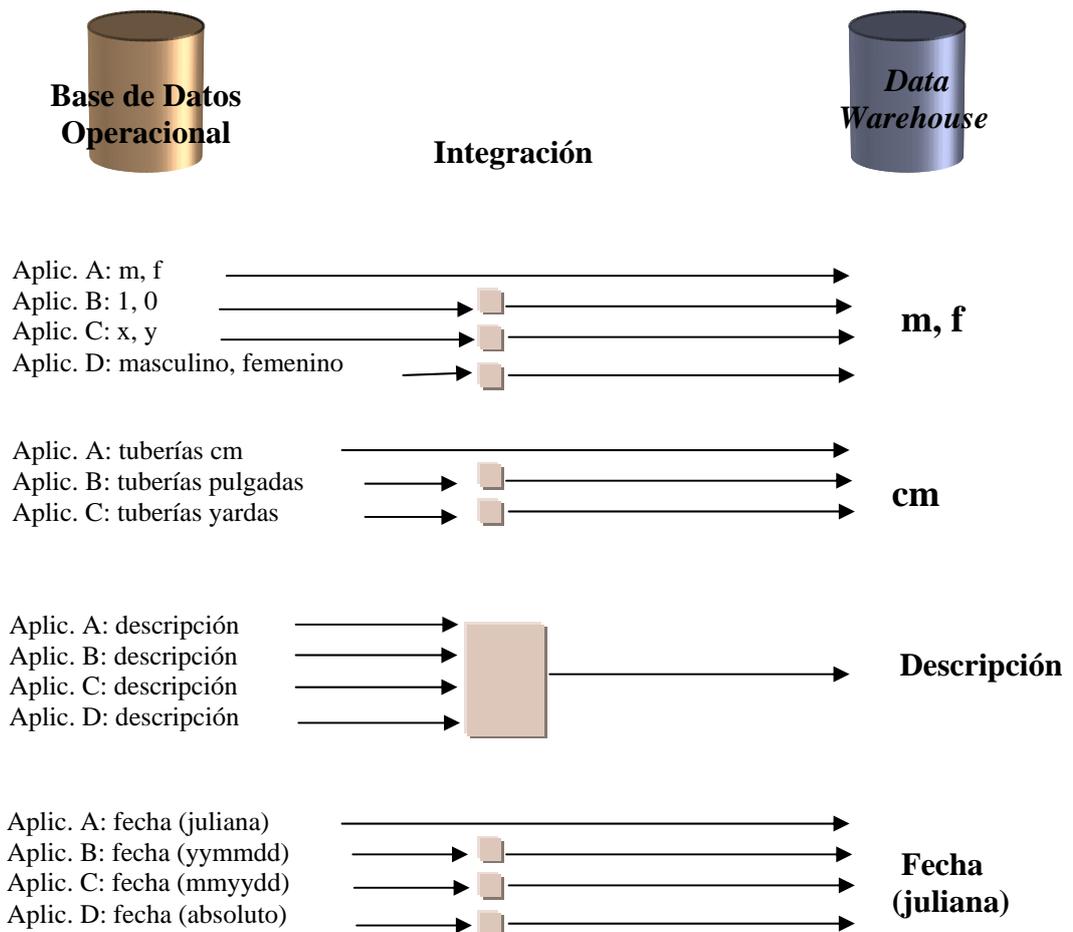


Gráfico 3.3. Integración de datos

3.3.4. Tiempo variante

Significa que la *Data* histórica es registrada, casi todas las consultas ejecutadas contra un *Data Warehouse* tienen algún elemento de tiempo asociado dentro de ellas. Ya se ha establecido que la mayoría de sistemas operacionales no retienen información histórica. Es casi imposible predecir lo que pasará en el futuro sin observar lo que sucedió en el pasado. Un *Data Warehouse* ayuda a dirigir este caso fundamental añadiendo una dimensión histórica a la *Data* tomada desde las bases de datos operacionales.

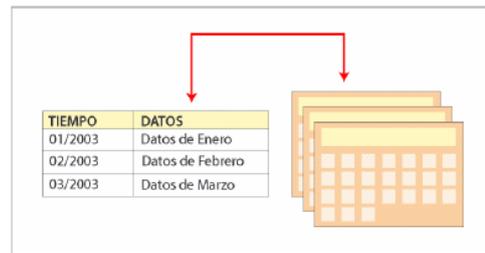


Gráfico 3.4. Tiempo Variante ³
www.exa.unne.edu.ar (consulta en línea)

3.4. Estructura del *Data Warehouse*

Hay niveles diferentes de esquematización y detalle que delimitan el *Data Warehouse*. En el gráfico 3.5. se muestran los diferentes componentes del *Data Warehouse* :

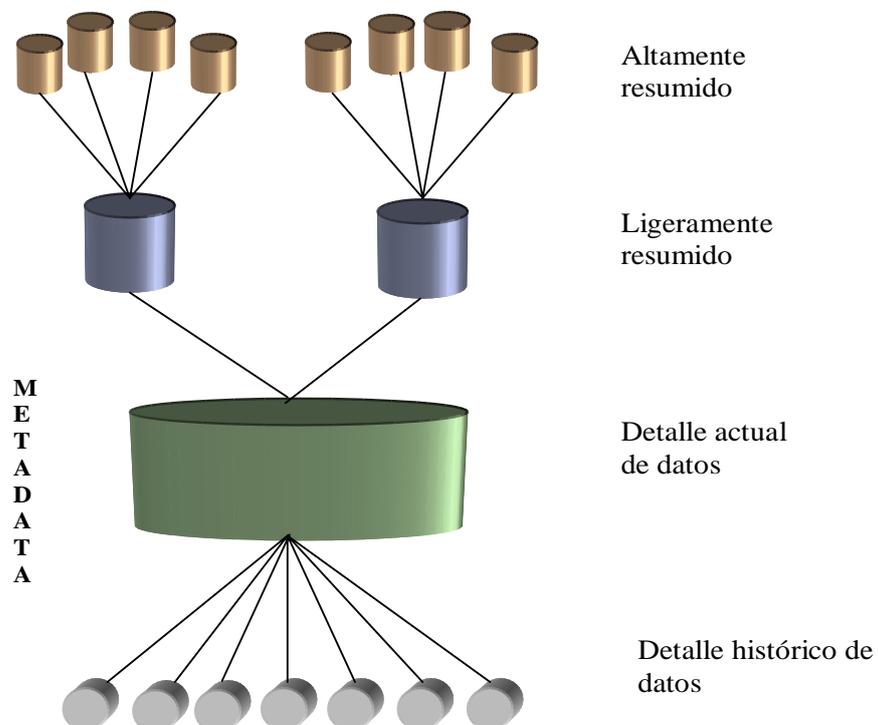


Gráfico 3.5. Niveles de esquematización que podrían encontrarse en un *DWH* ⁴
www.sqlmax.com/dataw1.asp (consulta en línea)

3.4.1. Detalle de datos actuales

En gran parte, el interés más importante radica en el detalle de los datos actuales, debido a que:

- Refleja las ocurrencias más recientes, las cuales son de gran interés.
- Es voluminoso, ya que se almacena al más bajo nivel de granularidad.

Casi siempre se almacena en disco, el cual es de fácil acceso, aunque su administración sea costosa y compleja.

3.4.2. Detalle de datos antiguos

La *Data* antigua es aquella que se almacena sobre alguna forma de almacenamiento masivo, a un nivel de detalle, consistente con los datos detallados actuales y no se accede con frecuencia por está razón y por el gran volumen de datos se puede almacenar en dispositivos externos y no en disco.

3.4.3. Datos ligeramente resumidos.

La *Data* ligeramente resumida es aquella que proviene desde un bajo nivel de detalle encontrado al nivel de detalle actual. Este nivel del *Data Warehouse* casi siempre se almacena en disco. Los puntos en los que se basa el diseñador para construirlo son:

- Que la unidad de tiempo se encuentre sobre la esquematización hecha.
- Qué contenidos (atributos) tendrá la *Data* ligeramente resumida.

3.4.4. Datos completamente resumidos

Estos datos son compactos y fácilmente accesibles. A veces se encuentra en el ambiente del *Data Warehouse* y en otros, fuera del límite de la tecnología que ampara al *Data Warehouse* (de todos modos, los datos completamente resumidos son parte del *Data Warehouse* sin considerar donde se alojan físicamente).

3.4.5. Metadata

De muchas maneras la metadata se sitúa en una dimensión diferente al de otros datos del *Data Warehouse*, debido a que su contenido no es tomado directamente desde el ambiente operacional. La metadata juega un rol especial e importante en *Data Warehouse* y es usada como:

- Un directorio para ayudar a ubicar los contenidos del *Data Warehouse*.
- Una guía para el *mapping* de datos de cómo se transforma del ambiente operacional al *Data Warehouse*.
- Una guía de los algoritmos usados para la esquematización entre el detalle de datos actual, con los datos ligeramente resumidos y éstos con los datos completamente resumidos, etc.
- La metadata juega un papel mucho más importante en un ambiente *Data Warehousing* que en un operacional clásico.

3.5. Elementos constituyentes de una arquitectura *Data Warehouse*

Una Arquitectura *Data Warehouse* (*Data Warehouse Architecture - DWA*) es una forma de representar la estructura total de datos, comunicación, procesamiento y presentación que existe para los usuarios finales que disponen de una computadora dentro de la empresa.

La arquitectura se constituye de un número de partes interconectadas:

3.5.1. Identificación de fuentes

Se identifican las bases, servidores ó archivos de los cuales se va a extraer la información. En caso de no existir información almacenada en una base de datos transaccional es posible utilizar Diccionarios de Datos, donde esté la información que se requiere. Se considerará la utilización de estos cuando se trate de un número de información pequeño.

3.5.2. Nivel de organización de datos

Incluye todos los procesos necesarios como seleccionar, editar, resumir, combinar y cargar datos en el depósito y acceder a la información desde bases de datos operacionales y/o externas. La organización de datos involucra con frecuencia una programación compleja, pero cada vez más, están creándose las herramientas *Data Warehousing* para ayudar en este proceso.

3.5.3. Nivel de acceso a la información

El nivel de acceso a la información de la arquitectura *Data Warehouse*, es el nivel del cual el usuario final se encarga directamente. En particular, representa las herramientas que el usuario final normalmente usa día a día. Este nivel también incluye el *hardware* y *software* involucrados en mostrar información en pantalla y emitir reportes de impresión, hojas de cálculo, gráficos y diagramas para el análisis y presentación. Actualmente, existen herramientas más sofisticadas para manipular, analizar y presentar los datos.

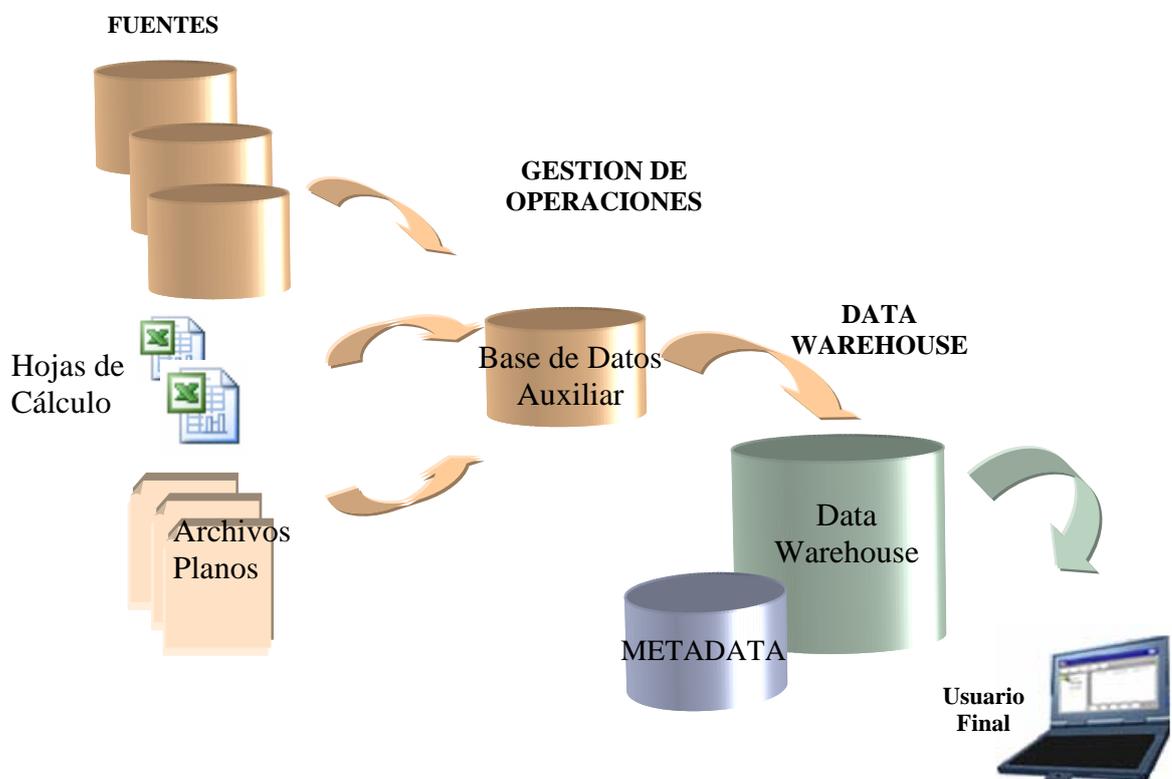


Gráfico 3.6. Elementos de la Arquitectura *Data Warehouse*

Después de conocer la arquitectura, se debe tomar en cuenta la gestión de procesos que tiene que ver con la programación de diversas tareas que deben realizarse para construir y mantener el *Data Warehouse* y la información del directorio de datos, depende del alto nivel de control de trabajo para muchos procesos (procedimientos) que deben ocurrir para mantener el *Data Warehouse* actualizado.

En el *Data Warehouse* (núcleo) es donde ocurre la *Data* actual, usada principalmente para usos estratégicos. En algunos casos, se puede pensar en el *Data Warehouse* simplemente como una vista lógica o virtual de datos. En muchos ejemplos, el *Data Warehouse* puede no involucrar almacenamiento de datos.

3.6. Operaciones dentro del ambiente *Data Warehouse* ⁵

3.6.1. *ETL* (Extracción, Transformación y Carga)

Es necesario cargar el *Data Warehouse* regularmente para que pueda servir a su propósito de facilitar el análisis de negocios. Para hacer esto, la *Data* de uno o más sistemas operacionales necesita ser extraída y copiada en el *Data Warehouse*. El proceso de extraer datos desde sistemas fuente y traerlos al *Data Warehouse* comúnmente es llamado *ETL*, el cual comprende la extracción, transformación y carga.

El acrónimo *ETL* es talvez muy simple, porque omite la fase de transporte e implica que cada una de las otras fases del proceso es distinta. Nos referimos al proceso completo incluyendo la carga de datos como *ETL*. Se debe entender que *ETL* se refiere a un amplio proceso, y no sólo a tres pasos bien definidos.

La metodología y tareas de *ETL* han sido bien conocidas por varios años y no son necesariamente únicas de los ambientes *Data Warehouse*: una gran variedad de aplicaciones y sistemas de bases de datos son la columna vertebral de las Tecnologías de Información de cualquier empresa. Los ambientes *Data Warehouse* no sólo tienen que intercambiar datos sino también integrar, reordenar y consolidar la *Data* sobre algunos sistemas, así proveen una nueva base unificada de información para inteligencia de negocios.

3.6.1.1. Proceso *ETL*

El proceso *ETL* comprende la extracción de la *Data* que se necesita en la empresa, después de extraer la *Data*, esta tiene que ser físicamente transportada al sistema de bodega de datos o a un sistema intermedio para su procesamiento. Dependiendo de la forma que se elija de transportación, algunas transformaciones pueden ser hechas durante este proceso, algo similar con lo que sucede en la extracción.

3.6.1.2. Operaciones diarias

Las cargas exitosas y transformaciones deben ser planificadas y procesadas en un orden específico. Dependiendo del éxito o fracaso de la operación o partes de ella, el resultado debe ser rastreado y subsecuentemente, procesos alternativos pueden ser iniciados. El control del progreso así como la definición de un flujo de trabajo de negocios de las operaciones son típicamente dirigidos por herramientas *ETL*.

3.6.1.3. Evolución del *Data Warehouse*

Al pertenecer a un sistema *IT* las fuentes y objetivos del *Data Warehouse* pueden cambiar. Estos cambios pueden ser mantenidos y rastreados a lo largo de la vida del sistema sin sobrescribir o borrar el antiguo proceso de flujo de información *ETL*. Para construir y conservar un nivel de confianza acerca de la información en el *Data Warehouse*, el flujo de proceso de cada registro individual en el *Warehouse* puede ser reconstruido en cualquier punto en el tiempo ó en el futuro en un caso ideal.

3.6.1.4. Extracción en *Data Warehouses*

La extracción es la operación que se encarga de tomar los datos de un sistema fuente para luego usarlos en un ambiente *Data Warehouse*. Este es el primer paso del proceso *ETL*. Diseñar y crear el proceso de extracción es a menudo una de las tareas que consumen más tiempo en el proceso *ETL* y de hecho en el proceso completo *Data Warehousing*. Durante la extracción, la *Data* deseada es identificada y extraída desde diferentes fuentes, incluyendo los sistemas de base de datos y aplicaciones. Muy a menudo, es imposible identificar a primera vista el subconjunto específico de

interés, por consiguiente más datos de lo necesario son extraídos, por ello la identificación de la *Data* es relevante.

Después de la extracción, esta *Data* puede ser transformada y cargada en el *Data Warehouse*. Los sistemas fuente para un *Data Warehouse* son típicamente aplicaciones de proceso de transacciones.

3.6.1.4.1. Consideraciones para la extracción

En la extracción los sistemas fuente pueden ser muy complejos y tener una documentación pobre y así determinar que *Data* es la que se necesita. La *Data* tiene que ser extraída normalmente no sólo una vez, sino varias veces de manera periódica para proveer toda la *Data* cambiada al *Data Warehouse* y conservarla actualizada a la fecha. Además, el sistema fuente normalmente no puede ser modificado, ni puede su ejecución y disponibilidad ser ajustada para acoplarse a las necesidades del proceso de extracción del *Data Warehouse*.

Dependiendo de las aptitudes del sistema fuente algunas transformaciones pueden tomar lugar durante este proceso de extracción. El tamaño de la *Data* extraída varía de cientos de *kilobytes* a *gigabytes*, dependiendo del sistema fuente y la situación del negocio. Así mismo dependiendo del sistema fuente el espacio de tiempo puede variar entre días/horas y minutos por acercarse a un tiempo real. Los archivos de transacciones (*log*) por ejemplo pueden fácilmente llegar a tener cientos de *megabytes* en muy corto tiempo. Estas son consideraciones importantes para la extracción y para *ETL* en general.

3.6.1.5. Transporte de datos

El transporte es la operación de mover datos desde un sistema a otro sistema. En un ambiente *Data Warehouse*, los requerimientos más comunes para la transportación están en mover datos desde:

- Una base de datos de limpieza a un *Data Warehouse*.
- Un *Data Warehouse* a un *Data Mart*.

- Un sistema fuente a una base de datos de limpieza o a una base de datos *Data Warehouse*.

El transporte es a menudo una de las partes más simple del proceso *ETL* y puede ser integrada con otras partes del proceso, ya que existe tecnología de consultas que proveen un mecanismo para la extracción, así como para el transporte de datos.

3.6.1.6. Transformación y carga de los datos

Las transformaciones de datos son a menudo lo más complejo y en términos de tiempo de procesamiento, la parte más costosa del proceso *ETL*. Es posible ir de simple conversiones de datos a técnicas complicadas de datos extremadamente complejos. Luego de realizar la extracción y transporte, es necesario manipular o transformar los datos antes de cargar los resultados en el *Data Warehouse*. Tomar los datos desde varias bases de datos operacionales y transformarlos en datos requeridos para el depósito, se refiere a la transformación o a la integración de datos. Las bases de datos operacionales, diseñadas para el soporte de varias aplicaciones de producción, frecuentemente difieren en el formato.

Los mismos elementos de datos, si son usados por aplicaciones diferentes que contribuyen al *DWH* o administrados por diferente *software DBMS*, pueden usar nombres de elementos inconsistentes, que tienen formatos inconsistentes y/o ser codificados de manera diferente. Todas estas inconsistencias deben resolverse antes que los elementos de datos sean almacenados en el *Data Warehouse* y los procesos de transformación de datos se desarrollen para direccionar estas inconsistencias.

La transformación de datos también se encarga de las inconsistencias en el contenido de datos. Una vez que se toma la decisión sobre que reglas de transformación serán establecidas, deben crearse e incluirse las definiciones en las rutinas de transformación. Se requiere una planificación cuidadosa y detallada para transformar datos inconsistentes en conjuntos de datos conciliables y consistentes para cargarlos en el *Data Warehouse*.

3.6.2. Elaborar la Metadata ⁶

La Metadata (es decir, datos acerca de datos) describe los contenidos del *Data Warehouse*. La metadata consiste de definiciones de los elementos de datos en el depósito. Es la información sobre los datos que se alimenta, se transforma y existe en el *Data Warehouse*.

Típicamente, Metadata incluye los siguientes ítems:

- Las estructuras de datos da una visión de los datos al administrador de datos.
- Las especificaciones de transformaciones de datos que ocurren tal como la fuente de datos se replica al *Data Warehouse*.
- El modelo de datos del *Data Warehouse* (es decir, los elementos de datos y sus relaciones).
- Un registro de cuando los nuevos elementos de datos se agregan al *Data Warehouse* y cuando los datos antiguos se eliminan o se resumen.
- Los niveles de sumarización, el método de sumarización y las tablas de registros de su *Data Warehouse*.

Algunas implementaciones de la Metadata también incluyen definiciones de las vistas presentadas a los usuarios del *Data Warehouse*. La Metadata es considerada como el corazón del ambiente *Data Warehousing*. Crear definiciones de metadata completa y efectiva puede ser un proceso que consuma tiempo, pero lo mejor de las definiciones y si se usan herramientas de gestión de *software* integrado, son los esfuerzos que darán como resultado el mantenimiento del *Data Warehouse*.

3.6.3. Dar acceso de usuario final

Los usuarios accesan al *DWH* por medio de herramientas de productividad basadas en *GUI* (*Graphical User Interface*-Interfaz gráfica de usuario). Existen muchos tipos de herramientas que pueden incluir *software* de consultas, generadores de reportes, procesamiento analítico en línea, *Data mining*, etc., dependiendo de los usuarios y requerimientos. Una sola herramienta no satisface todos los requerimientos, por lo que es necesaria la integración de una serie de herramientas.

En la realización de esta tesis se ha utilizado la herramienta *XLCubed* como complemento de Microsoft Excel y *Report Portal* para la presentación dinámica de reportes a los usuarios finales.

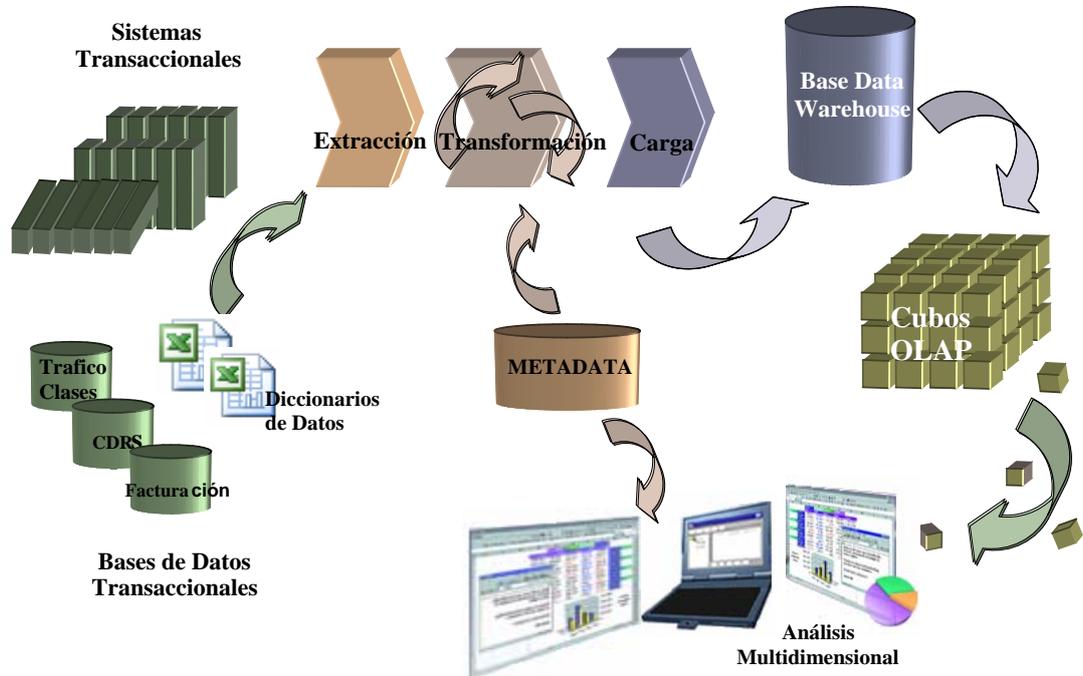


Gráfico 3.7. Esquema Conceptual *Data Warehouse*

3.7. *Data Marts*

Un *Data Mart* es una colección de temas de áreas organizadas para el soporte de decisiones basada en las necesidades de un departamento dado.

En algunas implementaciones *Data Warehouse*, un *Data Mart* es un *Data Warehouse* miniatura; en otros, es sólo un segmento del *Data Warehouse*. Los *Data Marts* a menudo son usados para proveer información a segmentos funcionales de la organización, como departamentos de ventas, embarque, inventario, etc. Por ejemplo una organización de servicios puede tratar los centros de operación regional como unidades individuales de negocios, cada una con su propio *Data Mart* contribuye al *Data Warehouse* maestro.

Los *Data Marts* son a veces diseñados como complemento individual de los *Data Warehouses* y contribuye a toda la organización como un miembro distribuido.

En otros diseños, los *Data Marts* reciben datos de un *Data Warehouse* maestro, a lo largo de actualizaciones periódicas, en cuyo caso la funcionalidad del *Data Mart* está a menudo limitada a la presentación de servicios para clientes.

A pesar de la funcionalidad provista por los *Data Marts*, deben ser diseñados como componentes del *Data Warehouse* maestro así los datos de la organización, formato y esquemas son consistentes en el *Data Warehouse*.

Diseños inconsistentes de tablas, mecanismos de actualización ó jerarquías de dimensión pueden evitar a la *Data* ser reutilizada en el *Data Warehouse* y puede resultar en reportes inconsistentes desde la misma *Data*. Por ejemplo, es poco probable que el resumen de reportes producidos en un *Data Mart* del departamento de finanzas que organiza las ventas llevado por la estructura de administración de reportes esté de acuerdo con los reportes resumen producidos en el *Data Mart* del departamento de ventas que organiza las mismas ventas llevado por la región geográfica. No es necesario imponer una vista de datos en todos los *Data Marts* para lograr consistencia, es usualmente posible diseñar esquemas consistentes y formatos de datos que permitan enriquecer las variedades de vistas de datos sin sacrificar la interoperabilidad. Por ejemplo, el uso de un formato estándar y la organización para el tiempo, cliente y la *Data* de producto no imposibilita a los *Data Marts* a presentar la información en las diversas perspectivas del inventario ventas o análisis financiero.

Los *Data Marts* deberían ser diseñados desde la perspectiva de que son componentes del *Data Warehouse* pese a su funcionalidad individual o construcción. Esto provee consistencia y usabilidad de información a lo largo de la organización. Las herramientas utilizadas para un *Data Mart* pueden incluir algunas de las herramientas usadas para *Data Warehouses* dependiendo de cómo el *Data Mart* ha sido diseñado. Si el *Data Mart* es creado y mantenido localmente y participa en el *Data Warehouse* de la organización como un contribuidor independiente, su creación y mantenimiento involucrará todas las operaciones de un *Data Warehouse*. Si el *Data Mart* es un punto de acceso local para los datos distribuidos desde un *Data Warehouse* central, sólo un subconjunto de las herramientas puede estar involucrado.

3.7.1. Distribución de la *Data* del *Data Warehouse* a los *Data Marts*

Cuando la *Data* es preparada y cargada en el *Data Warehouse* en el sitio central y luego distribuida a los *Data Marts* locales. Las herramientas *DTS* pueden ser usadas para ejecutar las transferencias de datos, incluyendo filtro de datos apropiados para el *Data Mart* y actualizar las tablas apropiadas en el mismo. Los paquetes *DTS* pueden también ser creados para actualizar los cubos *OLAP* en el *Data Mart* después de que la nueva *Data* es recibida desde el *Data Warehouse* central. Algunos escenarios de distribución *Data Warehouse* pueden también usar replicación para coordinar y mantener la *Data* del *Data Mart*.

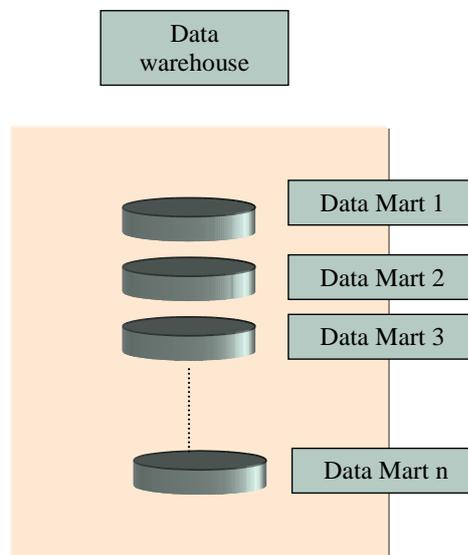


Gráfico 3.8. Distribución de los Datos a los *Data Mart*

3.8 Conclusión

El buen uso de los recursos de información eleva la eficiencia de la gestión administrativa, por esta razón las organizaciones hacen lo posible por conseguir información confiable que les permita tomar decisiones, el *Data Warehouse*, es considerado como una solución porque permite hacer un mejor uso de la información al reunir los elementos de datos apropiados desde diversas fuentes de aplicación en un ambiente integral centralizado, simplificando el problema de acceso a la información y en consecuencia, acelerando el proceso de análisis y consultas.

Los *Data Marts* son extensiones naturales del *Data Warehouse* se implementan primero por cuestiones de costos ya que resulta mucho más barato implementar un *Data Mart* que un *Data Warehouse*.

Es recomendable al realizar el análisis para el *Data Mart*, enfocarse a los procesos del negocio ó al objetivo que desea conseguir la empresa, debido a que existen muchas áreas que comparten los mismos indicadores y se cometería errores como el de duplicar la información al no compartir dimensiones.

REFERENCIAS

^{1,2,4,6} REALITECH. *Data Warehousing*. <http://www.sqlmax.com/dataw1.asp> Junio 2001, [consulta 15 de abril de 2006]

⁵ ORACLE CORPORATION. *Data Warehousing Guide*. <http://www.lc.leidenuniv.nl/awcouse/oracle/server.920/a96520/doc.htm> Marzo 2002, [consulta noviembre de 2005]

³ IBM. Variante en el Tiempo. Ibarra María de los Ángeles. http://www.exa.unne.edu.ar/depar/areas/informatica/sistemasoperativos_olap.pdf . 2005, [consulta 1 de abril de 2006]

CAPÍTULO 4

ANÁLISIS DIMENSIONAL

4.1. Introducción

Este capítulo presenta una definición del Modelo Dimensional, como una alternativa que busca la gente de negocios para conseguir información sumariada (resumida) a tiempo y en diferentes niveles.

Se presenta el cubo dimensional así como el esquema estrella utilizado para la representación de un modelo de más de cuatro dimensiones. Se define el esquema estrella como el modelo más utilizado de datos y se mencionan las consideraciones que se deben tomar en cuenta antes de construir dicho esquema.

Como parte de las dimensiones y los hechos (valores numéricos), se describen los diferentes componentes que los integran: jerarquías, niveles, atributos y datos sumariados, además de las agregaciones como resultado de la relación de dimensiones y hechos.

A lo largo del capítulo se proporcionan soluciones para mejorar el rendimiento en las consultas y opciones para la obtención de resúmenes, en la última parte se podrá encontrar una descripción breve del esquema copo de nieve como otra alternativa para el modelado de datos.

Como parte del análisis se especifica los detalles a tomar en cuenta al realizar la extracción, integración y consideraciones que deben ser revisadas, como introducción a la implementación física, que es redacta en el siguiente capítulo.

Algunos de los términos y definiciones establecidas han sido tomadas del libro “*Designing a Data Warehouse*”, redactado por *Chris Toddman*, quien describe en su libro una metodología utilizada para el diseño de un *Data Warehouse*, basado en ciertos conceptos establecidos por *Ralph Kimball*, el percusor del *DWH*.

4.2. Análisis Dimensional / Multidimensional

Una aproximación al diseño *Data Warehouse* es el desarrollo e implementación de un modelo dimensional. Esto ha dado lugar al análisis dimensional (a veces llamado análisis multidimensional).

El modelo dimensional es la representación de los requerimientos orientados al tema de negocio que desea ver en un reporte en formato entendible para la gente de negocios.

Cuando el *Data Warehouse* comenzó a ser desarrollado, quienes toman decisiones a fin de mejorar el rendimiento en sus empresas, pedían que las respuestas a todas sus preguntas, tuvieran las siguientes características:

- Sumarización de la información con la habilidad de “romper” los resúmenes en más detalles.
- Análisis de la información sumarizada a través de sus componentes organizacionales tales como “departamentos”, “regiones”.
- Habilidad para “dividir y tomar una parte” de la información en la forma en la que ellos elijan.
- Mostrar la información en ambas formas gráfica y tabular.
- Capacidad de ver su información en el tiempo.

El modelo dimensional nos lleva a hacer la siguiente observación:

“Existe generalmente un número de dimensiones diferentes desde las cuales un bloque dado de *Data* puede ser analizado. Esta perspectiva plural, o vista conceptual multidimensional, parece ser la forma en la que la mayoría de gente de negocios naturalmente ve sus empresas.” (*E.F. Coldd*, 1993). ¹

Así el concepto del análisis dimensional llega a ser un método para definir *Data Warehouse*. Se trata de determinar por entrevista a los tomadores de decisiones en una organización, cual es el área de tema en la que ellos están mayormente

interesados y cuales son las dimensiones más importantes del análisis. Hay que recordar que una de las características de un *Data Warehouse* es que está orientada a un tema.

4.3. Cubo multidimensional

En una consulta como por ejemplo el consumo mensual del tráfico saliente por tipo de destino en una central determinada de un *Data Mart* de tráfico telefónico, el área tema es el Tráfico saliente. Las dimensiones del análisis podrían ser el tiempo, el destino y la central. El requerimiento es analizar el consumo para el tráfico saliente por destino y el tráfico saliente mensual.

Este requerimiento es mostrado en el siguiente cubo tri-dimensional.

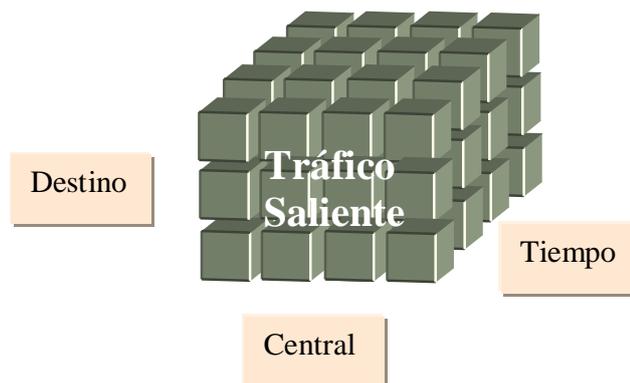


Gráfico 4.1. Cubo tridimensional

Esto significa que el tráfico saliente puede ser analizado por Destino y Central en el tiempo. Así cada elemento del cubo (cada minicubo) contiene un valor de tráfico saliente para un destino en particular, en una central en un punto particular en el tiempo. El cubo multidimensional muestra el tráfico telefónico con tres dimensiones de análisis. No existe un límite real para el número de dimensiones que pueden utilizarse en un modelo dimensional, aunque existe, por supuesto un límite para el número de dimensiones que pueden dibujarse.

Si se requiere para el análisis del tráfico saliente incluir una dimensión más como por ejemplo el cliente entonces las dimensiones de análisis serían: Destino, Central, Cliente y Tiempo.

Como no es posible dibujar modelos de cuatro dimensiones, el modelo dimensional conceptual se representaría como se muestra a continuación:

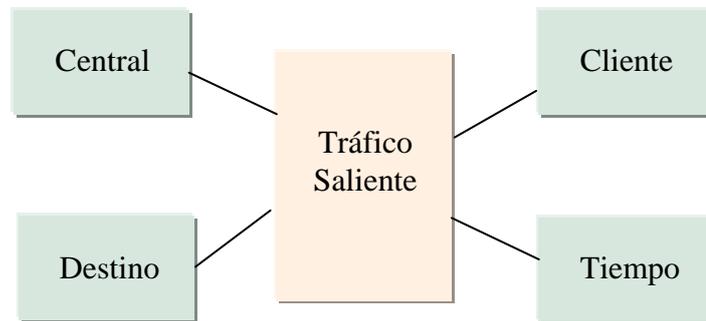


Gráfico 4.2. Modelo Dimensional “Consumo de Tráfico Saliente “

El gráfico anterior es a menudo referido como un Esquema Estrella porque el diagrama muestra la forma de una estrella. El área del tema es el centro de la estrella y las dimensiones del análisis forman los puntos de la estrella. El área tema es a menudo dibujada en una figura grande y delgada debido a que la tabla en sí es generalmente grande y contiene un número pequeño de columnas pero un número muy grande de filas. El Esquema Estrella es el diagrama más comúnmente utilizado para los modelos dimensionales.

4.4. Modelos de datos dimensionales

El modelado dimensional es generalmente una de las mayores luminarias del desarrollo del *Data Warehouse*. Los esquemas estrella y esquemas copo de nieve son ambos ejemplos de un modelo de datos dimensional.

El modelo dimensional es adoptado debido a las siguientes razones:

- Los modelos de datos dimensionales son fácil de entender. Por consiguiente, proveen una introducción ideal al tema.
- No son ambiguos.
- Reflejan la forma en la que en un administrador percibe sus negocios.
- La mayoría de productos *RDBMS* actualmente proveen un soporte directo para los modelos dimensionales.

4.4.1. Esquema Estrella

Un esquema en estrella es aquel que tiene una tabla de hechos que contiene los datos de análisis rodeada de las tablas de dimensiones. Este aspecto, de tabla de hechos (o central) más grande rodeada por tablas más pequeñas es lo que lo asemeja con una estrella.

4.4.1.1. Ventajas

- El esquema estrella es muy fácil de entender, incluso para gerentes de negocios no técnicos.
- El esquema estrella provee una mejor ejecución y tiempos de respuesta más pequeño.
- El esquema estrella es fácilmente extensible y manejará cambios futuros fácilmente.
- Este esquema es ideal por su simplicidad y velocidad para ser usado para el análisis. Permite acceder tanto a datos agregados como de detalle, además de permitir reducir el número de *joins* (uniones) entre tablas y deja a los usuarios establecer jerarquías y niveles entre las dimensiones.
- Este esquema es la opción con mejor rendimiento y velocidad pues permite indexar las dimensiones de forma individualizada sin que repercuta en el rendimiento de la base de datos en su conjunto.

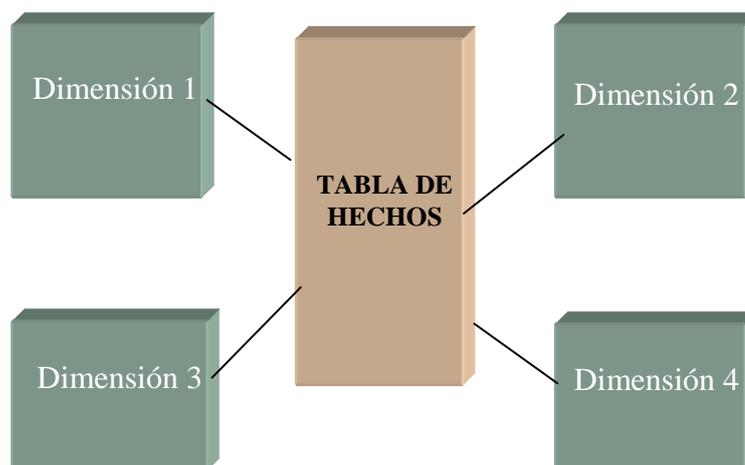


Gráfico 4.3. Esquema Estrella

4.4.1.2. Consideraciones

Es necesario explicar las consideraciones que se deben tener al utilizar un esquema estrella:

- No hay necesidad de registrar las relaciones y sus descripciones. En un esquema estrella existe una relación implícita entre los hechos y cada una de las dimensiones.
- El identificador lógico de los hechos es un identificador compuesto que incluye todos los identificadores de las dimensiones.
- Debe estar siempre presente la dimensión tiempo.

4.4.1.3. Dimensiones

Una dimensión es una jerarquía organizada de categorías, conocida como niveles, que describe la *Data* en las tablas de hecho *Data Warehouse*. Las dimensiones generalmente describen un conjunto similar de miembros en los cuales el usuario desea basar un análisis y éstos son un componente fundamental de cubos. Las dimensiones son un atributo estructural de cubos.

Cuando se crea una dimensión de una tabla, se seleccionan las columnas que la definen. El orden en el cual las columnas son seleccionadas es significativo debido a que este afecta a la ubicación de los miembros dentro de la jerarquía de la dimensión.

La clave primaria de cada tabla de dimensión se une a una llave foránea en la tabla de hecho del cubo. Las columnas clave son requeridas en la definición de la dimensión. Las dimensiones categorizan la *Data* numérica (medidas) en un cubo para el análisis.

Los valores alfanuméricos más pequeños alrededor del cubo son los miembros de las dimensiones. Los valores numéricos dentro del cubo representan la medida. Estos valores existen para todas las celdas en el cubo pero son mostradas sólo para aquellas en el primer plano.

Después de que se ha creado una dimensión es necesario procesarla. Después de que una dimensión es cambiada o su tabla es actualizada, usualmente debe procesar la dimensión. Sin embargo, el procesar la dimensión puede irrumpir el acceso a los usuarios finales a los cubos que incluyen la dimensión. ²

4.4.1.3.1. Estructura de una dimensión

La estructura de una dimensión que se crea determina el tipo de dimensión. Es posible crear una dimensión basada en columnas. Al definir una dimensión es posible:

- Seleccionar una o más columnas desde la fuente. Si selecciona múltiples columnas, todas las columnas deberían interrelacionarse de forma tal que todos sus valores puedan ser organizados en una jerarquía. Para definir la jerarquía, se debe ordenar las columnas desde lo más general a lo más específico. Por ejemplo, una dimensión tiempo es creada de las columnas Año, Trimestre, Mes y Día. Este modelo produce dimensiones regulares.
- Las columnas en una definición de dimensión aportan a los niveles de la dimensión. Los niveles son usualmente ordenados por detalle y organizados en una jerarquía que permite caminos lógicos para ver la información al nivel más bajo de especificación o granularidad (*drill down*). Por ejemplo, una dimensión tiempo puede permitir a los usuarios finales realizar un *drill down* de Año a Trimestre, de trimestre a mes y de mes a día. Cada *drill down* proporciona una mayor detalle.
- Las relaciones entre el número de columnas en una definición de dimensión y el número de niveles en la dimensión dependen de la estructura de la dimensión. En una dimensión regular, cada columna en la definición de la dimensión contribuye a un nivel.
- Cada nivel contiene miembros. Los miembros son valores en las columnas que definen los niveles. Por ejemplo, el nivel Trimestre puede contener cuatro miembros: Trimestre 1, Trimestre 2, trimestre 3 y Trimestre 4. Sin embargo, si la *Data* en la tabla abarca un tiempo de más de un año, el nivel Trimestre contiene más de cuatro miembros. Por ejemplo, si el nivel Año contiene tres

diferentes miembros, 1996,1997 y 1998 el nivel trimestre contiene doce miembros.

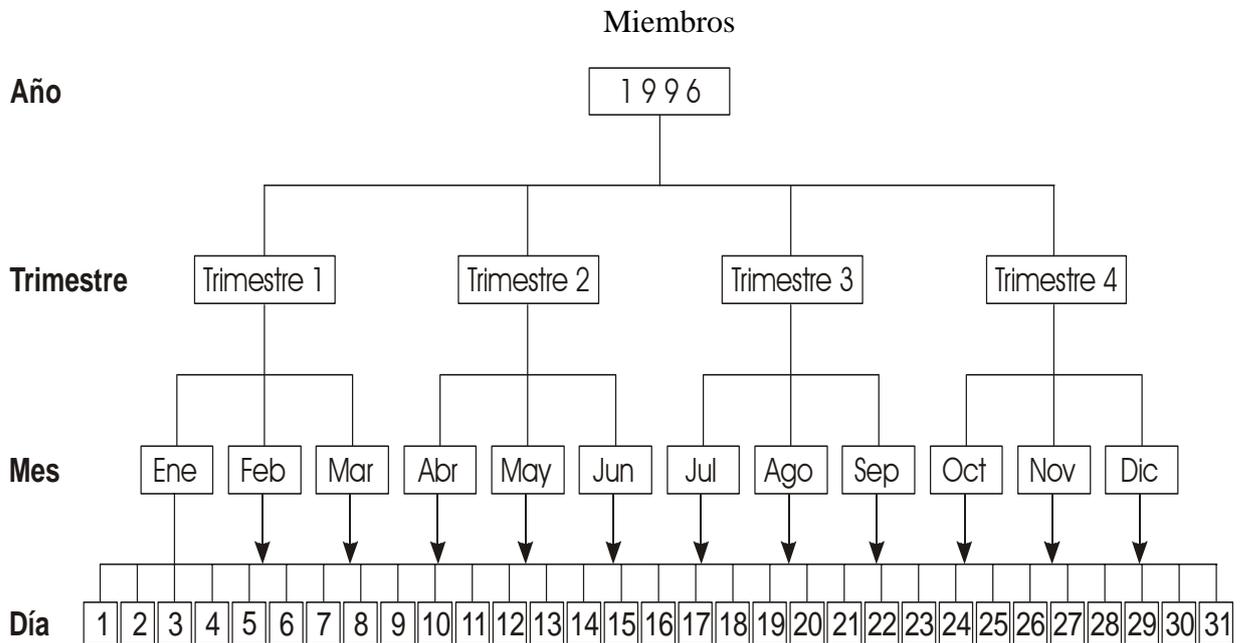


Gráfico 4.4. Relación entre los niveles y miembros en una dimensión ³

www.doc.ddart.net/mssql/sql2000 (consulta en línea)

La relación entre los niveles y miembros en una dimensión tiempo para un año es mostrada en el gráfico 4.4. Las flechas apuntando hacia abajo indican los miembros que no se muestran. Los miembros de día existen para cada mes, pero son mostrados sólo para Enero debido a las limitaciones de espacio.

4.4.1.3.2. Dimensiones privadas y compartidas

Una dimensión puede ser creada para usar un cubo o múltiples cubos. Una dimensión privada es una dimensión creada para un cubo. Una dimensión compartida es una dimensión que puede ser usada por múltiples cubos.

4.4.1.3.3. Jerarquías de dimensión

Una jerarquía es el conjunto de miembros en diferentes niveles en una dimensión. Las jerarquías son representadas como estructuras piramidales. La única excepción se da en las jerarquías en las cuales todos los miembros están en el mismo nivel.

Desde lo alto de la jerarquía piramidal a lo más bajo, los miembros son progresivamente más detallados. Por ejemplo en una dimensión Geografía definida con los niveles Continente, País y Ciudad, en ese orden, el miembro América aparece en el nivel alto de la jerarquía, el miembro Ecuador aparece en la mitad y Cuenca en el nivel bajo. Cuenca es más específica que Ecuador y Ecuador es más específico que América. Mientras más bajo es el nivel de la jerarquía piramidal, generalmente contienen más miembros. En el ejemplo dado existen más países que continentes y más ciudades que países. ⁴

4.4.1.3.4. Niveles y miembros

Un nivel es un elemento de una jerarquía de dimensión. Los niveles describen la jerarquía desde el más alto nivel (más sumariado) al más bajo (más detallado) nivel de granularidad de datos. Los niveles existen sólo dentro de las dimensiones. Están basados en columnas en una tabla de dimensión regular.

Los niveles son definidos dentro de una dimensión para especificar los contenidos y estructura de la jerarquía de la dimensión. Las definiciones de nivel determinan los miembros que son incluidos en la jerarquía. Los niveles son creados cuando se crea una dimensión. Por ejemplo, una dimensión calendario contiene los niveles Año, Trimestre y Mes. Las relaciones entre los niveles y miembros de la dimensión regular calendario, se muestra en la tabla 4.1.

Año	Trimestre	Mes
2006	Trimestre 1	Enero
2006	Trimestre 1	Febrero
2006	Trimestre 1	Marzo
2006	Trimestre 2	Abril
2006	Trimestre 2	Mayo
2006	Trimestre 2	Junio
2006	Trimestre 3	Julio
2006	Trimestre 3	Agosto
2006	Trimestre 3	Septiembre
2006	Trimestre 4	Octubre
2006	Trimestre 4	Noviembre
2006	Trimestre 4	Diciembre

Tabla 4.1. Relaciones entre los niveles y miembros de la dimensión regular ⁵

www.doc.ddart.net/mssql/sql2000 (consulta en línea)

4.4.1.4. Tablas de hecho

Cada *Data Warehouse* o *Data Mart* incluye una o más tablas de hecho, central al esquema estrella, la tabla de hecho captura la *Data* que mide las operaciones de negocio de la organización. Una tabla de hecho puede contener los eventos de una actividad del negocio. Generalmente contienen grandes número de filas, a veces cientos de millones de registros cuando contiene uno o más años de historia para una organización grande.

Una característica clave de una tabla de hecho es que contiene datos numéricos (hechos) que pueden ser sumados para proporcionar información acerca de la historia de la operación de una organización. Cada tabla de hecho también incluye un índice múltiple que contiene llaves foráneas de tablas de dimensión relacionadas, las cuales contienen los atributos de los registros de hecho. Las tablas de hecho no deberían contener información descriptiva o ninguna otra *Data* más que campos de medida numéricos y los campos índice que relaciona los hechos a las correspondientes entradas en las tablas de dimensión.

Las medidas más usadas y que son incluidas en una tabla de hecho son números que son sumables. Las medidas sumables permiten contar con información sumada que puede ser obtenida sumando varias cantidades de la medida.

En un cubo, una medida es un conjunto de valores que están basados en una columna en la tabla de hechos y generalmente son numéricos. Adicionalmente, las medidas son los valores centrales de un cubo que son analizados. Las medidas son la *Data* numérica de principal interés para los usuarios finales. Las medidas que se seleccionan dependen de los tipos de información que los usuarios requieren. Algunas medidas comunes pueden ser volumen de tráfico, minutos, valor promedio, etc. Para cada medida existe una celda que contiene un valor en el cubo. Así que no importa que combinación de miembros sea usada en una consulta, un valor de medida puede ser recuperado. El valor puede ser recuperado desde las agregaciones del cubo. Cada medida es derivada de una columna en una tabla de hecho. En un cubo que tiene sólo una tabla de hecho en su esquema, todas las medidas del cubo deben estar contenidas en ella.

Cada medida especifica una función agregada que determina como los valores en la columna fuente de la medida están agregados. Las funciones de agregación más utilizadas son: Suma, Mínimo, Máximo, Contar y Contar los distintos. Las columnas sumables pueden ser sumadas. Por ejemplo la columna minutos consumidos es sumable. Las columnas sumables son convenientes como medidas en un cubo sin importar la función de agregación que sea usada. Las columnas no sumables no pueden ser sumadas significativamente. Por ejemplo, una columna numérica contiene un identificador tal como el código de la central. Las columnas no sumables también pueden ser usadas como medidas en un cubo, pero para ser significativa deben ser sumadas por la función de agregación Contar (*Count*) ó Contar los distintos (*Count Distinct*) por ejemplo saber cuántas centrales existen para eso debería contar los distintos del código de la central (*cen_cod*).

Una medida puede ser derivada de múltiples columnas combinadas en una expresión. Por ejemplo, la medida promedio consumo por llamada es la división de dos columnas numéricas: consumo y número de llamadas. Los miembros calculados pueden ser usados como medidas. Los valores de miembros calculados son creados con fórmulas cuando el cubo es consultado, pero los valores no están almacenados.⁶

4.4.1.4.1. Presentación de medidas

Las medidas forman la parte principal de la información del cubo que es presentada a los usuarios finales. La presentación puede ser tabular (las medidas son mostradas en filas y columnas) ó gráfica (las medidas son mostradas en una variedad de formas, incluyendo líneas, colores, sombras, etc.) dependiendo de la aplicación con la cual los usuarios finales navegan en el cubo, pero las medidas son la información final en la cual se enfocan los usuarios.⁷

4.4.1.4.2. Agregaciones

Es el proceso de calcular la *Data* sumada desde registros detallados. Se trata de reducir el tamaño de la tabla de hecho mediante la agregación (registros sumados). Sin embargo cuando esto sucede, en la tabla de hecho la información detallada no está directamente disponible para el analista.

Si es necesaria la información detallada, las filas de detalle que fueron sumariadas tendrán que ser identificadas y localizadas posiblemente en el sistema fuente que proporcionó la *Data*. En la tabla de hecho la *Data* debería estar mantenida en el nivel más fino de granularidad posible. La *Data* de agregación en la tabla de hecho debería sólo estar realizada después de considerar las consecuencias.

Mezclar *Data* agregada y detallada en la tabla de hecho puede causar problemas y complicaciones al usar el *Data Warehouse*. Por ejemplo, Una orden de venta a menudo contiene varias líneas de *item* y puede contener un descuento, impuesto o costo de embarque que es aplicado al total de la orden y no a las líneas individuales de *ítems*. Las consultas sumariadas llegan a ser más complejas. Existen dos modelos que pueden ser utilizados en esta situación: Un modelo es asignar los valores de nivel de orden a líneas de ítem basados en valor, cantidad o peso de embarque. Otro modelo es crear dos tablas de hecho, una conteniendo datos en el nivel de línea de *item* y el otro conteniendo la información del nivel orden.

La llave de identificación de orden debería ser llevada a la tabla de hecho de detalle, así las dos tablas pueden estar relacionadas. La tabla de orden puede entonces ser usada como una tabla de dimensión para la tabla detallada, con los valores de nivel orden considerados como atributos del nivel orden en la jerarquía de la dimensión.

4.4.1.4.3. Mejorar los tiempos de respuesta (diseño físico/lógico)

Para completar el diseño del *Data Warehouse* es necesario reconocer el crecimiento de la tabla de hecho en el tiempo. Existe la posibilidad de que una consulta tenga que lidiar con cada una de las filas de la tabla de hecho para responder a una consulta en particular. De hecho es probable que algunas consultas requieran un gran porcentaje de filas, si no toda la tabla, para dar una respuesta al usuario. Esto tomaría mucho tiempo. Algunas consultas son bastante complejas, incluyendo muchos caminos de unión y esto incrementaría seriamente el tiempo tomado para que los resultados sean establecidos. El problema se vuelve realmente grave cuando varias personas están usando el sistema al mismo tiempo, cada uno con una consulta compleja por correr. Como diseñadores debemos tratar de acelerar las cosas para que los tiempos de respuesta mejoren.

Los índices pueden ayudar, pero las consultas requerirán acceder a más de la mitad de la *Data* y los índices son menos eficientes en esos casos que una búsqueda secuencial completa de las tablas. La respuesta estaría en los resúmenes, como se dijo antes las consultas sumarían grandes números de filas juntas y retornarían un conjunto resultado con un número más pequeño de filas. Si pudiéramos predecir en un cierto porcentaje, los tipos de consultas que los usuarios mayormente ejecutarán, se podría preparar algunas tablas de hecho sumariadas para que los usuarios puedan acceder a estas, en caso de satisfacer a los requerimientos de consulta. Donde los agregados no proveen la *Data* requerida, entonces el usuario puede todavía acceder al detalle. Si se cuestiona a los usuarios lo suficientemente cerca sería posible llegar a obtener un conjunto, talvez media docena o más, de tablas de hecho sumariadas. Los principios del esquema estrella se siguen aplicando, pero el resultado es que tenemos varias tablas de hecho en lugar de una sola.

Se debería enfatizar que esta es una consideración única de diseño físico, su único propósito es mejorar la ejecución de las consultas. La mayoría de veces la sumariación se realiza en el tiempo. Existen otras sumariaciones, pero las sumariaciones unidas al tiempo son la práctica más común en los *Data Warehouses*.

Una técnica muy usada por las personas que utilizan el *Data Warehouse*, es la habilidad de realizar operaciones como *drill down* (expande) y *drill up* (comprime), para llegar de un nivel de sumariación a un nivel de detalle más bajo. Estas habilidades son capacidades de reportes poderosos proporcionados por un *Data Warehouse* donde la sumariación es utilizada. El uso del *Data Warehouse* debe ser monitoreado para asegurar que los resúmenes están siendo usados por consultas que están realizándose en la base de datos. Si se encuentra que éstas no están siendo usadas deberían ser borradas y reemplazadas por otras que son de más uso.

4.4.2 Esquema copo de nieve

En el modelo Copo de Nieve (*Snowflake*) las tablas de dimensión están normalizadas con respecto al esquema estrella. Es una variante al esquema estrella en el cual las tablas de dimensión están normalizadas, es decir, pueden incluir claves que apuntan a otras tablas de dimensión.

En el gráfico 4.5 es posible ver un esquema copo de nieve donde una dimensión se une a la tabla de hecho a través de otra tabla de dimensión.

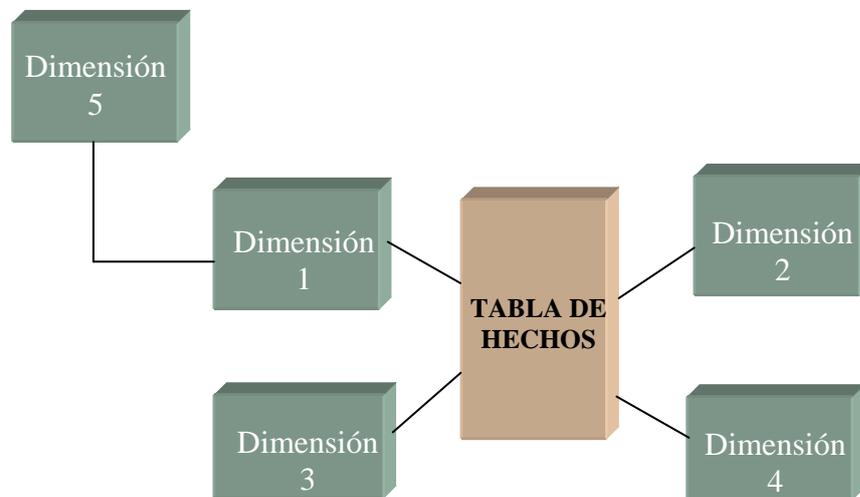


Gráfico 4.5. Esquema Copo de Nieve

Las ventajas de esta normalización son la reducción del tamaño y redundancia en las tablas de dimensión y un aumento de flexibilidad en la definición de dimensiones. Sin embargo, el incremento en la cantidad de tablas hace que se necesiten más operaciones *JOINS* (uniones) para responder a las consultas, lo que empeora la ejecución. Otra desventaja es el mantenimiento que requieren las tablas adicionales.

4.5. Análisis para la extracción

Para realizar la extracción se debe tener en cuenta las consultas que se quieren obtener. El primer problema es extraer información desde los sistemas fuentes así los datos pueden ser ingresados al *Warehouse*. Se debe tener un mecanismo para identificar los datos y conocer su semántica. Habiendo identificado la *data*, ésta debe ser capturada y almacenada de tal forma que pueda ser establecida dentro del *Warehouse* en el tiempo apropiado.

Se debe tener cuidado en asegurar lo que significa un término de información. Si un determinado término tiene varias transiciones y en cada una de ellas es visto de una forma diferente, es necesario identificar claramente el momento en que los datos estarán listos para el ingreso a la bodega de datos.

Los altos funcionarios decidirán cuando sucederá esto. Lo que significa que siempre se identificará primero los datos y su cambio de estado antes de su consecuente registro en el *Data Warehouse*. La única forma de lograr esto es hacer algún cambio en la aplicación, así el cambio de estado es reconocido y capturado.

La mayoría de organizaciones no están preparadas para considerar el re-desarrollo de sus sistemas operacionales para introducir un *Data Warehouse*. En lugar de ello, el *software* de aplicación tiene que ser modificado para capturar los datos requeridos en almacenes temporales de forma que ellos puedan ser ubicados dentro del *Data Warehouse* subsecuentemente.

En el caso de las aplicaciones basadas en lenguajes de tercera generación los cambios son implementados alterando los programas para que así ellos creen almacenes de datos temporales como archivos secuenciales, por ejemplo: los sistemas de administración de bases de datos relacionales a menudo proveen la capacidad para “*triggers*” para ser almacenados en la base de datos. Estos *triggers* son sentencias *SQL* de tablas relacionadas que son ejecutadas (a veces expresadas como disparo, de allí el nombre *trigger*) cada vez que una fila es insertada, actualizada o borrada. Podemos crear *triggers* para examinar el estado de una orden cada vez que la orden es actualizada. Tan pronto como el estado indica que una venta ha ocurrido, el *trigger* lógico puede registrarse en una tabla temporal para la subsiguiente colección dentro del *Warehouse*. El uso de aplicaciones empaquetadas a veces requiere que los desarrolladores originales lleguen a estar involucrados en hacer los cambios requeridos para que los datos estén disponibles para el *Data Warehouse*. Es también necesario extraer la información relacionada a las dimensiones del análisis, los puntos de la estrella y la tabla de hechos en el esquema. La información del cliente está sostenida en una base de datos relacional. Así obtener un archivo inicial de clientes sería algo útil.

Las tablas en cuestión deberían ser copiadas a un archivo usando el programa de utilidad de exportación del *RDBMS*. Obtener detalles de nuevos clientes y cambios para clientes existentes es ligeramente más delicado pero es probablemente mejor al ejecutar por el uso de *triggers* de la base de datos como se describió antes.

Cada vez que los detalles de los clientes son actualizados, el *trigger* de actualización es disparado y podemos registrar el cambio (a veces llamado el delta) en un archivo secuencial para la transferencia subsiguiente dentro del *Warehouse*.

Puede pensarse que existe una tabla de clientes perfectamente utilizable ya establecida en la base de datos. Frases como duplicación innecesaria e incontrolada de datos, esto es, redundancia de datos, pueden venir a la mente. Estas son preocupaciones perfectamente correctas y razonables.

Usualmente, algunos arreglos pueden ser realizados, por medio de lo cual la compañía de *software* altera su producto o nos provee la suficiente información, así podemos escribir algunos programas adicionales para capturar los datos que necesitamos.

4.6. Análisis para la integración

Existen, algunos aspectos importantes de la integración:

1. Formato de Integración.
2. Semántica de Integración.

4.6.1. Formato de integración

El problema del formato de integración se refiere principalmente al aseguramiento de que el dominio de integración sea restaurado donde éste ha sido perdido. En la mayoría de organizaciones, existen usualmente algunos casos donde los atributos en un sistema tienen diferentes formatos para el mismo ó atributos similares en otros sistemas. Por ejemplo:

- Números de cuentas de bancos o números telefónicos pueden ser almacenados como tipo “*string*” en un sistema y como tipo “numérico” en otros.

- En diferentes sistemas, son usados diferentes tamaños para valores de cadenas tales como nombres, direcciones, y descripciones de productos, etc.
- Fechas, pueden mantenerse en algunos formatos incluyendo “ddmmaa”, ”aammdd”, ”aaaammdd”, ”ddmesaa”, ”ddmesaaaa”. Algunos sistemas almacenan diferentes tipos de dato de fecha.
- Los atributos monetarios son también un problema. Algunos sistemas almacenan dinero como valores enteros y esperan a que la aplicación inserte los puntos decimales. Otros han incluido puntos decimales.

Los formatos de integración no coincidentes son muy comunes. Esto es verdad especialmente para la data extraída cuando:

1. El sistema operativo es diferente.
2. La aplicación de *software* es diferente.

El procedimiento de integración consiste de una serie de reglas que están diseñadas para asegurar que los datos que están cargados en un *Data Warehouse* estén estandarizados. Así todas las fechas son del mismo formato, los valores monetarios están siempre representados de la misma manera, etc.

Esto es importante porque se trata de usar el *Data Warehouse* por ejemplo si se desea una lista de todos los empleados agrupados por sexo, edad y salario promedio, donde ninguno de esos atributos fueron apropiadamente estandarizados, sería una tarea muy difícil de emprender para un programador experimentado. Esto sería imposible para una persona no técnica.

También, un *Data Warehouse* acepta información de una variedad de fuentes en una tabla simple. Esto es factible sólo si el dato es de un tipo consistente y tiene un formato consistente.

La regla de integración establecida es usada como un tipo de mapa que especifica como la información que ha sido extraída de los sistemas fuentes tiene que ser convertida antes de que ésta sea permitida dentro del *Data Warehouse*.

4.6.2. Semántica de integración

La semántica se refiere al significado de los datos. Típicamente para el *Data Warehouse* se extrae información de diferentes sistemas operacionales que son usados por diferentes personas en una organización, esto es lógico, los sistemas financieros es más probable que sea usado por los departamentos de contabilidad, mientras que, los sistemas de control de stock es usado por el personal de bodega. Puede ser muy confuso para un analista de base de datos tratar de entender los tipos de información que esta siendo sostenida en los sistemas. El problema es combinado, porque, a menudo, los usuarios de los sistemas y la información están desprevenidos del problema.

En las conversaciones diarias, se discute sobre diferentes puntos que pueden no ser obvios para ellos y la mayoría del tiempo no existen repercusiones serias.

Al construir un *Data Warehouse* no es posible darse el lujo de ignorar estas diferencias generalmente sustanciales en la semántica porque la información producida por las consultas que realiza el *Data Warehouse* será usada para tomar decisiones de soporte en los niveles más altos de la organización.

Es vital, por consiguiente, que cada *item* de datos que es insertado en el *Data Warehouse* tenga un significado preciso que sea entendido por todos. Con este fin, un *Data Warehouse* debe proveer un catálogo de información que describa precisamente cada atributo en el *Warehouse*. El catálogo es parte del *Warehouse* y es, de hecho, un repositorio de datos que describe otros datos. Estos “datos acerca de datos” son usualmente llamado metadatos.⁵

4.7. Conclusión

Después de explorar los diferentes modelos de datos y de conocer cada uno de los componentes de dimensiones (jerarquías, niveles, atributos) y hechos (valores numéricos) estaremos en capacidad de tomar una decisión acerca de cual modelo es el más adecuado para cada empresa donde se implemente el *Data Warehouse*, así como conocer los conceptos básicos para iniciar el desarrollo de los cubos.

Se tomarán en cuenta aspectos como espacio en disco y la velocidad con la que se ejecutan las consultas para elegir entre los dos modelos de datos mencionados. Para la realización de esta tesis, el modelo de datos que se ha utilizado es el esquema estrella por las ventajas que presenta, aunque el esquema copo de nieve es una alternativa para ahorrar espacio de almacenamiento, se ha seleccionado el primero por la rapidez en la ejecución de las consultas, debido a que en los cubos de tráfico telefónico se habla de millones de registros por lo que una baja ejecución de las consultas no resultaría beneficioso para los usuarios, quienes buscan información estandarizada, resumida e inmediata para brindar un mejor servicio a los abonados. Sin embargo no se tomará en cuenta sólo este aspecto, debido a que existen otros factores que permitirán que el *Data Warehouse* cumpla con su objetivo.

REFERENCIAS

¹ TODDMAN, Chris, “*Designing a Data Warehouse*”, Hall PTR, New Jersey, Octubre 2001, pp 34-35.

² MICROSOFT CORPORATION. *Dimensions Hierarchies*.
http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions_42ib.htm, 1988, 2000, [consulta 20 de abril de 2006]

³ MICROSOFT CORPORATION. *Introduction To Dimensions*.
http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions_30mr.htm, 1988, 2000, [consulta 21 de abril de 2006]

⁴ MICROSOFT CORPORATION. *Levels and members*.
http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/aglevels_7w4z.htm, 1988, 2000, [consulta 21 de abril de 2006]

⁵ MICROSOFT CORPORATION. *Dimensions*.
http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions_12pf.htm, 1988, 2000, [consulta 20 de abril de 2006]

⁶ MICROSOFT CORPORATION, “*Database creation, warehousing, and optimization*”, Microsoft Press, United States of America, 2001. pp. 275-277.

⁷ MICROSOFT CORPORATION. *Measures*.
http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agmeasures_9fsj.htm, 1988, 2000, [consulta 20 de abril de 2006]

⁵ TODDMAN, Chris, “*Designing a Data Warehouse*”, Hall PTR, New Jersey, Octubre 2001, pp 39-44.

CAPÍTULO 5

TRILOGÍA - MODELO CONCEPTUAL/ LÓGICO /FÍSICO

5.1. Introducción

En este capítulo se presenta el modelo conceptual, lógico y físico como parte fundamental del desarrollo *Data Warehouse*, dando a conocer el análisis que se debe realizar para desarrollar el modelo conceptual general. En un primer plano se presenta el modelo donde se estudiará las circunstancias y comportamientos de los datos, en caso de considerarse necesario el almacenamiento de históricos se debe hacer un análisis de:

- Atributos cambiantes o no como parte de las circunstancias y
- La relación del cliente con la empresa como parte del comportamiento.

Se presenta en gran detalle el *Dot Modeling* (Modelamiento de punto) como un modelo simple de captura de requerimientos de información y como un método para permitir a los gerentes definir la información requerida y a los desarrolladores basarse en dichos requerimientos para desarrollar posteriormente el modelo lógico.

Como un complemento se presenta las hojas de trabajo, así como la metadata como parte del modelo.

Se destaca el desarrollo de aplicaciones colectivas *JAD*, con sus dos talleres de trabajo como ayuda para obtener los requerimientos de usuario a través del esfuerzo colectivo de desarrolladores y gente de negocios.

El esquema lógico es incluido como una ayuda para clarificar el modelo conceptual y no tiene la intención de prescribir en un modelo físico.

En la implementación física se presentará la importancia de la validación de datos al ser extraídos, así como su integración y mapeo.

5.2. Modelo Conceptual, Lógico y Físico

Una de las fuerzas de la trilogía es que las decisiones relacionadas al diseño lógico y físico de la base de datos podrían ser tomadas e implementadas sin afectar el modelo abstracto que reflejaba los requerimientos del negocio, modelo conceptual.

En las bases de datos relacionales el modelo conceptual, lógico y físico son tratados de la misma forma. Esto significa que algunos cambios que sean hechos para el diseño, por razones de mejoramiento de ejecución son reflejados en el modelo conceptual así como en el modelo físico. Así se puede llegar a la conclusión de que los requerimientos de negocios están siendo cambiados por razones de ejecución.

En circunstancias normales, en las bases de datos *OLTP* por ejemplo, podemos debatir en los pros y contras de este modelo debido a que los usuarios de negocios no tienen cerca el modelo de datos de interés para ellos.

Los *Data Warehouses* son diferentes. No puede haber debate, los usuarios absolutamente tienen que entender los datos en el *Data Warehouse* ó al menos tienen que entender la parte que ellos utilizan. Por esta razón, el modelo de datos conceptual debe ser reintroducido como una parte necesaria del ciclo de vida del desarrollo de *Data Warehouse*.

Existe otra razón por la que se necesita reinventar el modelo de datos conceptual para el desarrollo de los *Data Warehouses*. Muchos productos *OLAP* son no relacionales, y sus manifestaciones lógicas y físicas son completamente diferentes del modelo relacional.

5.3. Modelo Conceptual

Es necesario conocer las circunstancias y comportamientos de las partes que integrarán el *Data Warehouse*. La mayoría del tiempo se analiza el comportamiento. La tabla de hecho en un esquema dimensional tradicional usualmente contiene una interacción de cualquier dimensión con el negocio. Esto es la forma como se comporta en torno a dicha organización. Por ejemplo si se realiza el análisis de una

dimensión cliente, se podría considerar que pueden existir diferentes circunstancias en torno a éste que podría afectar al negocio como:

- Que el cliente se cambie su dirección ó se cambie a una área diferente.
- Puede ser que el cliente haya cambiado su estado civil.
- Puede ser que la economía del cliente altere la relación con la empresa.

Si el análisis se realiza para un *datamart* de tráfico telefónico, el tipo de información que nos mostraría el comportamiento del cliente es:

- Tipos de llamadas realizadas (locales, larga distancia, etc.).
- Duración de las llamadas.
- Valor cargado a la llamada.
- Hora de la llamada.
- Destinos de las llamadas.
- Tipos de conexiones utilizadas al realizar las llamadas.
- Cambio de Categoría.

Esta información nos presentará los mayores indicadores en diferentes áreas. Aunque un inesperado comportamiento del cliente es el efecto de cambio en alguna circunstancia.

Los principios de causa-efecto pueden ser utilizados para ciertos problemas que se puedan detectar con dichos indicadores. El cambio en las circunstancias determinaría la causa de ciertos problemas. Si analizamos el comportamiento, esto simplemente nos dirá algo que ya sabemos y que es obvio.

Se necesita poner más atención en ser más rigurosos al rastrear los cambios en las circunstancias, en lugar del comportamiento aunque esto no implica dejar la información de comportamiento a un lado. Esto es justo lo que necesitamos para cambiar el énfasis para algunos tipos de información que son absolutamente críticas para el éxito.

Todos los esquemas dimensionales por naturaleza están orientados al comportamiento, pero para desarrollar un modelo orientado a las circunstancias se debe tener una aproximación diferente.

Si se va a construir un modelo tal, entonces se necesita conocer primero el mayor tipo de información de comportamiento y circunstancias. En el ejemplo del cliente y enfocándose a las circunstancias, es necesario primero crear una dimensión cliente.

La dimensión cliente tendría algún tipo de identificador de cliente y un conjunto de atributos. Pero a más de esto se debe poder analizar, medir y clasificar el efecto de cambios en las circunstancias de los clientes.

Esto implica que se debe enfocar en un cambio de circunstancias que simplemente significa un **cambio en el valor de algún atributo**. Sin embargo, no todos los atributos son objeto de cambio como parte del curso ordinario del negocio. Algunos atributos cambian y otros no. Incluso pueden existir atributos que cambien, pero que no sea necesariamente significativo el cambio ó de real interés para el negocio. Es importante distinguir entre los dos tipos de cambio que deben ser de real interés, como:

- Cambios donde necesitamos estar en capacidad de ver los valores anteriores del atributo, así como el nuevo valor.
- Cambios donde los valores previos del atributo pueden perderse.

Lo que se tiene que hacer es agrupar a los atributos en estos dos tipos. Así tendremos un modelo con dos entidades como en el gráfico 5.1.

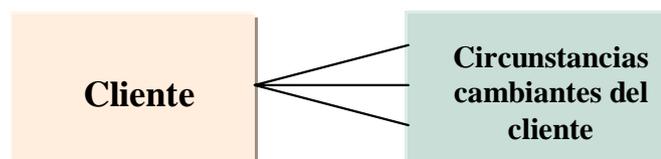


Gráfico 5.1. Modelo general para un cliente con circunstancias cambiantes

De esta forma se empieza con la construcción del modelo conceptual general (*GCM*).

Para cada cliente, tenemos un conjunto de atributos que pueden cambiar así como un conjunto de atributos que no cambian y si lo hacen, no es necesario conocer los valores previos.

La cardinalidad es de uno a muchos. Esto no indica que existen muchos atributos que puedan cambiar, sino que cada atributo puede cambiar muchas veces. Por ejemplo la dirección de un cliente puede cambiar muy frecuentemente en el tiempo. La fecha de nacimiento y el sexo son atributos que se consideran mayormente como no cambiantes ó por lo menos que sus valores anteriores se pueden perder. La dirección y el estado civil por ejemplo, son atributos que requerirían ser conservados.

El modelo conceptual general que se intenta desarrollar debe incluir información de comportamiento. En el ejemplo anterior, la información del comportamiento viene de la interacción del cliente con la organización la misma que muestra que existen muchas instancias de comportamiento en el tiempo.

Para poder obtener las respuestas que queremos es posible usar algunos elementos dimensionales, es decir, por ejemplo juntar nuestra dimensión de cliente con otras que estén en esquemas estrella separados.

Hay que recordar que el propósito de una dimensión, principalmente, es restringir consultas contra la tabla de hecho. El principal propósito de una dimensión común es proveer una facilidad de *drill across* de una tabla de hecho a otra. Aún así siguen siendo modelos basados en comportamiento que son diferentes a un modelo diseñado basado en las circunstancias.

Un modelo conceptual puede constar de tres segmentos. Los dos primeros están basados en las circunstancias del cliente y el comportamiento. El tercer tipo de segmento es conocido como segmento derivado que podría ser el valor de tiempo de vida, esto es el ejemplo de clasificar a los clientes. La inclusión ó clasificación de un cliente en estos segmentos está determinada por algún proceso de cálculo tal como un modelo predictivo. Normalmente no se asignaría un cliente a una clasificación en un segmento derivado sin antes realizar un análisis del valor de algún atributo. Se debería tratar de tener segmentos derivados como parte del modelo conceptual.

5.3.1. Definición

El modelo conceptual permite definir a través de una metodología los requerimientos de usuarios. El modelo *GCM* nos proporciona una plantilla desde la cual todos los modelos conceptuales en el futuro pueden ser derivados.

5.3.2. Requerimientos

Para poder realizar el modelo conceptual es necesario proveer los siguientes requerimientos:

- El modelo debe ser simple de entender y ser usado por personas no técnicas.
- Soportar el modelo general conceptual y soportar el tiempo.

Los modelos más ampliamente usados son el modelo entidad relación. Estos modelos son ampliamente entendidos por profesionales *IT* pero no son fáciles de entender por personas no que no conocen sobre dichos modelos, además tienen reglas de cardinalidad, condiciones de participación, relaciones de inclusión y exclusión, relaciones n-arias y subtipos de entidades. Estas reglas de sintaxis son necesarias para modelos de sistemas operacionales. En muchos aspectos, los requerimientos de diagrama para *Data Warehouses* dimensionales son más simples que los modelos *ER* tradicionales. Las reglas para los modelos dimensionales, a usar para modelar el comportamiento de un cliente son:

- La estructura de un modelo dimensional es predecible. Existe una tabla de hecho simple en el centro del modelo. La tabla de hecho tiene una o más tablas de dimensión relacionadas a ella. Cada dimensión tiene cero, una ó más tablas jerárquicas relacionadas a ella.
- No se requiere que las filas de la tabla de hecho tengan un único identificador.
- Superentidades o subentidades no son tan importantes en los modelos dimensionales.

- Las relaciones generalmente no son complejas. Las relaciones son siempre de “uno a muchos” en una configuración donde la dimensión está en el lado “uno” de la relación y la tabla de hecho en el lado “muchos”. Las relaciones “uno a uno” y “muchos a muchos” son raras, aunque esta cambia cuando el tiempo es introducido. No existe una necesidad real para modelar la cardinalidad (grado) de las relaciones.
- Las condiciones de participación no necesitan ser especificadas. La dimensión en el lado uno de la relación siempre tiene una participación opcional. La condición de participación para la dimensión ó hecho, en el lado “muchos” es siempre obligatoria.
- No existe requerimiento para nombrar ó describir las relaciones mientras su significado es implícito. Es importante mostrar como las jerarquías dimensionales están estructuradas, pero esta es la única información que se necesita para describir las relaciones.

Los *Data Warehouses* no están diseñados para aplicaciones operacionales, estos son diseñados para ayudar a la gente de negocios a tomar decisiones. Es importante, sin embargo, que el *Data Warehouse* contenga la información correcta. A menudo, la gente de negocios está incapacitada para expresar claramente sus requerimientos en términos de información. Este es a menudo el caso en que ellos sienten que tienen problemas pero no están seguros de donde es que recae dicho problema. La mayoría de administradores de negocios tienen una serie de objetivos de negocio. Un *Data Warehouse* puede ser diseñado para ayudarles a lograr sus objetivos de negocio, si ellos están en capacidad de expresarles claramente y describir el tipo de información que necesitan para ayudar a quienes toman las decisiones, en seguimiento de sus objetivos.

Dot Modeling es un método que lleva a los gerentes a través de un proceso de definición de objetivos y requerimientos de información. Lo que se necesita es una abstracción que permita a los requerimientos del negocio enfocarse en un modelo participativo. La gente de negocios debe estar en la capacidad de construir, validar, modificar ó incluso reemplazar el modelo por sí solos. Sin embargo el modelo debe ser lo suficientemente poderoso como para permitir a los diseñadores de *Data Warehouse* continuar luego con el desarrollo del modelo lógico.

Posteriormente se presentará la notación del *dot modeling* como un modelo de captura de requerimientos de información en una forma en la que la gente de negocios puede entender. Existe un requerimiento fundamental: la gente que usa el *Data Warehouse* debe entender como está estructurado.

El modelo conceptual debería ser fácil de entender por parte de gente no técnica para que con un entrenamiento muy pequeño, tales personas puedan producir sus propios modelos. Si es aceptado que el modelo dimensional, debido a su simplicidad, sea un método apropiado para describir la información que va a contener un *Data Warehouse*, entonces debería ser sensible para asegurar que la simplicidad sea mantenida y que el modelo no añada complejidad.

Otro requerimiento del modelo conceptual es que debe retener su forma dimensional logrando un modelo claramente entendible al leer. Trataríamos de asegurar que la forma radial esencial sea retenida incluso en ejemplos relativamente complejos. También existe una necesidad dentro del modelo de registrar la información semántica acerca de cada atributo.

5.3.2.1. Tratamiento de las circunstancias y la retrospección

En reconocimiento de que se necesita establecer más énfasis en el tratamiento de la historia dentro del Modelo Conceptual General, tenemos que examinar el modelo en detalle para evaluar como cada uno de los elementos deberían ser clasificados. Cada componente del modelo que es sujeto de cambio será evaluado para determinar la forma en la cual los valores históricos deberían ser tratados. Un componente significa:

Entidad: Un conjunto de circunstancias o una dimensión.

Relación: Por ejemplo, una jerarquía.

Atributo: Por ejemplo, dirección del cliente.

Para cada componente será dada una clasificación. Esta es llamada retrospección del componente. La retrospección tiene tres posibles valores: Verdadero, Falso y Permanente

La retrosección verdadera significa que el objeto reflejará el pasado fielmente. Esto permite a las consultas devolver subconjuntos de los datos reflejados en los valores históricos.

Cada dimensión, relación y valor de atributo tendrá en efecto un espacio de vida que describe la existencia del objeto. Un objeto puede tener un espacio de vida discontinuo, esto es, algunos períodos de actividad, con períodos de inactividad. La retrosección verdadera es la descripción más precisa de la vida del objeto *DWH*.

La retrosección falsa indica que la vista de la historia será alterada cuando el valor del objeto cambie. En términos simples, cuando los cambios ocurran, los valores anteriores serán sobrescritos y por consiguiente, perdidos. Se podría pensar que los valores antiguos nunca han existido.

La retrosección permanente significa que el valor del objeto no cambiará en el tiempo.

El estado de retrosección verdadera dirigirá a los diseñadores de la base de datos lógica a proveer algún tratamiento especial, con respecto al tiempo, al objeto afectado.

El tipo de tratamiento variará en cada caso y dependerá del tipo de sistema de administración de base de datos que sea implementado. La inclusión de un atributo existente es también implícita para entidades, pero no para relaciones y atributos, donde el valor de retrosección es falso. La provisión del tiempo, cuando la retrosección es falsa, es usualmente más simple de implementar que cuando la retrosección es verdadera. Para relaciones y atributos, es simplemente un caso de reemplazo de valores anteriores con un valor nuevo, en otras palabras, una simple actualización.

5.3.3. *Dot Modeling*

Esta es una metodología para el desarrollo de modelos conceptuales para *Data Warehouses*. La metodología es llamada *dot modeling* o modelado punto. Esta

metodología está basada en los requerimientos para los modelos dimensionales. Es una metodología que permite a la gente no técnica construir su propio modelo conceptual que refleja la percepción personal de su organización en términos dimensionales. Esto también provee una forma estructurada para construir un modelo lógico desde el conceptual.

El modelo ha recibido comentarios positivos de gente no técnica en ambientes donde ha sido implementado. *Dot* no es un acrónimo viene de la característica de que en el centro de la parte de comportamiento del modelo, los hechos, están representados por un punto.

El método fue desarrollado como un tipo de evolución usando conceptos dimensionales y ha evolucionado para adaptarse a los requerimientos del Modelo Conceptual General. Se inicia modelando el comportamiento. El gráfico 5.2. representa el diseño de un reporte tabular de dos dimensiones. Este tipo de reporte es familiar para todos y es una forma común de mostrar la información.

Reporte de Tráfico Saliente

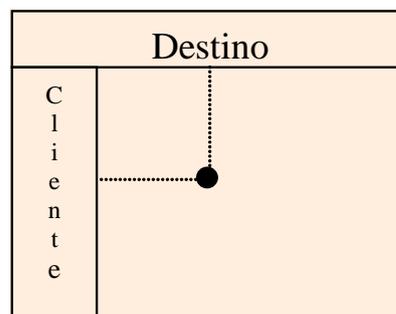


Gráfico 5.2. Ejemplo de un reporte de dos dimensiones

La intersección de los ejes en este ejemplo, mostrada como un punto, representaría la información acerca del tráfico de un cliente en un destino particular. La información representada por el punto es generalmente numérica. Podría ser un valor atómico (no derivado), tal como llamadas realizadas ó podría ser complejo e incluir otros valores como la duración promedio de la llamada.

Si se requiere incluir una dimensión adicional tal como el tiempo en el reporte, entonces el punto representaría una llamada particular de un cliente a un destino en un tiempo particular. La información contenida en el punto sigue siendo la misma de antes. Lo que ha cambiado es que existen más dimensiones por las cuales la información representada por el punto puede ser analizada. Esto se refleja en el gráfico 5.3.

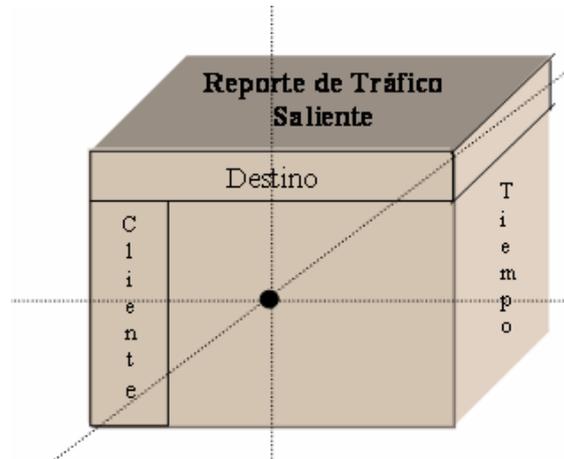


Gráfico 5.3. Reporte tráfico saliente con tres dimensiones

El punto continuará representando la misma información. Sin embargo no es posible representar más de tres dimensiones con este diagrama. En efecto, el punto está atrapado dentro del diagrama tridimensional. Para permitir dimensiones adicionales de análisis a ser representadas en el diagrama, el punto debe ser removido a un diferente tipo de estructura donde tales restricciones no se apliquen. Esta es la razón detrás del desarrollo de la metodología de modelado punto. En este modelo el punto está ubicado en el centro del diagrama y las dimensiones son colocadas alrededor.

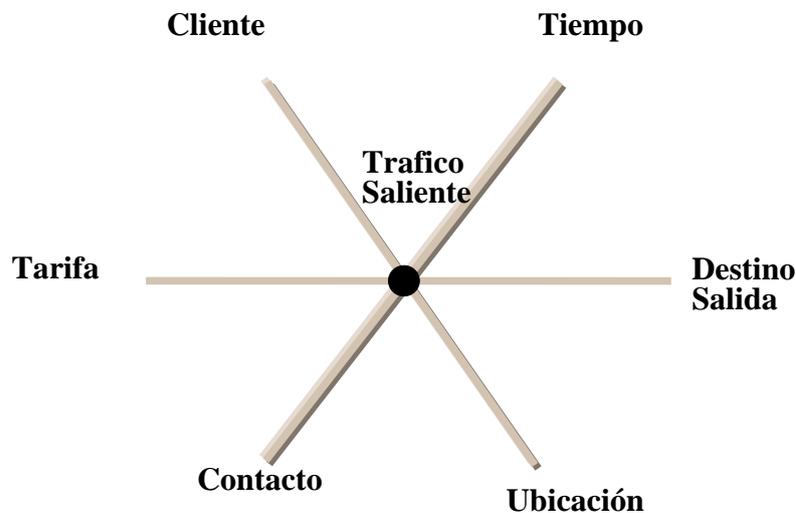


Gráfico 5.4. Modelo Multidimensional Punto Sencillo

5.3.3.1. Componentes de un Modelo Punto de Comportamiento

Existen tres componentes básicos para un diagrama de modelo punto:

Punto: representa los hechos. El nombre del área tema del modelo dimensional es aplicado a los hechos. En el ejemplo anterior los hechos están representados por el Tráfico Saliente.

Nombres de dimensión: cada una de las dimensiones es mostrada en el modelo y contiene un nombre.

Conectores: están ubicados entre los hechos y las dimensiones para mostrar las dimensiones en primer nivel. De forma similar, los conectores están ubicados entre dimensiones y grupos para mostrar la estructura jerárquica.

Debe establecerse énfasis en la simplicidad ya que no existen reglas de notación para el diagrama principal. Es general la ubicación del punto cerca del centro del diagrama y las dimensiones radiando del punto. Esto muestra una forma dimensional legible.

Los atributos para los hechos y dimensiones no son mostrados en el diagrama. Los atributos son descritos en hojas de trabajo de soporte. De forma similar, los requerimientos temporales son representados en hojas de trabajo en lugar del diagrama. El método utiliza un conjunto de hojas de trabajo.

Algunas hojas de trabajo son completadas durante la etapa del diseño conceptual del desarrollo y algunas son completadas durante la etapa del diseño lógico. La primera hoja de trabajo es la hoja de trabajo del modelo de datos en sí. Esta contiene lo siguiente:

- Nombre de la aplicación ó modelo (por ejemplo, Tráfico Telefónico).
- El diagrama del comportamiento.
- Una lista de los atributos del hecho (por ejemplo, llamadas, minutos consumidos, etc.).

Para cada atributo de hecho se describe información que es registrada, bajo lo que comúnmente se conoce como “metadata”. Su propósito es documentar la definición de negocios del atributo. Esto es para resolver los problemas de las diferentes personas, dentro de la organización, teniendo diferentes vistas acerca de las semánticas de atributos particulares. Las descripciones deberían ser expresadas en términos de negocios.

Una segunda hoja de trabajo, la hoja de trabajo de entidades, es usada para registrar lo siguiente:

- Dimensiones de comportamiento.
- Circunstancias.
- Segmentos Derivados.

Esta parte del método contiene algo de la más compleja información en el modelo. El nombre del modelo en cada página asegura que las partes del documento no se mezclen de forma equivocada con otros documentos del modelo. El propósito de la hoja de trabajo de entidades es ayudar a los diseñadores del sistema a entender los requerimientos para ayudarles en el diseño lógico.

Para cada entidad los siguientes *items* de información son registrados:

- Nombre de la dimensión como es entendida por la gente de negocios. Por ejemplo “abonado”.
- Retrospección de la existencia de la entidad.
- Atributo existente para la entidad. Para las entidades con retrospección permanente.
- Frecuencia de la captura de cambios para la existencia de la dimensión. Esto ayudará a establecer si la dimensión será objeto de errores de una sincronización temporal.

Para cada dimensión, un conjunto de atributos también se define en una hoja de

trabajo separada. La siguiente descripción se refiere a las propiedades de los atributos no existentes. Así, para cada atributo, la siguiente información es registrada:

- Nombre dimensión: Nombre como la persona de negocios se refiere.
- Nombre atributo: Es el nombre como la persona de negocios se refiere a éste.
- Retrospección: Si se debe considerar valores históricos.
- Frecuencia: Se refiere a la frecuencia con la cual la *data* es registrada en el *DWH* para determinar con exactitud los procedimientos a seguir.
- Dependencia: Identifica otros atributos del cual depende este atributo.
- Atributo identificado: Este indica si el atributo es atributo identificado o si forma parte de un atributo.
- Metadata: Una descripción del atributo del negocio.
- Fuente: Es un mapeo de regreso a la fuente, describe el origen del atributo.
- Transformaciones: Cualquier proceso que se aplica al atributo antes de que éste sea llevado al *DWH*. Ejemplos: utilizar los mismos formatos, sustitución de valores por defecto en lugar de blancos o nulos.
- Tipo de Dato: Se refiere al tipo de dato y precisión del atributo.

La información acerca de las jerarquías dimensionales es capturada en la hoja de trabajo de jerarquías. La hoja de trabajo muestra los nombres de los componentes mayores y menores de la jerarquía. La siguiente información es capturada:

- Retrospección de la jerarquía.
- Frecuencia de captura.
- Metadata describiendo la naturaleza de la jerarquía.

5.3.3.2. Modelo punto y el Modelo Conceptual General

En el gráfico 5.5 se muestran por separado modelos dimensionales que comparten algunas dimensiones. Esto muestra que incluso con situaciones muy complejas, todavía resulta fácil determinar los modelos dimensionales individuales utilizando la notación punto porque la forma radial de cada modelo individual es todavía discernible.

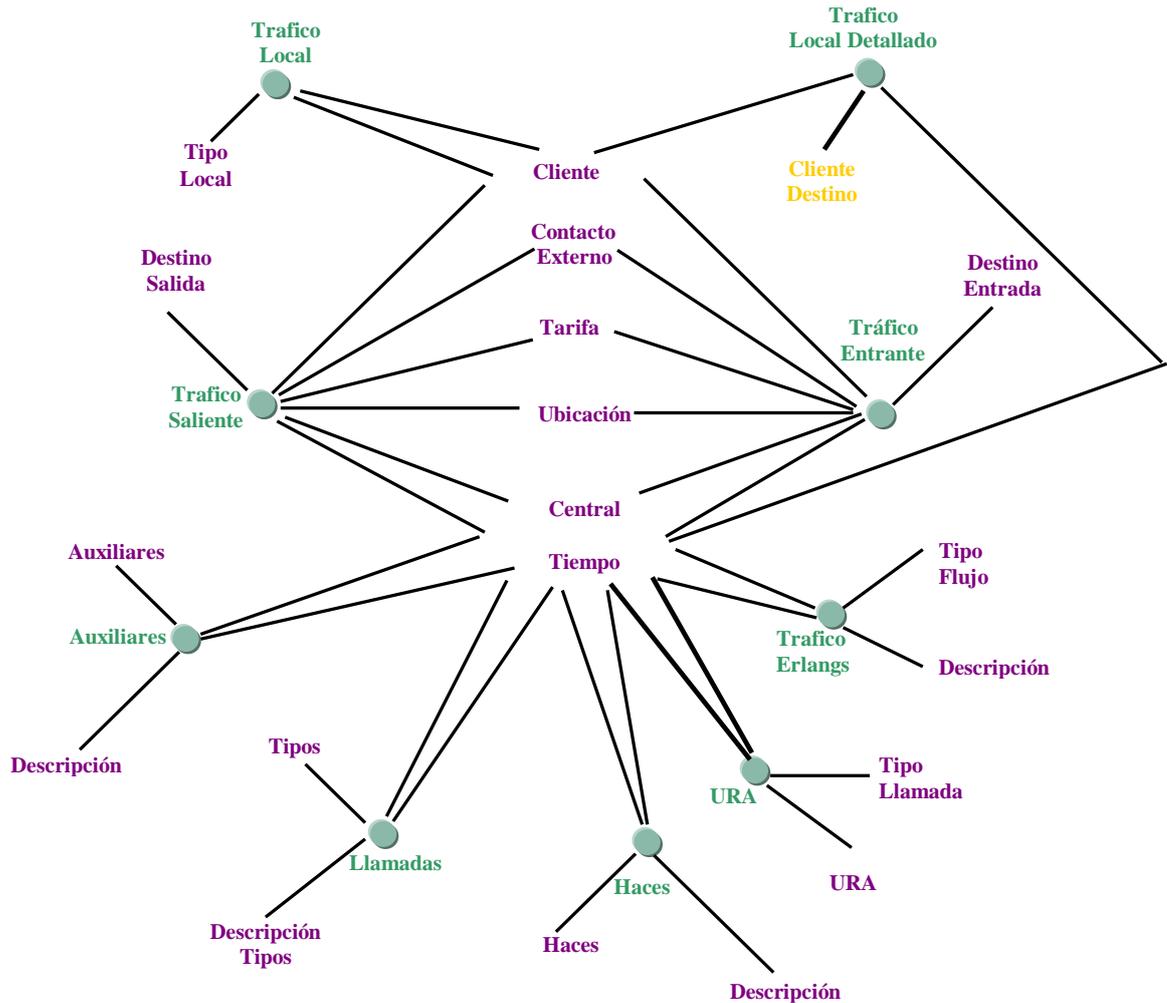


Gráfico 5.5. Modelo Punto para el Área de Tráfico Telefónico (E.T.A.P.A)

5.3.4. Talleres de trabajo del modelado punto

La etapa del diseño conceptual consiste del desarrollo de un modelo de datos usando el método del modelo punto. El modelo lógico es desarrollado como una extensión de dicho modelo.

5.3.4.1. Metodología del Modelado Punto

Existen un número grande de procesos asociados con la construcción de un modelo punto completo. Es posible conducir los procesos a través del uso de dos talleres de trabajo altamente estructurados. El uso de estos talleres de desarrollo de aplicaciones colectiva (*JAD*) es ahora una práctica bien conocida y consta de una o más sesiones de trabajo. Es una buena forma de obtener los requerimientos de usuario.

El propósito de los talleres de trabajo *JAD* es permitir desarrollar mejor los sistemas, reuniendo a la gente de negocios y a los desarrolladores del sistema. La idea es que el diseño del sistema sea un esfuerzo colectivo.

5.3.4.1.1. Taller de trabajo de información estratégica

El objetivo de este taller es permitir a la gente de negocios desarrollar su propio modelo punto para reflejar su percepción de negocios y trabajar conjuntamente.

Participantes

Debe existir, al menos, dos consultores en las sesiones del taller de trabajo, un facilitador organizado y un consultor de negocios que entienda de la metodología y la gente de negocios. Es muy útil contar con una persona para registrar los procedimientos de las sesiones. También es útil tener una persona extra que asista a las revisiones, esto es, chequear que todos entienden lo que se requiere y que el progreso está siendo realizado. El éxito ó fracaso de los talleres tiene que ver con el contenido técnico, pero tiene mucho más que ver con las siguientes circunstancias:

- Ambiente.
- Relaciones entre participantes.
- Cordialidad de los presentadores.
- Comodidad de los participantes.

Cuando el lenguaje de los participantes es diferente del lenguaje nativo de los consultores, se debe tener cuidado en asegurar que el mensaje, en ambas direcciones sea claro.

Sería ideal que al menos uno de los consultores debería hablar el lenguaje de los participantes. De no ser posible, entonces se debería contar con un intérprete. En este caso el taller se alargaría cerca de un 50% de tiempo de lo normal (2 días).

Cliente

La información estratégica del taller requiere una atención mixta de la gente de negocios y gente *IT*. Las proporciones ideales es de dos tercios de participación de la gente de negocios y un tercio de gente *IT*. Es muy importante que la sesión sea dominada por el personal *IT*. Deben estar presentes algunos representantes de niveles superiores de administración. Esto es el personal cuyo ingreso depende del éxito del negocio, ellos están en capacidad de influenciar en la estrategia y dirección de negocio. La clave para el éxito del *Data Warehouse* es enfocarse en los negocios y no en la tecnología. No es suficiente establecer meramente la información que es estratégica. Esta tendencia de citar los requerimientos de negocio como fundamental para el éxito es muy utilizada. Cualquier información del sistema debería existir solamente para soportar los firmes objetivos del negocio y es crítico que las metas del negocio sean definidas claramente.

5.3.4.1.1.1. Proceso del taller de trabajo de información estratégica

Diferentes consultores tienen formas diferentes de facilitar los talleres de trabajo. Un método que ha tenido éxitos en el pasado consiste de:

La organización del lugar donde se realiza el taller:

1. Las mesas de trabajo deberían estar organizadas en forma de una herradura en lugar del estilo de una clase. Así se alienta a la iteración.
2. Mucho espacio de escritura debe estar disponible, para registrar puntos importantes.
3. Se debe crear una atmósfera donde todos contribuyan.

5.3.4.1.1.1.1. Primer paso: Introducción al Taller

Presentación: Explicación del porqué el equipo está reunido y cuáles son los objetivos.

Objetivo: Al fin del taller de trabajo de información estratégica se habrá construido el modelo conceptual de los requerimientos de información para apoyar a la dirección de negocios de la organización.

No asumir que cada uno en el lugar de trabajo se conocerá con los demás. Se debe tratar de impulsar a que todos se conozcan. Esto es un método para romper el hielo que ha sido utilizado anteriormente con éxito. Se debe preguntar algo como:

- Nombre, Posición en la organización y
- Breve descripción de sus responsabilidades.

También se debe pedir que digan algo acerca de si mismos que no esté relacionado con el trabajo. Romper el hielo es muy importante en talleres de trabajo y no debe hacerse de prisa. Una forma de hacer esto, podría ser, permitir a todos hablar 10 minutos con la persona sentada a su lado y luego dejar que ellos se presenten unos a otros. Con esto conseguiremos *relax* en las personas. Puede recomendarse hacer una nota de quienes son gente de negocio y quien es gente *IT*. Es una buena idea presentarse uno mismo y dejar a los demás que sigan luego.

Reglas:

- Siempre se debe dar la autoridad colectiva de cambiar o añadir reglas.
- Sólo una persona está permitida a hablar en un momento determinado.
- **El silencio indica aprobación.** Si las personas no presentan objeciones a las decisiones, entonces indica que ellos las aceptan. No es útil que alguien, durante una sesión en el día dos, diga algo como: “No estuve de acuerdo con las ideas desde un principio”.
- **Las sesiones deben comenzar a tiempo.** Se debe establecer claramente y de forma precisa cuando comenzará la siguiente sesión. Es responsabilidad de todos estar a tiempo y no llegar tarde a las sesiones.
- **Teléfonos móviles.** No deben ser permitidos.
- **No, a las preguntas tontas.** Esto no significa que no estén permitidas las preguntas tontas. Esto de hecho significa que ninguna pregunta debería considerarse inválida.

- **Críticas Personales no son permitidas.** Esta bien para ellos criticar a los consultores ó el método, pero no a cada uno de los otros.

Delinear Agenda

El siguiente paso que se debe seguir es delinear la agenda para el taller de trabajo completo. A las personas les gustaría conocer, la hora de ingreso a los talleres, hora de salida al almuerzo, hora de regreso y hora de término de reuniones. Es siempre valioso preguntar si alguien tiene restricciones en el tiempo y tratar de acomodar de la mejor manera los horarios.

5.3.4.1.1.2. Segundo Paso: Primera sesión

Metas de Negocio

Esta es la primera sesión real de negocios. La siguiente pregunta debe ser planteada. ¿Tiene su organización metas de negocio?, está es una pregunta directa que debería ser realizada a la gente de negocios de alto nivel. Se debe tomar en consideración el lenguaje corporal rápidamente porque existen algunas personas que pueden sentirse:

- Apenadas por no saber cuales son sus metas de negocio.
- Incomodas por compartir sus metas de negocio con los demás.

La mayoría de personas tiene una idea vaga acerca de las metas de negocio y ofrecería cosas como:

- Incrementar una parte del mercado.
- Incrementar la fidelidad de clientes.
- Incrementar las utilidades.

Lo que se necesita es conseguir estar en posibilidad de articular las metas de negocio que tienen las siguientes propiedades:

- Medible.
- Tiempo limitado.
- Enfocado al cliente.

La tercera propiedad no es en absoluto un requerimiento, pero la mayoría de las metas de negocio están relacionadas de alguna forma con los clientes. Un ejemplo sería: Incrementar la fidelidad de los clientes en un 5% anual para los próximos tres años.

Si es posible conseguir una meta clara de la gente de negocio, es bueno. Si se pueden obtener dos o más, sería excelente. El resultado más probable es que no se consiga ninguna.

Pensar acerca de la estrategia de negocios

Este es el primer ejercicio real del taller de trabajo. Los participantes forman grupos de tres o cuatro personas. Para conseguir lo mejor del ejercicio, la composición de grupos es importante. Por eso es útil, conocer quienes son la gente de negocios y quienes son gente *IT*.

Es necesaria una persona de negocios de alto nivel, si es posible, en cada grupo. Las demás personas de negocios y la gente *IT* deben esparcirse de la mejor manera posible a los grupos. No se debe permitir a la gente *IT* llegar a ser dominante en cualquier grupo porque ellos tienden a ser los coordinadores en estas sesiones. Una forma para mejorar esto es que durante la sesión de introducción, el facilitador que no está hablando anote los nombres y títulos de trabajo de cada uno de los participantes y pueda formar grupos independientemente. Luego los participantes son informados a cerca del grupo al cual han sido asignados.

Los objetivos del ejercicio en esta etapa son:

1. Decidir acerca de las metas de negocios (asumiendo que ninguna se estableció anteriormente). Cada equipo trabajará en su conjunto de metas, en lugar de que todos establezcan metas comunes. Cada una de las personas de

negocios en el lugar tendrá sus propias metas de negocios, diferentes a la de los otros, una o dos son suficientes. Cada meta debe poseer las tres propiedades listadas anteriormente.

2. Pensar acerca de la estrategia de negocio les ayudará a lograr las metas. Esto es un conjunto priorizado de iniciativas que la organización necesite implementar. Es poco probable que un *item* simple, por sí mismo, sea suficiente para hacer que la meta sea cumplida.
3. Que pasos necesitaría tomar para poner la estrategia en operación. Los pasos serán conducidos por la priorización de los componentes estratégicos.

Se debe tener cuidado en permitir el tiempo suficiente para esto. Este proceso lidera a la discusión seria de grupos ya que cada uno tiene una visión de cada simple cosa que su organización necesita hacer para llegar a ser más exitosa. Estas discusiones, son buenas y se debería permitir su curso. Cerca de una hora debería permitirse para esta parte. Es valioso recomendar que cada persona tenga 15 minutos para pensar acerca del problema, así cada uno tendrá una contribución que hacer posteriormente cuando se formen los grupos.

Modelo punto inicial

El objetivo es lograr la comprensión de todos acerca de cómo usar el sistema del modelado punto. La gente *IT*, entenderá más rápido que la gente de negocios que no lo es.

Primero, en una pizarra se podría realizar una explicación con una hoja de cálculo. Todos conocen una hoja de cálculo y pueden relacionarla a los ejes, las celdas, etc. Luego, se podría construir un modelo frente a ellos.

Es vital que al menos alguno de los miembros del equipo entienda como funciona. Es útil también responder preguntas, especialmente de gente *IT*, para convencerse de que cada grupo está en capacidad de desarrollar un modelo rudimentario. En esta etapa no se espera modelos perfectos ya que existe una etapa de refinamiento posterior. Sin embargo es importante que conozcan la diferencia entre un hecho y una dimensión. Esto ayuda a dibujar una analogía de esquema estrella.

La mayoría de la gente *IT* estará en capacidad de relacionarlo de forma directa. La gente de negocios lo entenderá más lentamente.

Comportamiento

En esta parte se realizará la explicación del comportamiento. Estos son los atributos del punto en sí y representan el enfoque del modelo. Los hechos típicos que se pueden encontrar en la mayoría de áreas son los valores y las cantidades. Se debe tener cuidado de incluir hechos derivados tales como utilidades, retornos sobre la inversión, etc. Cada atributo de hecho debe tener las propiedades de ser:

- Numérico.
- Sumable o semisumable.

Cada hecho medible debe ganar su lugar en el *Warehouse*. Quien sugiera un hecho debería demostrar que éste tiene la capacidad de contribuir con la estrategia de negocios. Adicionalmente, las preguntas deberían ser formuladas, en forma de consultas (en el lenguaje natural, no *SQL*). Para mostrar como el hecho será utilizado.

El hecho, como será llamado en el *Data Warehouse*, deberá ser registrado. Es vital que cada uno tenga una idea de lo que significa los hechos y tiene sentido gastar un poco de tiempo en el enfoque de la semántica de la información. Todos deben tener un conocimiento claro de la *data* antes de que esta sea registrada. También se requiere conocer la granularidad de tiempo del hecho medible. Esto debería ser de forma ideal el nivel más bajo de granularidad de tiempo que es práctico. El mejor nivel posible es conocido como tiempo válido, el mismo que significa el tiempo en el que el evento ha ocurrido en la vida real. Esto generalmente se traduce al nivel de transacción. En una aplicación de telecomunicaciones donde el *Data Warehouse* registra llamadas telefónicas, el tiempo válido relacionado al tiempo de llamada sería establecido como el segundo. En otras aplicaciones, el tiempo de granularidad puede ser el Día.

Generalmente, todos los hechos compartirán el mismo tiempo de granularidad. Sin embargo, es importante conocer cuando éste no es el caso, así se registra el tiempo de granularidad para cada atributo de hecho medible. Para cada atributo de hecho, la metadata que soporta el atributo debería ser registrada, esto significa una descripción precisa del atributo.

Las Dimensiones

Se debe tratar de que los practicantes *Data Warehouse* entiendan este término y puedan estar en capacidad de explicar el término a gente no técnica. Una forma de lograr esto es dibujar una hoja de cálculo en dos dimensiones con ejes de cliente y destino por ejemplo y explicar que es un ejemplo de una vista de dos dimensiones del tráfico saliente.

Cada intersección ó celda, en la hoja representa el tráfico de un cliente a un destino particular. Se puede dibujar una intersección usando líneas y un punto donde se interceptan. El punto representa el hecho medible. La vista en dos dimensiones puede extenderse a una vista tridimensional transformando el diagrama en un cubo y etiquetando el tercer eje, como el tiempo por ejemplo. Entonces se tendrá una intersección de las tres dimensiones, el punto, representa el tráfico de un cliente a un destino en un momento particular. Entonces, se necesita un método en forma de diagrama que libere el punto de su limitación tri-dimensional y aquí es donde el modelo punto ayuda a describir las dimensiones como ejes de las hojas de cálculo multidimensionales.

Crear el Modelo Punto inicial

Tomando las metas de negocio, la estrategia y los pasos que ellos hayan decidido en la sesión de grupo, los grupos tendrán que seguir lo siguiente:

1. Usando su estrategia de más alta prioridad, decidir que información es necesaria para cumplir sus metas. Es bueno guiarlos con un ejemplo.
2. Formular preguntas que ellos crean poder estar en posibilidad de contestar.
3. Crear un modelo punto que sea capaz de contestar estas preguntas.

Los facilitadores deben estar cerca todo el tiempo para ayudar cuando se requiera. En esta etapa siempre tendrá preguntas como “¿Esto es una dimensión o un hecho?”. También la gente *IT* comenzará ya ha hacer preguntas acerca de cosas como:

- Disponibilidad *data* de sistemas fuente.
- Arquitecturas de la base de datos *Warehouse*.
- Problemas en la calidad de datos, etc.

Podrían ser un poco exigentes en conocer estos puntos. Es importante que no restrinja el proceso. Lo que se trata es de establecer la información que se necesita y no lo que está disponible actualmente.

Presentaciones de Grupo

Cada grupo deberá luego presentar, a los otros grupos:

1. Metas de negocios.
2. Pasos en su estrategia.
3. Información requerida para soportar los pasos.
4. Modelo de punto para soportar los requerimientos de información.
5. Algunas preguntas ejemplo de lo que el modelo soporta.
6. Como cada pregunta es relevante para su estrategia.

La información debería presentarse en un proyector o una pizarra.

Los facilitadores deben proveer retroalimentación. La retroalimentación en esta etapa debe siempre ser positiva con algunas sugerencias de cómo el modelo podría ser mejorado. No se debe comentar acerca de corrección e integridad de los modelos. El proceso de refinamiento resolverá estos problemas.

El proceso de refinamiento

Primero se debe decidir si los grupos continuarían trabajando en sus modelos ó si todos los grupos adoptarían, el mismo ó talvez un modelo compuesto. No existen

preferencias reales de forma en esta parte. En cualquier caso, el trabajo está todavía a cargo de los grupos. Entonces se explicará acerca de:

- Combinación de dimensiones, jerarquías dimensionales e
- Inclusión de sugerencias de otros.

El punto uno se refiere a las dimensiones en el modelo original que no son realmente dimensiones del todo. Se debe analizar cada una para saber si realmente es una dimensión o un atributo de una dimensión. La creación de jerarquías de dimensión se debe tomar en cuenta. Los grupos deberían ahora redefinir los modelos utilizando un tiempo aproximado de 15 minutos.

Presentar los Modelos Refinados

La información que cada grupo debe exponer:

- Como el modelo se ha desarrollado del modelo original como resultado de un proceso de revisión.
- Como algunos mejoramientos, de una retroalimentación previa, ha sido incorporada.

Documentar los Modelos Refinados

Los modelos son documentados, en su estado refinado, usando el modelo de datos y la primera parte de las entidades y hojas de trabajo de segmentación (las partes restantes de las hojas de trabajo de entidades y segmentación son completadas durante el segundo taller de trabajo).

Las entidades incluyen las circunstancias y las dimensiones de los modelos punto de comportamiento. La segmentación se relaciona con los segmentos derivados, lo que constituye la parte final del Modelo Conceptual General. Compartir una hoja de trabajo para estos dos componentes del modelo es una buena idea porque puede ser difícil para la gente de negocios separar las dos cosas en su mente. Se debe alentar a cada miembro de equipo a completar sus propias hojas de trabajo.

Estos modelos proporcionan la entrada primaria para el Taller de trabajo del Componente de Análisis.

5.3.4.1.1.2. Taller de trabajo cubierto

Es importante sumarizar el proceso anterior y felicitar a los participantes por haber dado los primeros pasos hacia la creación de una información que lidera el modelo de negocios para la organización:

1. Se debería explicar como el modelo constituye el primer paso y cómo llega a ser la entrada a la siguiente etapa. Brevemente describir los componentes del taller de trabajo del componente de análisis.
2. Tratar de asegurar el compromiso de parte de la mayoría de personas para apoyar al segundo taller
3. La organización para el segundo taller debería ser confirmada, si es posible.
4. Realizar un prospecto que apoye una forma de retroalimentación y pedir a cada persona que lo complete. Lo que se trata de hacer es mejorar el servicio que se ofrece, por lo que se necesita la retroalimentación

5.3.4.1.2. Taller de trabajo del Componente de Análisis

El principal objetivo del taller del componente de análisis es poner algo en el modelo que se creó en el taller de estrategia de la información.

Participantes:

Profesionales: de ser posible, la gente de consultaría que estuvo en el primer taller, debería estar presente. El ejercicio es más técnico y algunas habilidades adicionales son requeridas, la continuidad del personal debe preservarse.

Clientes: de forma similar, de ser posible, las personas anteriores deben estar presentes. Sin embargo, como se ha establecido, este taller es más técnico y algunas personas de negocios de alto nivel deberían estar permitidas a enviar personal

autorizado para trabajar en su lugar. Las metas de negocios y la primera parte del modelo punto ha sido establecida, así que, su parte de la tarea está completada.

Las proporciones para la gente de negocios y la gente *IT* puede invertirse en este taller, así tal vez las dos terceras partes de la gente *IT* estaría bien. Es, sin embargo vital que algunas personas de negocio estén presentes porque es importante la continuidad.

5.3.4.1.2.1. El proceso del taller de trabajo para el componente de Análisis

La organización de este segundo taller es muy similar a la primera. Las facilidades de escritura y el lugar donde se desarrolla el taller deben tener las mismas características de antes. No se necesita un lugar extra para el trabajo en grupo esta vez.

Revisión del Modelo Anterior

El propósito de este primer ejercicio es refrescar las mentes de los participantes para establecer el modelo punto. Desde el taller de trabajo anterior, es bastante probable que los participantes hayan pensado en otros refinamientos para el modelo. Si los refinamientos deben ser adoptados depende de la autoridad de las personas presentes. Mientras los refinamientos contribuyan a las metas del negocio deben ser incluidos donde sea apropiado. La magnitud del refinamiento es difícil de evaluar en términos generales. Idealmente, el líder consultor del proyecto estará en capacidad de estimar el tiempo requerido para conducir esta parte del ejercicio. Si no hay cambios requeridos, entonces esta parte no debería tomar mucho tiempo.

Definición de atributos

Este puede ser un proceso que consuma tiempo. El objetivo es construir una lista de atributos para cada dimensión en el modelo. La hoja de trabajo de soporte de las entidades del modelado punto debería ser usada para este ejercicio. Se comenzará completando las entidades en la hoja de trabajo de entidades en el primer taller de trabajo. Ahora puede completarla añadiendo más detalles de los atributos.

El formulario de la hoja de trabajo puede estar con mucha información, sin embargo, alguna información puede ser añadida luego en la etapa del modelo lógico. Luego se registrarán los nombres de los atributos y la metadata de negocios.

El nombre de atributo, como será conocido en el *Data Warehouse*, debería ser registrado. El nombre del atributo debe especificar claramente lo que representa, para la gente de negocios. En este nivel se considera una concisa pero precisa descripción del atributo en términos de negocio.

El facilitador llamaría a propuestas de atributos de las hojas de trabajo de los participantes. Cada atributo debe ganar su lugar en el modelo y demostrar como el atributo será usado en forma de consultas (lenguaje natural, no *SQL*). Como se ha establecido, este ejercicio puede ser un consumidor de tiempo y algo tedioso. Es recomendado que se realice con varios intervalos de descanso.

Análisis Dimensional de los Hechos

Para cada atributo de hecho medible se examina cada dimensión para determinar la magnitud de cual de las funciones aritméticas estándar puede ser aplicada. Esto nos permitirá distinguir entre los hechos completamente sumables y los hechos semi sumables. Por ejemplo:

- Las cantidades de un supermercado pueden ser sumadas por producto y no por cliente.
- Los balances de bancos pueden ser sumados en un punto particular del tiempo pero no a través del tiempo.
- Retorno de la inversión y algunos otros porcentajes, no pueden ser añadidos del todo, pero los máximos, mínimos y promedios pueden estar bien.

Cada uno de los hechos medibles debería someterse a una examinación para determinar las funciones aritméticas *standard* que pueden ser aplicadas de forma segura. La hoja de trabajo usada es llamada hoja de trabajo de uso de hecho, y un ejemplo se muestra en el Gráfico 5.6.

Uso de Hecho del Modelado Punto					
Nombre del modelo: Tráfico Saliente					
Nombre del Hecho: Fac. trafico entrante			Frecuencia: Mensual		
Dimensiones	Suma	Cuenta	Promedio	Min	Máx
1. Cliente		X			
2. Destino	X		X		
3. Ubicación	X		X		
4. Tarifa	X		X		

Gráfico 5.6 Hoja de trabajo de hecho¹
TODDMAN, Chris, "Designing a Data Warehouse"

Jerarquías y Agrupaciones

Sirve para proveer algo de metadata para describir las relaciones entre los diferentes niveles en la jerarquía. La metadata es requerida para proveer una descripción precisa de la relación en términos de negocio. La hoja de trabajo que se usa es llamada hoja de trabajo de las jerarquías y agrupaciones, en el gráfico 5.7 se muestra un ejemplo. La jerarquía o agrupación es identificada registrando los nombres de las dimensiones en el diagrama de jerarquía en la parte de arriba de la hoja.

Jerarquías y Agrupaciones del Modelado punto						
Nombre del Modelo: Tráfico Saliente						
Retrospección	VERDADERO	Frecuencia	Mensual			
Metadata						
Esta jerarquía registra la relación entre un cliente y el tráfico el cual el cliente cursa. Cada cliente tiene una o más llamadas de salida.						

Gráfico 5.7 Ejemplo de jerarquía²
TODDMAN, Chris, "Designing a Data Warehouse"

Retrospección

Es necesario presentar el concepto de retrospección a quienes asisten a las sesiones para que ellos puedan asignar valores de retrospección a cada uno de los componentes *Data Warehouse*. Es importante también comprender los problemas que se presentan alrededor de la retrospección, generalmente no es necesario que los clientes tengan el mismo nivel de entendimiento. Se debería presentar a ellos algo

que les permita alcanzar el significado de negocios del área sin llegar a enredarse en los detalles técnicos. El significado de retrospección varía ligeramente cuando se aplica a diferentes tipos de objetos *Data Warehouse*. Cuando es aplicado a dimensiones, el valor de retrospección se relaciona con la existencia de la dimensión en cuestión. Por ejemplo si necesitamos conocer cuantos clientes tenemos este momento, entonces debemos estar en capacidad de distinguir entre los clientes actuales y los anteriores. Si deseamos contestar: ¿Cuántos clientes tenemos este año comparado con el año anterior?. Entonces necesitamos conocer exactamente cuando un cliente llegó a ser un cliente y cuando dejó de serlo. Algunos clientes puede que hayan tenido un espacio de vida discontinuo. Para algunos períodos de tiempo ellos pueden ser clientes activos y durante el período donde se formula la pregunta, pueden no serlo.

Si es requerido registrar esta información fielmente entonces la retrospección es verdadera, se aplicaría a la dimensión cliente. Si un cliente se muda de una región a otra ó cambia algún atributo como el número de hijos por ejemplo, entonces, dependiendo de la aplicación, esto puede ser importante para poder rastrear la historia de estos cambios y asegurar que las consultas retornen resultados actuales.

Si ese es el caso entonces retrospección verdadera debería ser aplicada a los atributos de la jerarquía. En otros casos puede requerir registrar sólo los últimos valores, tales como el nombre de la esposa del cliente. En este caso, retrospección falsa debería aplicarse. En otros casos los valores no cambiarán nunca. Por ejemplo la fecha de nacimiento. En estos casos el valor de retrospección permanente será aplicado. El valor dado a cada objeto del *DWH*. Debe recordarse que los requerimientos relacionados a la retrospección son requerimientos de negocio y no técnicos.

Como determinar el valor de retrospección

Para establecer el valor de retrospección de cada objeto *Data Warehouse*, se debe investigar el uso de los objetos dentro de la organización. Una forma de hacerlo es preguntar al personal adecuado lo siguiente: Si este objeto cambiara ¿le gustaría rastrear la historia con precisión?.

Generalmente la respuesta a esta pregunta es la afirmación. Pero se debe analizar que la provisión de soporte temporal es costosa en términos de recursos y adversamente afecta a la ejecución, es muy importante que tal soporte sea proporcionado sólo donde éste agregue valor a la información.

Granularidad y la tabla punto de tiempo

Para dimensiones, jerarquías y atributos, la granularidad de tiempo se relaciona a la frecuencia con la cual los cambios de valores y cambios de existencia, son notificados y aplicados al *Data Warehouse*. Así, si por ejemplo, las actualizaciones de cliente van a ser enviadas al *Warehouse* en una base mensual, entonces la granularidad de tiempo es mensualmente.

Para completar con precisión la granularidad de tiempo para captura de cambios en las dimensiones debe ser lo mismo que la granularidad del tiempo para capturar los eventos de hecho medibles.

En la práctica existe un retraso inevitable en el tiempo entre el evento cambio y la captura de este cambio en el sistema fuente. Existe, generalmente, un retraso adicional entre la implementación del cambio en el sistema fuente y la subsiguiente captura del cambio en el *Data Warehouse*. El desafío para los diseñadores *Data Warehouse* es minimizar los retrasos en la captura de cambios.

La hoja de trabajo del punto tiempo es simplemente una lista de todas las cabeceras relacionadas con el tiempo por la cual los usuarios necesitarán agrupar los resultados de sus consultas. Los requerimientos variarán de una aplicación a otra. Es importante también establecer cuanta historia la aplicación desearían que tenga. Esto tiene dos principales beneficios:

1. Permite a los diseñadores saber cuantos registros tendrán que ser creados para poblar la tabla Punto _ tiempo.
2. Esto le da a los diseñadores algunas indicaciones del último tamaño de la base de datos cuando fue poblada por completo.

5.3.4.1.2.2. Taller de trabajo cubierto

Es importante sumarizar el proceso que ha sido conducido sobre el trabajo del segundo taller. Ahora se contará con un modelo conceptual completo dirigido a los negocios que los diseñadores pueden usar para ayudar a construir el *Data Warehouse*. Ahora se contará con todos los componentes del modelo conceptual y gran parte de la *data* ayudará a desarrollar el modelo lógico, el cual es el siguiente paso del proceso. ³

5.4. Modelo lógico

5.4.1. Introducción al modelo lógico

La mayoría de *Data Warehouses* no son implementados en bases de datos relacionales sino en sistemas de bases de datos dimensionales que son generalmente conocidos como Sistemas de Procesamiento Analítico en Línea. Actualmente no existen estándares para ayudar a los diseñadores en la producción de modelos lógicos para sistemas de bases de datos propietarias de *OLAP* como existen para los sistemas operacionales.

Cuando se trata de formular soluciones potenciales, los siguientes requerimientos necesitan ser satisfechos:

- Reporte exacto de los hechos. Es muy importante que, en un modelo dimensional, cualquiera que sea la entrada del hecho esté unida a una entrada de dimensión. El hecho debe unirse a la dimensión correcta con respecto al tiempo.
- Registro exacto de los cambios en entidades para el soporte de consultas involucradas en las circunstancias del cliente y la navegación de dimensión. Consultas como éstas representan un componente significativo del uso del *Data Warehouse*. Es importante asegurar que los períodos de existencia de dimensiones, relaciones y atributos sea registrado de forma exacta con respecto al tiempo donde esto haya sido identificado como requerimiento de negocios.

Una regla general es que:

“Cada consulta ejecutada en un *Data Warehouse* debe tener una restricción de tiempo”.⁴

Si una consulta no tiene una restricción de tiempo explícita, entonces el tiempo de período que se deduce es “todo el tiempo”. Las consultas que contienen todo el tiempo, en la medida en la que le concierna al *Data Warehouse*, puede ser generalmente consideradas como un absurdo porque “todo el tiempo” simplemente indica el lapso de tiempo que la base de datos existe.

Al realizar la consulta: ¿Cuántos clientes tenemos?, aunque esta consulta tiene una restricción de tiempo implícito, se pregunta realmente: ¿Cuántos clientes tenemos en este preciso punto del tiempo ó cuantos clientes existen actualmente?.

5.4.2. El uso de los atributos de existencia

Es necesario explorar sobre los atributos de existencia como una aproximación a la implementación de la retrospección. La existencia del atributo que se explicó en el modelo conceptual es una aplicación general de todos los componentes *Data Warehouse*. Existen varias formas de representar la existencia de un atributo. En el nivel más simple de la entidad, un atributo de existencia podría añadirse para registrar cada ocurrencia de si está activo o no. Si embargo, si se requiere un soporte completo de historia, entonces se requiere de un atributo de existencia compuesto donde se registre la fecha de inicio y de fin para denotar un período activo actual. Así cada atributo de existencia podría ser implementado como sigue:

1. Para soporte temporal de una entidad que pueda tener una existencia discontinua, se requiere una tabla separada donde esté la llave primaria de la entidad y un período de existencia. Si la existencia discontinua no es posible, el período de existencia puede ser añadido a la entidad.
2. Al igual que el anterior, para establecer una relación entre entidades, se requiere una tabla separada, que contenga las llaves primarias de ambas entidades que participan juntas con un período de existencia.

3. Para el soporte temporal de un atributo de una entidad, se requiere una tabla separada que contenga una llave primaria de la entidad, un período de existencia y el valor del atributo.

El uso de atributos de existencia en la forma de tablas adicionales añade complejidad, así que deben ser permitidos sólo para aquellos elementos (circunstancias, segmentación ó la estructura dimensional del *Data Warehouse*) donde exista una clara necesidad para el soporte temporal.

5.4.3. El uso de la dimensión tiempo

El propósito de la dimensión de tiempo es proporcionar un mecanismo para restringir y agrupar los hechos, como hacen las otras dimensiones en el modelo. Se trata de proveer un método para soportar el tiempo en las dimensiones, jerarquías así como también en los hechos.

Kimball proscribe el uso de la dimensión tiempo con otras dimensiones porque se enfoca en que las semánticas del tiempo y las dimensiones son diferentes de la semántica de los hechos. La vista de que la dimensión del tiempo no debería ser usada en las consultas del navegador es soportada de forma explícita por el modelo de datos convencional llamado esquema estrella, que muestra la dimensión del tiempo relacionada con la tabla de hechos nada más. No existen relaciones entre la dimensión del tiempo y alguna otra dimensión. En consideración de esto dos puntos emergen:

- Primeramente, la dimensión del tiempo provee una interfaz física para usuarios cuando se formulan consultas. Previniéndolos de usar la dimensión del tiempo con otras entidades lo que indica que los usuarios estarán en capacidad de establecer restricciones de tiempo seleccionando términos en consultas involucrados con la tabla de hecho pero no con otras entidades tales como las dimensiones. En este tipo de consultas los valores de tiempo explícitos deben estar codificados.
- Segundo, algunas consultas del navegador dimensional son mucho más fáciles de expresar si una unión es permitida entre las dimensiones.

Cuando la dimensión tiempo se una a las otras dimensiones, con atributos de existencia usados, permitirá tener una expresión más simple de algunas consultas temporales.

En un modelo así existirían dos relaciones entre la tabla del tiempo y una tabla de dimensión, una para el tiempo de inicio y otra para el tiempo final. Un problema causado por este cambio en el modelo convencional es que los modelos dimensionales perderán inmediatamente su forma simple y el modelo completo llegará a ser muy complejo. La simplicidad es uno de los requerimientos establecidos en el modelo. Un problema que se presentaría sería que el modelo no debe perder su forma simple y claramente no se puede presentar a los usuarios tal diagrama.

La solución sería que la tabla de dimensión del tiempo sea removida del modelo. Sin embargo dicha dimensión siempre está incluida como parte del modelo dimensional porque el tiempo es siempre una dimensión de análisis en un *Data Warehouse* que registra la historia. La metodología punto sería adoptada para acoplarse a los requerimientos. La dimensión tiempo debe tener atributos específicos para cada aplicación. Tal vez se podría requerir información sobre el tiempo, tal como:

- Cerrar al medio día.
- Efectos de abrir tarde debido al entrenamiento de personal.
- Abrir las 24 horas.

Así, existe una necesidad de especificar los atributos para el tiempo en las bases de cada aplicación. En la metodología del modelado punto es posible resolver este problema utilizando una tabla que satisfaga los requerimientos anteriores manejados por dimensiones de tiempo explícitas. Cada aplicación tendrá sus propios requerimientos los cuales serán cubiertos con una tabla que se recomienda siempre debe tener:

- Fecha, Nombre del día.
- Número de semana, Nombre del mes.

- Número de mes, Trimestre
- Año.

Se podría añadir un valor a los clientes proporcionando una tabla tiempo de inicio que puede contener, 10 años de historia y 10 años de fechas futuras. Esto parece una gran cantidad de datos, pero en realidad es menos de 8000 filas donde la granularidad es diaria dependiendo de la organización. Para niveles más finos de granularidad, por ejemplo, segundos, es sensible proporcionar dos tablas de tiempo. La primera contiene todos los días requeridos como antes y el otro contendría una entrada para cada segundo de un día simple (por ejemplo, de 00:00:00 a 23:59:59). Es entonces un asunto simple unir a una tabla, en el caso de cambios dimensionales ó ambas tablas en el caso de llamadas telefónicas. De esta forma, múltiples granos de tiempo pueden ser adaptados. Se podrían proporcionar columnas adicionales como fines de semana, fechas de días festivos, etc. dependiendo del negocio que utilice el modelo.

5.4.4. Esquema lógico

El esquema lógico ha sido incluido como una ayuda para clarificar y no tiene la intención de prescribir un modelo físico. Los problemas de ejecución tendrán que ser considerados cuando se implemente la solución y algo de denormalización debería aplicarse. Para la implementación de una retrospección verdadera para cualquier entidad, debe considerarse las indicaciones establecidas en el Modelo Conceptual General.

5.4.5. Consideraciones de ejecución

Las consideraciones de ejecución tienen que ser hechas en algún punto y esa ejecución puede ser considerada más importante que la precisión en las decisiones de diseño que son tomadas. Es recomendable pasar valores a la tabla de hecho, para resolver los problemas de ejecución. Los atributos de la tabla de hecho siempre tienen la propiedad de una retrospección permanente. Así la precisión de los resultados no se vería comprometida y la ejecución mejoraría considerablemente, incluso mejor que el original.

Sin embargo las relaciones de ciertas dimensiones deben ser dejadas intactas debido a que las jerarquías no pueden ser reconstruidas desde la tabla de hecho.

5.4.6. Seleccionar una solución

En la persecución de precisión relacionada con el tiempo, necesitamos saber si la *data* que se está recibiendo para ser establecida en el *DWH* actualmente refleja el tiempo en que el cambio ocurrió realmente.

Los cambios de algunos atributos son más importantes que otros con respecto al tiempo. Para los cambios más importantes, se espera registrar los tiempos en que los cambios ocurrieron. Algunos sistemas están en la posibilidad de proveer el tiempo válido de cambios en los atributos, pero la mayoría no están equipados para hacer esto. Así que hay que enfrentarse al problema de deducir los cambios por algún tipo de proceso de comparación que periódicamente examine los valores actuales comparándolos con los valores anteriores para determinar precisamente: qué ha cambiado y cómo.

La única clase de tiempo disponible en este escenario es el tiempo de transacción. Bajo circunstancias normales, el tiempo de transacción es el tiempo en que los cambios son registrados en el sistema operacional. Sin embargo, los cambios del tiempo de la transacción no son realmente registrados en ningún lado por la aplicación. Los cambios, como una dirección de cliente, simplemente resultan en el reemplazo de la dirección anterior por la dirección nueva, sin un registro de cuando el cambio fue realizado. Otros sistemas tratando de detectar el cambio, por el método de archivos de comparación, no tienen una forma real de conocer cuando el cambio ocurrió en el mundo real o cuando fue registrado en el sistema.

Así, en el ambiente *Data Warehouse*, existen dos desfases de tiempo a ser considerados. El primero está entre el tiempo en el que ocurrió en el mundo real, el tiempo válido y el tiempo en que el cambio fue registrado en un sistema operacional (tiempo de transacción). Generalmente, la organización está desprevenida del tiempo inválido de un evento de cambio. En cualquier caso, el tiempo válido es raramente registrado.

El segundo desfase de tiempo es el tiempo que toma para el cambio, una vez que este ha sido registrado en el sistema operacional, para encontrar su camino en el *Data Warehouse*. La solución es tratar de minimizar los desfases de tiempo inherentes en este proceso. Aunque esto es a menudo más fácil de decir que hacer, el objetivo de los diseñadores debe ser identificar y procesar cambios tan rápido como sea posible para que el aspecto temporal de los hechos y dimensiones pueda ser sincronizado.

5.5. La implementación física

Existe un flujo tradicional de datos desde los sistemas fuente al *Data Warehouse*. El proceso operacional comienza por la extracción periódica de datos, tiene que probarse la calidad de la *data*, ésta es una de las partes más importantes que deben tomar en cuenta los desarrolladores del *Data Warehouse*. La *Data* pobre en calidad amenaza el éxito el proyecto completo.

También, es necesario considerar lo que sucede en el *Data Warehouse* cuando los cambios ocurren en los sistemas fuentes. Los cambios pueden ser muy ligeros, teniendo poco o nada de impacto ó pueden ser masivos, como en el caso de la introducción de un nuevo sistema de planeación de recursos (*ERP*) que cambiará significativamente la fuente de datos del *Data Warehouse*. Diseñar un *DWH* que pueda cubrir esto es similar a diseñar una construcción de 100 años de historia en una zona de terremotos. Cada día, nueva *data* es extraída, la calidad es revisada, cargada y sumariada.

5.5.1. Modelo de desarrollo *Data Warehouse*

Ha existido una consideración enorme en la industria acerca de varios modelos de desarrollo de los *Data Warehouses*. La discusión se enfoca en los beneficios de los *Data Marts* ó sistemas de aplicación centrados en aplicación, a veces conocidos como niveles de departamento del *Data Warehouse*. La razón de esto es debido a que los *Data Warehouse* empresariales son de gran riesgo y a menudo, no justifican los extensos y costosos tiempos de entrega. Un *Data Mart* departamental, por otro lado, puede ser implementado rápida y fácilmente con un bajo riesgo.

El gran inconveniente de éstos sistemas es que no existe consistencia en su modelo y no existe generalmente una forma de integrar la *data* para que cualquier pregunta a lo largo de la empresa pueda ser respondida. Son atractivos por su bajo riesgo y esto es debido a que tienen un bajo costo relativamente. Los beneficios que ellos traen son igualmente pocos.

Lo que se necesita son sistemas de información que ayuden al negocio en la persecución de sus metas, esto es, el *Data Warehouse*. El truco es adoptar un modelo de bajo riesgo y bajo costo, con un sistema completamente integrado. Esto se logra adoptando un modelo incremental, para lo cual se requiere revisar lo referente a los sistemas de bases de datos. La introducción a los sistemas de administración de bases de datos tuvo un cambio en el énfasis desde las aplicaciones céntricas hacia un modelo de datos para desarrollar sistemas. Los principales objetivos de los sistemas de administración de bases de datos son el mejoramiento de:

1. Evolución: La habilidad de que la base de datos se adapte a las necesidades cambiantes de la organización y a la comunidad de usuarios. Es decir que tendrá la capacidad de crecer en volúmenes de datos, aplicaciones y usuarios.
2. Disponibilidad: Asegurar que la *data* tiene estructura y está en capacidad de ser revisada de diferentes formas por aplicaciones diferentes y usuarios con requerimientos específicos y no específicos.
3. Compartir: Reconocer que la *data* pertenece a toda la organización y no a usuarios individuales o grupos de usuarios.
4. Integridad: Mejorar la calidad, mantenimiento de existencia y aseguramiento de privacidad.

Algunos sistemas son mejores en estos aspectos, que otros. Sería difícil argumentar que algunos de los sistemas analíticos en línea (*OLAP*) con estructuras de bases de datos propietarias logran algunos de éstos cuatro criterios en términos empresariales.

El punto es que aún contando con sofisticados sistemas de administración de bases, somos culpables de desarrollar sistemas centrados en aplicación en los cuales los datos no pueden ser fácilmente accesados por otras aplicaciones, lo que es aceptable en sistemas transaccionales, donde tales interfaces son de beneficio y se pueden

construir procesos de integración para permitir el intercambio de información. Con los sistemas informacionales, sin embargo esta restricción no es aceptada.

Por otra parte los objetivos de Evolución, Disponibilidad, Compartimiento, e Integridad son aún completamente válidos e incluso más con los sistemas de soporte de decisiones donde lo siguiente existe:

1. La naturaleza del acceso de usuario tiende a ser no estructurada.
2. La información es requerida para un amplio número de usuarios con diferentes necesidades.
3. El sistema debe estar en capacidad de responder a los cambios en la dirección de negocios de modo oportuno.
4. Los resultados de consultas deben ser confiables y consistentes.

El desarrollo de una solución centrada en aplicación no soportará tales objetivos. La única forma para asegurar que la base de datos posea estas cualidades es diseñarlas desde un inicio. Este requerimiento ha dado lugar al *EASI* (*evolvability/availability, sharability, integrity*) evolución/disponibilidad, compartir e integridad. El cimiento de *EASI* es el bloque de datos, es decir la *data* que contiene el nivel más bajo de detalle (granularidad). Idealmente, esta debería estar al nivel de transacción.

El bloque de la *data* es alimentado por los sistemas operacionales de la organización. El bloque de datos soporta la información y sistemas de soporte de decisión (aplicaciones) que están en capacidad de tomar la *data* desde el bloque, así como almacenar la *data* para su propiedad.

Las aplicaciones son independientes tienen sus propios ciclos de procesamiento y sus propios usuarios. Las aplicaciones pueden, proveer *data* para otros sistemas ó pueden incluso actualizar los sistemas fuentes. El punto más importante es notar que el bloque de *data* es un recurso que está disponible para todas las aplicaciones. Este es completamente consistente con los objetivos originales de los sistemas de administración de sistemas.

5.5.1.1. La capa de validación, integración y mapeo (VIM)

Para que la *data* sea permitida entrar al bloque, cada *item* de datos debe ajustarse a restricciones rigurosas considerando su estructura e integridad. Cada *item* de datos debe pasar por una serie de pasos de proceso conocido como *VIM*. *VIM* indica validación, integración y mapeo. Los procesos *VIM* forma la base de la calidad de los datos asegurando la estrategia para el *Data Warehouse*. Cercanamente asociado con el proceso *VIM* está una capa extensa de la metadata.

5.5.1.1.1. Validación de datos

Es la primera parte del aseguramiento de calidad de datos. La validación de datos es uno de los pasos más importantes en la carga de la *data* dentro del *Data Warehouse*. Aunque esta es un área que es a menudo subestimada en la planeación del *DWH*, desafortunadamente la mayoría de organizaciones no tiene una idea de la calidad de sus datos y son siempre sencillos en este aspecto. Si las organizaciones dicen que es casi perfecta la calidad, lo más seguro es que la calidad sea baja. Claramente, si la calidad de la *data* es pobre, entonces la exactitud de cualquier información puede estar derivada de ésta, entonces será altamente sospechosa y no se debe permitir decisiones de negocios que estén basadas en información de exactitud que se pueda cuestionar. Si esto se llegara a conocer, entonces los usuarios perderían la confianza en el *Data Warehouse* y dejarían de usarlo. Existen diferentes problemas de calidad, la *data* puede ser:

1. Pérdida (o parcialmente perdida).
2. Tener errores.
3. Estar fuera de tiempo.
4. Ser inconsistencia.

5.5.1.1.1.1. Pérdida de datos

La principal razón para que la *data* se pierda es el insuficiente rigor en las rutinas de procesos de colección de datos de los sistemas operacionales. Este puede ser campos en una ventana de entrada o un documento dónde, para propósitos de la aplicación la

data no es clasificada como obligatoria, mientras que, en el ambiente *Data Warehouse* es absolutamente obligatoria.

Resolver el problema de la pérdida de datos puede ser un problema grave. Idealmente, los valores que se pierden deberían ser clasificados dentro del sistema fuente antes de que la *data* sea extraída y transferida al *Warehouse*. A veces esto no es posible hacer y a veces no es deseable. En este caso el diseño *Warehouse* tiene que incorporar algún método para resolver esto. Es muy riesgoso permitir *data* en el *Data Warehouse* con pérdida de valores, especialmente si esto resulta en valores nulos.

El modelo de corrección de valores perdidos es un problema de negocios. ¿Cuanta importancia pone el negocio en este elemento de datos particular?, donde existe un alto nivel de valor puesto en esto, entonces se debe tener algún proceso de definición para manejar los registros donde los valores son perdidos. Existen dos modelos principales:

1. Uso de un valor por defecto.
2. Rechazo del registro.

Usar valores por defecto significa que el rechazo de registros puede ser evitado, ellos pueden ser usados efectivamente en campos numéricos ó de texto y en campos de llaves foráneas. Es posible usar el término “desconocido” o “no definido”. En el caso de llaves foráneas, entonces se necesita agregar un registro en la tabla de dimensión con una llave principal “UN” (única) y una descripción correspondiente. Un beneficio de los valores por defecto es que siempre podemos ver la magnitud con la cual el valor por defecto ha sido aplicado. La aplicación de valores por defecto debería, sin embargo, limitarse a conjuntos de tamaño conocido.

Cuando la columna tiene valores continuos numéricos, el valor por defecto no puede ser “desconocido”. La única forma de tratar esto es prevenir a tal registro de ingresar al *Data warehouse*. Así que deben rechazarse. Rechazar registros indica que ellos no son permitidos en el *DWH* en su estado actual. Existen tres principales modos de rechazar los registros:

1. **Rechazo permanente:** En esta situación, el registro es simplemente alejado y nunca entra al *Data Warehouse*. El problema obvio con esta solución es que existe un compromiso de simplicidad de operación versus la precisión de la información. Donde los volúmenes son pequeños en proporción al todo, esto puede ser aceptable pero debe ser realizado con extremo cuidado. Este tipo de solución puede ser aplicado sólo para registros de comportamiento. Por ejemplo si se rechaza un registro de cliente permanentemente porque la dirección se perdió, entonces cada subsiguiente llamada telefónica que el cliente podría tener entonces tiene que ser rechazada porque no existe un registro de cliente con el cual relacionarla.
2. **Rechazo para arreglo y readmisión.** Con este modelo, los registros perjudicados son puestos de forma separada en un archivo que está sujeto a algunos procesos. Una vez que estos problemas han sido solucionados, los registros son reingresados en el *DWH*. La forma más simple de hacer es unir los registros corregidos con el nuevo lote de registros que sean recolectados de los sistemas operacionales para que sean revalidados automáticamente como antes. Si se elige este proceso se debe construir un conjunto de proceso, casi como un subsistema, para manejar los rechazos, permitiéndolos ser cambiados y modificados de alguna forma y ubicados de vuelta a la cola. En este caso se ven cientos de millones de registros rechazados y el formato de los registros han cambiado y es poco probable que puedan ser corregidos. El archivo de rechazo contiene algunos registros buenos y una pila entera de basura. Esta situación requiere una restauración y un reentrenado.
3. **Rechazo con restitución automática.** Este es un tipo de variación de los anteriores tipos de rechazo. Algunas veces simplemente se restituyen los registros perjudicados. Un ejemplo de esto podría ocurrir en una empresa de telecomunicaciones con un cliente nuevo que adquiere una línea telefónica. En cuanto la línea sea adquirida se realizarán llamadas, sin embargo puede no estar registrado el cliente aún. Por lo que las llamadas en el *Data Warehouse* no encontrarán un cliente con el cual coincidir. Por lo que los datos perdidos en este caso son las circunstancias del cliente: nombre, dirección, etc. Así las llamadas serán rechazadas al ingresar. Pero no se requiere que pasen por ningún proceso de cambio, sino se necesita que los datos del cliente sean

registrados en el *Data Warehouse*. Cuando esto suceda, las llamadas serán aceptadas sin intervención adicional.

5.5.1.1.1.2. Data errónea

Existen diferentes tipos de *data* errónea:

Valores fuera del rango válido. Estos son usualmente fáciles de reconocer, este tipo de error ocurre cuando por ejemplo en el campo sexo, se ha establecido valores F o M por ejemplo y aparece un campo con un valor X. Esto también se puede aplicar a valores numéricos como la edad donde se pueden permitir valores de 18 a 100 años donde una edad caiga en un rango fuera, entonces tendrá errores.

Errores referenciales. Este es un error que viola una restricción de integridad referencial. Un registro es presentado con una llave foránea que no coincide con una llave maestra. Esto podría ocurrir si, un registro es borrado y subsecuentemente la transacción registrada no tiene nada con que coincidir. Esto se manejará de acuerdo a las políticas internas de la organización, pero se recomienda que los registros que contengan errores referenciales no sean subidos al *DWH*, ya que una vez establecidos dentro del *Data Warehouse*, serán difícil de encontrar y serán la mayor causa de resultados inconsistentes.

5.5.1.1.1.3. Data fuera de tiempo

Información fuera de tiempo en el *Data Warehouse* es casi imposible de ser detectada: valores que han ingresado incorrectamente, pero que todavía son válidos. Por ejemplo, una edad de cliente ha sido ingresada como 26 en lugar de 62. Se debería tratar de recibir cambios de los sistemas operacionales en el *Data Warehouse* lo más frecuentemente posible.

5.5.1.1.1.4. Data inconsistente

La *data* inconsistente en el *DWH* es un resultado de dependencias ó cambios casuales de los cuales los sistemas operacionales pueden ser ignorantes. Por ejemplo cuando

un cliente se cambia de domicilio a otro, eventualmente el *DWH* debe ser informado acerca de esto y tendrá que realizar apropiadamente el ajuste a las circunstancias del cliente.

5.5.1.1.2. Integración

La siguiente parte del procesamiento *VIM* es la capa de integración. La capa de integración asegura que el formato de la *data* sea correcto. La integración de formato significa que, por ejemplo todo se vea igual (por ejemplo el formato *yyyymmdd*) ó talvez algunos códigos de productos de algunos sistemas pueden necesitar ser convertidos para cumplir con una vista estándar de productos.

Para hacer la integración más flexible, en la medida de lo posible debería ser implementada en la *metadata*.

5.5.1.1.3. Mapeo

La capa de mapeo es un proceso de datos conducido que toman los datos fuente y los lleva al formato objetivo. El propósito es de facilitar la transición cuando sea requerida, por ejemplo, reemplazar un sistema fuente con otro ó aceptar alimentar *data* alterada. El objetivo es desarrollar una capa de mapeo formal que haga fácil la implementación de cambios en fuentes donde la funcionalidad no adicional es añadida (por ejemplo un resultado de actualizaciones de sistemas fuente).

5.6. Conclusión

Al haber analizado el modelo conceptual, lógico y físico, estamos en capacidad de desarrollar nuestro *Data Warehouse*, tomando en cuenta las necesidades de la empresa en la cual se implementa la bodega de datos.

El estudio de las circunstancias como ha sido presentado sería válido para el desarrollo de un modelo orientado al cliente, el cual no ha sido aplicado a la realización de esta tesis, debido a que en el área de la empresa en donde se desarrolló el proyecto las circunstancias del cliente no agregan valor, por lo que no se utilizó la

Administración de las Relaciones con el Cliente (*CRM*). Sin embargo en algunos casos se tomarán en cuenta los aspectos relevantes del modelo de comportamiento, debido a que se almacena un histórico de los datos cambiantes. Es vital obtener información a tiempo y estandarizada, de esta forma la empresa se esfuerza por crecer y dar un mejor servicio a los abonados.

En el proyecto realizado se ha registrado información histórica tomando en cuenta algunos aspectos de la retrospectiva. Se ha cubierto en un 100% los aspectos referentes a la validación e integración de la data, en la fase de desarrollo.

Los aspectos que no se han tomado en cuenta para el desarrollo del *Data Mart* de Tráfico Telefónico en nuestra tesis, se han incluido porque pueden ser adoptados por otros proyectos y han sido tomados del libro “*Designing a Data Warehouse*” del autor *Chris Toddman*.

REFERENCIAS

¹ TODDMAN, Chris, “*Designing a Data Warehouse*”, Hall PTR, New Jersey, Octubre 2001, pp 152.

² TODDMAN, Chris, “*Designing a Data Warehouse*”, Hall PTR, New Jersey, Octubre 2001, pp 153

³ TODDMAN, Chris, “*Designing a Data Warehouse*”, Hall PTR, New Jersey, Octubre 2001, pp 117-215.

⁴ TODDMAN, Chris, “*Designing a Data Warehouse*”, Hall PTR, New Jersey, Octubre 2001, pp 161.

CAPÍTULO 6

PROCESAMIENTO ANALÍTICO EN LÍNEA

6.1. Introducción

El Procedimiento Analítico en línea (OLAP) es necesario para el análisis de información para la toma de decisiones (objetivo del *DWH*), utilizando bases de datos multidimensionales donde se almacena grandes cantidades de información, la misma que es vista como cubos de donde se extraen los datos para brindar los reportes necesarios a los usuarios.

En este capítulo se conocerá más sobre *OLAP* y sus operaciones, necesarias para extraer la información de los cubos. Se podrá encontrar las diferentes alternativas de almacenamiento, así como la arquitectura de cada una de ellas, ventajas y desventajas. Esto permitirá poder elegir entre un modo de almacenamiento, dependiendo de las necesidades de la empresa y el factor que se considere más importante.

Se deberá tomar en cuenta una vez más el almacenamiento y el tiempo de ejecución de consultas.

6.2. Bases de datos OLAP

OLAP (Online Analytical Processing) es el proceso de análisis de información útil para la toma de decisiones desde fuentes de datos especialmente estructuradas (*Data Warehouse* y *Data Mart*). Es un proceso que le permite al usuario extraer y ver datos en una forma fácil y selectiva.

Para facilitar el análisis, *OLAP* se almacena en una base de datos Multidimensional la misma que está diseñada para analizar grandes grupos de registros. Las bases de datos relacionales pueden ser consideradas como de dos dimensiones, mientras que las multidimensionales consideran cada atributo (producto, región, período de tiempo, etc.) como una dimensión separada.

Las aplicaciones *OLAP* pueden ubicar la intersección de las dimensiones y mostrarlas, es decir, los datos son clasificados en diferentes dimensiones y pueden ser vistas unas con otras en diferentes combinaciones para obtener diferentes análisis de los datos que contienen. Atributos como períodos de tiempo, pueden ser segmentados en subatributos. En este modelo los datos son vistos como cubos, los cuales están compuestos de categorías descriptivas (dimensiones) y valores cuantitativos (medidas).

Una base de datos *OLAP* no necesita ser tan grande como una bodega de datos, dado que no toda la información transaccional se requiere para el análisis de tendencias. Mediante el uso de conectividad de bases de datos abierta (*ODBC*), los datos se pueden importar de bases de datos relacionales existentes para crear una base de datos multidimensional para *OLAP*, la misma que se puede utilizar para minería de datos ó para el descubrimiento de relaciones antes no conocidas entre datos.

6.3. Tipos de almacenamiento de datos

6.3.1. MOLAP

MOLAP (Multidimensional Online Analytical Processing) es un proceso analítico en línea (*OLAP*) en el cual la información se almacena multidimensionalmente, para ser visualizada multidimensionalmente. Una aplicación *OLAP* considera los datos en forma multidimensional, el usuario es capaz de ver diferentes aspectos o facetas de agregados de datos, como el tráfico por destino, cliente y categoría.

La información en una base de datos relacional, se puede ver en forma multidimensional, pero sólo mediante el acceso secuencial y el procesamiento de una tabla para cada dimensión o aspecto que se quiera ver. Los procesos *MOLAP* usan información almacenada en un arreglo multidimensional, en la cual todas las posibles combinaciones de datos se reflejan. Cada dato en una celda puede ser accesado directamente. Por ésta razón, *MOLAP* es, para la mayoría de los usos, más rápida y más ágil que el *ROLAP* para responder al usuario. El *MOLAP* se usa particularmente en proyectos de bodegas de datos.

6.3.1.1. Arquitectura *MOLAP*

El sistema *MOLAP* utiliza una arquitectura de dos niveles: La bases de datos multidimensionales y el motor analítico.

- La base de datos multidimensional es la encargada del manejo, acceso y obtención del dato.
- El nivel de aplicación es el responsable de la ejecución de los requerimientos *OLAP*. El nivel de presentación se integra con el de aplicación y proporciona un interfaz a través de la cual los usuarios visualizan los análisis *OLAP*.

La información procedente de los sistemas operacionales, se carga en el sistema *MOLAP*, mediante una serie de rutinas *batch*. Una vez cargado el dato elemental en la Base de Datos Multidimensional (*MDDDB*), se realizan una serie de cálculos en *batch* para calcular los datos agregados, a través de las dimensiones, rellorando la estructura *MDDDB*. Posteriormente se generan unos índices y algoritmos de tablas *hash* para mejorar los tiempos de accesos a las consultas. Una vez que el proceso de compilación ha concluido, la *MDDDB* está lista para su uso. Los usuarios solicitan informes a través de la interfaz y la lógica de aplicación de la *MDDDB* obteniendo el dato.

6.3.1.2. Ventajas

- Mejora la ejecución en el procesamiento de consultas.
- Menor tiempo de demora al realizar cualquier cálculo.
- Posibilita hacer cálculos más complicados.

6.3.1.3. Desventajas

- Tamaño limitado para la arquitectura del cubo.
- No es posible acceder a datos que no están en el cubo.

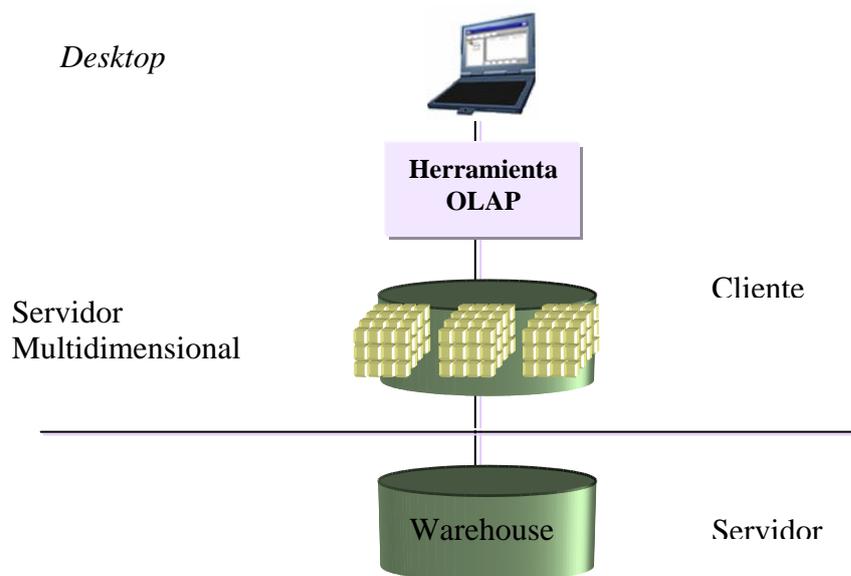


Gráfico 6.1. MOLAP

6.3.2. ROLAP

Relational Online Analytical Processing es una forma de procesamiento analítico en línea (*OLAP*) que ejecuta el análisis multidimensional sobre datos almacenados en una base de datos relacional (*OLAP Relacional*), en lugar de una base de datos multidimensional, como se considera el estándar de *OLAP*.

6.3.2.1. Arquitectura

El procesamiento de los datos puede suceder en el sistema de bases de datos, un servidor de capa media ó en el cliente. En una arquitectura de dos niveles, el usuario envía una consulta en *SQL* a la base de datos y recibe el resultado de la misma.

En una arquitectura de tres niveles, el usuario envía el requerimiento para el análisis multidimensional y el motor *ROLAP* lo convierte a *SQL* para ser enviado a la base de datos.

Posteriormente se ejecuta la operación en sentido contrario. El motor convierte los resultados del *SQL* a un formato multidimensional antes de servírselo al cliente para su análisis.

6.3.2.2. Ventajas

- Uso total de la seguridad e integridad de la base de datos.
- Escalable para grandes volúmenes de información.
- La data puede ser compartida con aplicaciones SQL.
- Contiene una *Data* y una estructura más dinámicas.

6.3.2.3. Desventajas

- Consultas más lentas.
- Construcción costosa.
- Los cálculos están limitados a las funciones de las bases de datos.

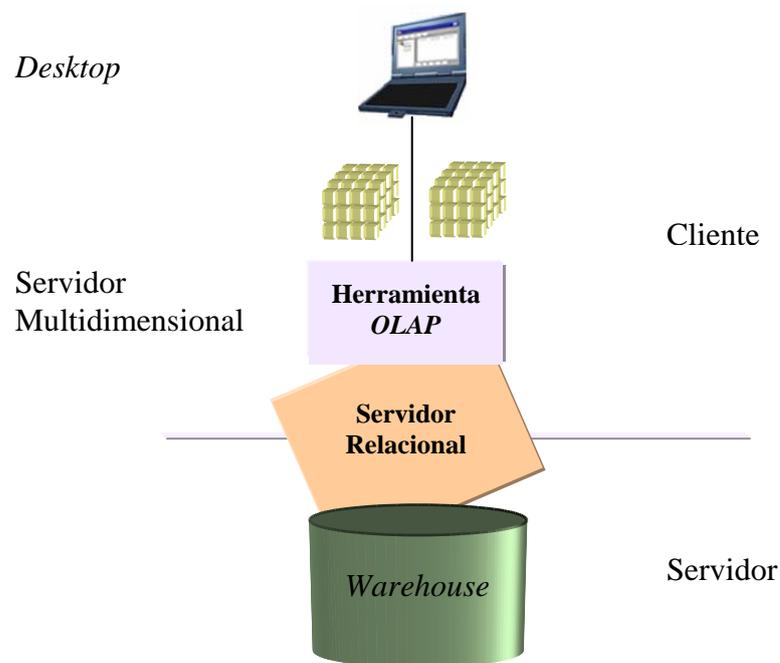


Gráfico 6.2. ROLAP

6.3.3. HOLAP

El *OLAP* Híbrido, es una combinación de *ROLAP* Y *MOLAP*. Los datos agregados y precalculados se almacenan en estructuras multidimensionales y los registros de menor nivel de detalle en una relacional. Requiere un buen trabajo de análisis para identificar cada tipo de dato. Son conocidos como híbridos debido a que contiene las características de los sistemas de los que está compuesto.

6.4. Seleccionar el mejor modo de almacenamiento

En un *Data Warehouse* a gran escala, es crítico diseñar cubos *OLAP* para proveer una ejecución y optimización máxima. Al diseñar los cubos, se debe tomar decisiones importantes acerca del modo de almacenamiento de información y el nivel de agregación.

Las agregaciones son sumas precalculadas de datos de tablas de hechos con alguna combinación de niveles de cada dimensión y son usadas para responder consultas y crear agregados adicionales. Al escoger la cantidad (porcentaje) de agregación para incluir en un cubo, se debe considerar los problemas de espacio de almacenamiento y el tiempo de respuesta. Precalcular todos los posibles agregados incrementaría enormemente los requerimientos de almacenamiento para una base de datos. Por otra parte, calcular todos los agregados en el momento de la consulta resultaría en un tiempo de respuesta de consulta muy lento.

6.4.1. Tipo de almacenamiento

En pruebas realizadas con varios proyectos, se han seleccionado porcentajes de nivel de 0, 30, 60 y 90, aunque la mayoría de aplicaciones usan niveles de porcentaje de agregación entre 30-60, los valores 0 y 90 son incluidos para proveer una comparación de datos. Hay que recordar que el número de porcentaje de agregación representa el mejoramiento esperado en la ejecución de consultas comparado con tener agregaciones precalculadas.

6.4.2. Tiempos de proceso para cada tipo de almacenamiento

Los resultados indicados a continuación se obtuvieron al procesar cubos estructurados idénticamente usando diferentes modos de almacenamiento de datos y niveles de agregación.

Los tiempos de procesamiento para cada tipo de almacenamiento serían:

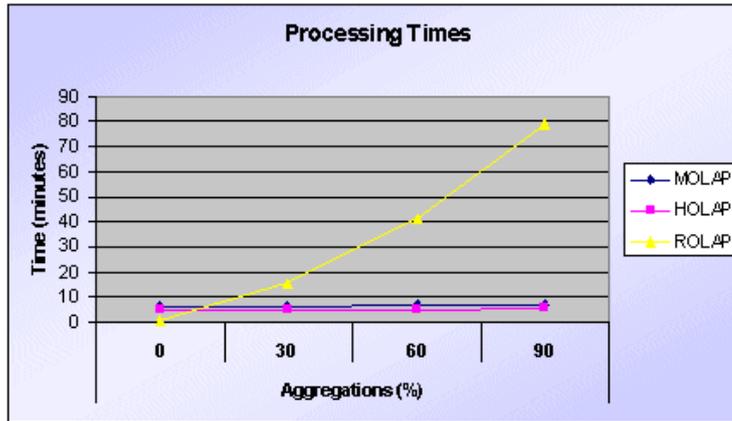


Gráfico 6.3. Tiempo de procesamiento para cubos ¹
www.msdn.microsoft.com/library "Analysis Services"

- En 0 por ciento de agregación, *ROLAP* toma un mínimo de tiempo para procesar el cubo. La data de la tabla de hecho y dimensión no está incluida en el cubo y no existen agregaciones precalculadas, sin embargo si el nivel de agregación incrementa, tomará más tiempo el procesamiento.
- La diferencia 30%-60% de agregación entre los cubos *MOLAP* y *HOLAP* no es substancial.
- Los tiempos de proceso se incrementa en 60% y 90% de agregación entre *MOLAP* y *HOLAP*, pero no es un incremento substancial.
- El tiempo de proceso se incrementa exponencialmente entre un porcentaje de agregación de entre 60 y 90 para *ROLAP*.

6.4.3. Requerimientos de espacio en disco para cada tipo de almacenamiento

La siguiente ilustración muestra los requerimientos de espacio (en *Megabytes*) para cada tipo de almacenamiento:

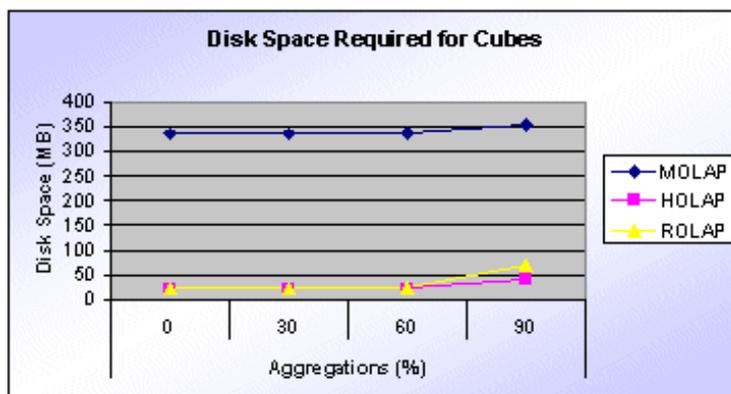


Gráfico 6.4. Espacio en Disco requerido para cubos ²
www.msdn.microsoft.com/library "Analysis Services"

- El almacenamiento *MOLAP* ocupa más espacio que *HOLAP* Y *ROLAP* ya que contienen una copia de los hechos y dimensiones originales.
- La diferencia entre 0%- 60% de ocupación no es substancial para *MOLAP* y *HOLAP*. Este se incrementa mientras que el nivel de agregación se aproxima a un 90%.
- EL almacenamiento *HOLAP* usa un espacio mínimo, debido a que no se tiene una copia de hechos originales y dimensiones en una base de datos *OLAP* y las agregaciones están almacenadas en la base de datos *OLAP* en un formato eficiente multidimensional.
- Al almacenamiento *ROLAP* requiere espacio adicional mientras el nivel de agregación exceda el 30% y mientras se aproxime al 90%. El espacio *ROLAP* mostrado incluye el espacio requerido para agregaciones en la base de datos relacional.

6.4.4. Requerimientos de disco para cubos *MOLAP* comparado con el Esquema estrella

Esta tabla compara la cantidad de espacio requerido por los cubos *MOLAP* al tamaño del esquema estrella original (tabla de hecho y tablas dimensionales) en el *RDBMS*.

Nivel de agregación (como porcentaje)	Espacio en disco usado por el cubo <i>OLAP</i>	Tamaño del esquema estrella (tablas de hechos y medidas con índices)	Porcentaje de compresión de datos hecho al construir los cubos <i>OLAP</i>
60	335.75	5188	93.53
90	353.11	5188	93.19

Tabla 5.1. Comparación de almacenamiento de espacio ³
www.msdn.microsoft.com/library "Analysis Services"

El espacio tomado por los cubos *MOLAP* en comparación al tamaño de las tablas del esquema estrella original es de aproximadamente el 7%. Incluso al 90% de nivel de agregación, logra aproximadamente la misma cantidad de compresión de datos. Generalmente se comprime en lugar de extender *data*. El espacio adicional requerido para construir cubos *MOLAP* varía con el número de niveles en la dimensión, el número de medidas y el tipo de *data*. ⁴

6.5. Operaciones OLAP

Una consulta a un almacén de datos consiste generalmente en la obtención de medidas sobre los hechos parametrizadas por atributos de las dimensiones y restringidas por condiciones impuestas sobre las dimensiones.

Por ejemplo si se desea conocer el consumo mensual de tráfico saliente por tipo de destino en una central determinada. La medida es el consumo mensual, el tráfico saliente es el hecho, la restricciones es el consumo durante un mes y el parámetro de consulta es el tipo de destino, central. ⁵

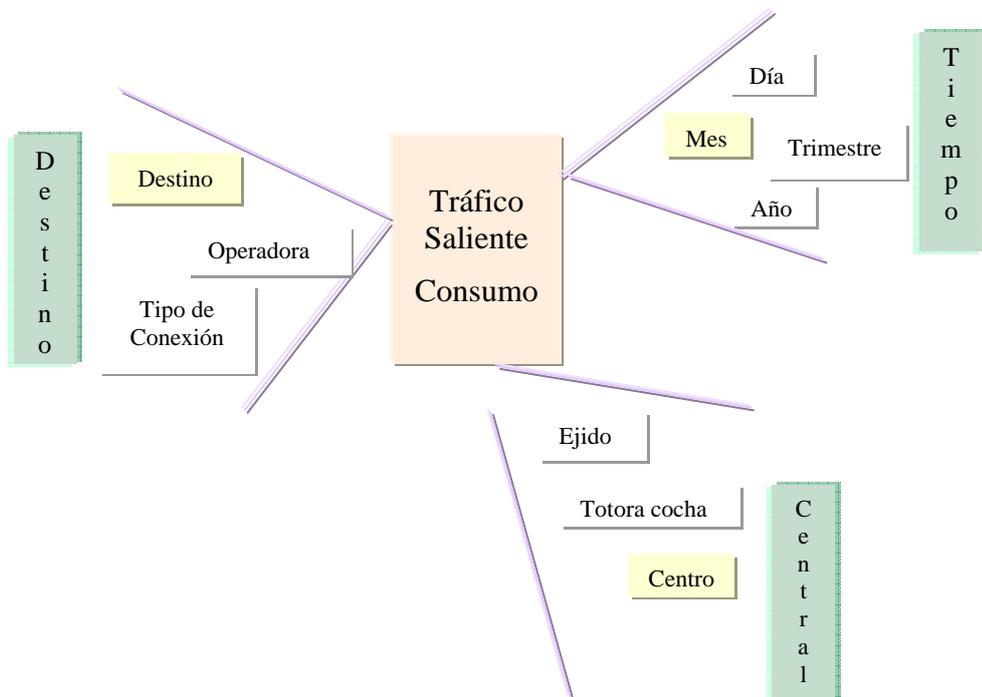


Gráfico 6.5. Consulta OLAP

Normalmente es posible realizar consultas que, en cierto modo, pueden hacerse con selecciones, proyecciones, concatenaciones y agrupamientos tradicionales, pero en las herramientas OLAP lo realmente interesante son sus operadores de refinamiento o manipulación de consultas, entre los que se puede mencionar.

- *DRILL*
- *ROLL*
- *SLICE & DICE*
- *PÍVOT*

El carácter agregado de las consultas en el Análisis de Datos aconseja la definición de nuevos operadores que faciliten la agregación (consolidación) y la disgregación (división) de los datos:

Roll (agregación): permite eliminar un criterio de agrupación en el análisis, agregando los grupos actuales.

Drill (disgregación): permite introducir un nuevo criterio de agrupación en el análisis, disgregando los grupos actuales.

Las operaciones de disgregación (*Drill*) y agregación (*Roll*) se pueden hacer sobre:

- Atributos de una dimensión sobre los que se ha definido una jerarquía: ***Drill-Down, Roll-Up***.

Destino - Conexión - Operadora (Destino Salida)

Año - Trimestre – Mes - Día (Tiempo)

Drill Down: La operación *Drill Down* permite ir desde niveles resumidos a niveles detallados de la información y viceversa.

Roll Up: Esta operación es inversa a la anterior va desde niveles detallados a resúmenes de información. Muestra el mayor nivel de agregación en la dimensión.

- Sobre dimensiones o cubos independientes : *Drill-Across, Roll-Across*

Drill Across: La operación *Drill Across* permite moverse entre cubos, es decir, incorpora ó cruza más de una tabla de hechos. Permite moverse lateralmente a otro cubo que comparte dimensiones.

Drill-Through: Se usa para ir a los datos en el *DWH* a partir de los cuales se calcula una celda del cubo. Permite obtener datos almacenados en el *DWH* para profundizar un análisis sin necesidad de que estén incluidos en la base *OLAP*.

Slice & Dice: Esta operación selecciona y proyecta datos en el informe.

Pivot: La operación *pivot* reorienta las dimensiones en el informe. Selecciona el orden de visualización de las dimensiones.

6.6. Conclusión

Después de conocer lo que es *OLAP* y los diferentes modos de almacenamiento estamos en capacidad de decidir sobre la forma como se llevará la información almacenada en el *Data Warehouse*.

De entre los diferentes tipos de almacenamiento, en este trabajo de tesis, se ha seleccionado *MOLAP*, debido a que la información que se extrae proviene de un cubo prediseñado. Aunque existiría la ventaja de ahorrar espacio en disco al utilizar otro modo de almacenamiento, hemos considerado como factor importante los tiempos de respuesta de las consultas debido a la gran información que manejan los cubos de tráfico telefónico (Políticas de la Empresa). Sin embargo las operaciones *OLAP* han permitido el poder obtener de los cubos la información sobre indicadores y reportes importantes, necesarios en las áreas donde será utilizado el *Data Warehouse*.

REFERENCIAS

^{1,2,3} MICROSOFT CORPORATION. *Optimizing Cube Performance Using Microsoft SQL Server 2000 Analysis Services*. <http://www.msdn.microsoft.com/library>. *Analysis Services*. 2001, [consulta 15 de abril de 2006]

⁴ IBM. *Sistemas OLAP*. Ibarra María de los Ángeles. http://www.exa.unne.edu.ar/depar/areas/informatica/sistemasoperativos_olap.pdf. 2005, [consulta 15 de abril de 2006]

⁵ DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN UNIVERSIDAD POLITÉCNICA DE VALENCIA, Almacenes de Datos. Hernández Orayo José. http://www.dsic.upv.es/~jorayo/cursoDWDWM/dwdm_11.pdf, [consulta 24 de abril de 2006]

CAPITULO 7
ENTREGABLES MSF

7.1. Fase I: Documento visión/alcance

Conocimiento del área del tráfico telefónico

7.1.1. Reuniones de levantamiento (visión global fase I)

Fecha:	27/10/2005
Ing. Miguel Ángel Zúñiga	Aseguramiento de Ingresos
Ing. Juan Pablo Bermeo	Aseguramiento de Ingresos
Fecha:	31/10/2005
Ing. Juan Pablo Astudillo	Estación Terrena
Fecha:	14/11/2005
Ing. Vicente Pinos	Planificación
Eco. Zaida Gutiérrez	Planificación
Fecha:	16/11/2005
Ing. Luís Espinoza	Conmutación
Ing. Fernando Vélez	Conmutación
Fecha:	17/11/2005
Ing. Luís Espinoza	Conmutación
Ing. Fernando Vélez	Conmutación
Ing. Miguel Ángel Zúñiga	Aseguramiento de Ingresos
Ing. Cristina Carpio	Sistemas
Ing. Fernando Uyaguari	Sistemas
Fecha:	20/12/2005
Ing. Miguel Ángel Zúñiga	Aseguramiento de Ingresos
Ing. Juan Pablo Bermeo	Aseguramiento de Ingresos
Ing. Fabiola Redrován	Coordinación con otras operadoras
Dr. Diego Ponce	Gerencia de Telecomunicaciones
Fecha:	26/12/2005
Ing. Fernando Vélez	Conmutación
Ing. Luís Espinoza	Conmutación
Ing. Paúl Mancheno	Conmutación

7.1.2. Introducción

En la primera fase del proyecto *Data Warehouse* referente al Visionamiento y Alcance se genera el presente entregable que especifica el “Conocimiento del Área del Tráfico Telefónico”, el mismo que incluye las reglas del negocio que permitirán interpretar de mejor manera la información, así como la identificación de las variables cuantitativas y cualitativas que representa la información que fluye en los procesos del área mencionada y el conocimiento de los indicadores que agregan valor a la empresa.

Se han identificado las actividades de la siguiente manera:

1. Identificación de los factores claves de éxito del proyecto:
 - a. Requerimientos Generales.
 - b. Riesgos del Proyecto.
 - c. Información necesaria para la toma de decisiones.
2. Identificación de procesos funcionales del negocio de Telecomunicaciones para el módulo de Tráfico telefónico:
 - a. Procesos Generales de Conmutación.
3. Identificación de los datos involucrados en los procesos de Tráfico Telefónico:
 - a. Requerimientos Generales.
4. Identificación de expectativas y necesidades:
 - a. Visión, Misión del Proyecto
 - b. Requerimientos Generales
 - c. Reportes
 - d. Fuera del alcance
5. Identificación de los usuarios del sistema de información, según lo definido por E.T.A.P.A.:
 - a. Perfiles de Usuarios

Además el documento contiene otra información de importancia.

7.1.3. Oportunidad del negocio.

7.1.3.1. Descripción de la oportunidad

E.T.A.P.A., entidad Pública Municipal ubicada en la ciudad de Cuenca, cuenta con dos principales líneas de negocio:

- Agua Potable (Agua Potable, Alcantarillado y Saneamiento)
- Telecomunicaciones (Telefonía pública, Telefonía fija y Banda Ancha).

E.T.A.P.A. brinda y pone a disposición sus servicios con la más completa infraestructura, exigencia de calidad, atención responsable a los casi 500.000 habitantes, ubicados tanto en áreas urbanas como rurales del Cantón Cuenca. La misma procura permanentemente la renovación tecnológica orientada hacia la excelencia en el servicio, a fin de cumplir con su compromiso de lograr una óptima rentabilidad en la operación, que permita contribuir al desarrollo económico y social del país. Para que el proyecto *Data Warehouse* de E.T.A.P.A. tenga éxito, es necesario conocer hacia donde va la Institución, por lo que, a continuación se detalla la visión y misión de la misma.

Visión: Ser una empresa modelo en el ámbito nacional en la prestación de servicios públicos de telecomunicaciones, agua potable y saneamiento, complementarios y conexos; con los más altos niveles de calidad e innovación de sus productos y servicios, buscando la fidelidad de sus clientes.

Misión: Garantizar la prestación de los servicios de Agua Potable y Telecomunicaciones en Cuenca, con responsabilidad ambiental, calidad, honestidad y vocación de servicio, como proyectarse a nivel nacional.

Para cumplir con la misión de la empresa dentro del plan estratégico se han incluido planes de acción tácticos, uno de los cuales es contar con una base de información única que permita la mejor toma de decisiones que resuelva los problemas que se presentan al recolectar la información produciendo inconsistencias.

7.1.4. Planteamiento del problema

La empresa “E.T.A.P.A.” actualmente no cuenta con una base de información única centralizada por lo que los reportes pueden ser extraídos e interpretados de distintas formas, además en el área de Conmutación no se cuenta con un sistema que permita la elaboración de reportes.

7.1.5. Visión del proyecto

El proyecto está orientado a construir un repositorio central de información de Tráfico Telefónico con la aplicación e implementación de la metodología y tecnología *Data Warehouse*, con el objetivo de contar con herramientas para un eficiente control y monitoreo, así como, proveer información a las áreas de Planificación, Aseguramiento de Ingresos y Conmutación de forma oportuna, ágil y confiable para la toma de decisiones eficientes y productivas.

Para conseguir el objetivo de esta primera fase de acuerdo con la visión del proyecto, se han establecido reuniones con las personas encargadas de los procesos funcionales de Conmutación, con quienes se analiza la información referente al tráfico telefónico y usuarios claves para conocer la información que fluye en dichos procesos, los reportes que son utilizados a diario, así como los indicadores.

7.1.6. Concepto de la solución.

Para el desarrollo e implementación del *Data Warehouse* se hace uso de *Microsoft Solution Framework (MSF)* que constituye una serie de modelos flexibles e interrelacionados que pueden guiar a una organización en la disposición de los recursos, gente y técnicas requeridas para crear una infraestructura de tecnología en línea, con objetivos de negocios. Es posible utilizar *MSF* junto con sus propias herramientas y técnicas para planear y administrar un proyecto. Puede constituir un complemento junto con el Proceso *Rational* (Proceso Unificado) para el modelamiento del sistema (*UML*).

Es importante recalcar que se contará con una plataforma tecnológica (Herramientas *ETL Microsoft*) que realizará los procesos de extracción de información desde diferentes fuentes en forma eficiente y en tiempo real, que tenga la capacidad de transformar, homologar y consolidar dicha información en un almacén de datos.

También se hará uso de la Herramienta *Analysis Services* (extracción desde la bodega de Datos) para la construcción de cubos.

La plataforma mencionada anteriormente, proporcionará al personal de sistemas de E.T.A.P.A., un ambiente integrado y simplificado para construir y administrar la bodega de datos y los *Datarmarts*, mientras facultan a los analistas, y usuarios de la organización para obtener la información correcta en el momento adecuado para mejorar la toma de decisiones y hacerlo en el menor tiempo posible.

7.1.7. Metas, objetivos, asunciones, y limitaciones

7.1.7.1. Metas, objetivos, asunciones

- Cliente Satisfecho: cumplir las expectativas del cliente, de acuerdo a los parámetros mencionados en la propuesta correspondiente.
- Incrementar eficacia del usuario: contar con información oportuna, a través de una herramienta amigable, el usuario dedicará mayor tiempo al análisis de información y no a la obtención y preparación de la misma.
- Mejorar los procesos de: análisis de información, rentabilidad de los equipos telefónicos, evitar saturación de la red, análisis para el dimensionamiento, análisis de la calidad de servicio, análisis de la calidad de circulación de tráfico, análisis del tráfico de abonados.
- Contar con una herramienta eficiente para toma de decisiones.
- Proveer información relevante, oportuna, periódica, confiable, y veraz a usuarios internos, que permita ir cumpliendo la estrategia de la organización.
- Optimizar la información del área del Tráfico Telefónico de E.T.A.P.A.
- Liberar tiempo en la generación de información.
- Integración de la información del área del tráfico telefónico en una sola fuente de información.

- Incrementar la productividad y eficiencia de E.T.A.P.A.
- Acceso fácil, rápido y consolidado en el *Data Warehouse* por parte de usuarios.
- Brindar información consistente a los usuarios, esto es que los resultados de las consultas sean interpretados de la misma manera.

7.1.7.2. Limitaciones

- Falta de colaboración y disponibilidad por parte de quienes laboran en las áreas involucradas.
- Desconocimiento de reglas de negocio.
- Desconocimiento de procesos de las áreas involucradas.
- Falta de una base de datos operacional de la cual se pueda extraer la información necesaria.
- Falta de especificación puntual de requerimientos por parte de quienes analizan la información.

7.1.8. Análisis de uso

Uno de los objetivos principales del proyecto motivo de este documento es que, los usuarios podrán obtener información a través de herramientas de fácil uso.

- *Accelerator for Excel*: herramienta es un complemento de Excel, y permite una navegación dinámica de la información, para la creación de reportes avanzados para el respectivo análisis.
- *Accelerator web*: permite tener reportes predefinidos, diseñados y validados por los usuarios, lo que hace que se pueda acceder dentro o fuera de la institución. De la misma manera, se podrá publicar reportes que los usuarios diseñen en el tiempo

7.1.9. Perfiles de usuario

Los usuarios que accederán a la información del área del tráfico telefónico son:

- Conmutación.
- Coordinación con otras operadoras.
- Planificación.
- Aseguramiento de Ingresos.

7.1.10. Requerimientos

7.1.10.1. Requerimientos de negocio

Tráfico Telefónico

De acuerdo a las reuniones llevadas a cabo con usuarios clave de las mencionadas áreas en el punto anterior se determinó en forma general los siguientes requerimientos de información:

Tráfico

En las áreas de Aseguramiento de Ingresos, Planificación y Coordinación con otras operadoras, se considera el tráfico saliente (Tráfico que sale de la red de E.T.A.P.A.) y el tráfico entrante (Tráfico que ingresa a la red de E.T.A.P.A.). Para el área de Conmutación se consideran los diferentes flujos de tráfico: Tráfico interno, Tráfico en tránsito, Tráfico de llegada, Tráfico de salida y Tráfico indefinido.

Destino

Se refiere a la cobertura de las llamadas es decir el alcance de la telefonía a las diferentes áreas: local, celular, regional, nacional e internacional.

Centrales

La central se refiere al lugar físico donde se genera el tráfico. En el área de Conmutación el tráfico puede ser analizado por:

Central:

A este nivel es posible obtener el porcentaje de tráfico en Erlangs y la cantidad de llamadas generadas en la Central: llamadas eficaces e ineficaces con sus causas.

URA:

Es la unidad de conexión del abonado. Están dentro de las centrales y puede existir una o más. A nivel de URA es posible obtener las llamadas presentadas y rechazos de llamadas.

Haces

Son los circuitos que enlazan las centrales, también nos indica por donde pasan las llamadas. Hay diferentes tipos de haces: Entrante/Saliente, Bidireccional ó Mixto

La información requerida sobre **Haces** es la siguiente:

- Carga
- Tráfico
- Tomas
- Rechazos de Tomas
- Rendimiento
- Ocupación
- Llamadas eficaces

Auxiliares

Son recursos de la red telefónica que sirve para dar tono cuando el abonado levanta el auricular, es decir permite realizar una llamada, la información requerida es:

- Intento de Tomas
- Rechazos
- Circuitos ocupados

Servicios especiales

En el área de Aseguramiento de Ingresos y Coordinación con otras operadoras (conciliación) es necesario conocer sobre los servicios especiales entre los cuales se distingue las llamadas con servicio 1800 #telefónico y servicio gratuito tres dígitos por ejemplo bomberos, policía, etc.

Es importante hacer una distinción de estos tipos de servicios ya que son parte de los destinos a donde van las llamadas.

7.1.11. Consideraciones importantes

- Es necesario estandarizar el nombre de las Operadoras para que la información pueda ser manipulada por las diferentes áreas.
- La información para el área de Coordinación con otras operadoras es necesaria tenerla mensualmente, pero dicha información pertenecerá al consumo (minutos) del mes anterior al actual.
- La información de tráfico telefónico para el área de Aseguramiento de Ingresos es necesaria diariamente y los resúmenes de la información tienen que ser detallados de forma mensual.
- El análisis del tráfico en el área de Aseguramiento de Ingresos se lo realiza por categorías.
 - Este análisis permite: La recategorización de las líneas telefónicas en base a una media obtenida sobre el número de contactos diferentes a los que se realiza las llamadas.
- Detectar fraudes ocasionados por un *Bypass*.
- Es necesario conocer la población proporcionada por el INEC, para el cálculo de la densidad.
- La información proporcionada por la Central está basada en muestras que se realizan en una hora específica.
- Toda la información del área de Conmutación que se hará constar será la existente en la base de datos de Oracle, debido a que la información necesita un proceso de decodificación (*EBCDIC – ORACLE*), el mismo que es realizado por dicha área y no constituye una responsabilidad para los desarrolladores del *Data Warehouse*.
- La información de Conmutación sobre observaciones permanentes (Clases) que aparecerá en el *Data Warehouse*, dependerá de las actualizaciones y la correcta administración que se da a la base de datos de Oracle.
- El administrador del sistema transaccional en el área de Conmutación debe otorgar los permisos necesarios para la extracción de los datos en forma mensual. Para evitar este problema unos días antes de la fecha prevista a la otorgación de los permisos se enviará un mensaje a través del correo electrónico.

- No se incluye el negocio de agua potable y alcantarillado.
- No se incluye negocio de Internet.

En cuanto a las observaciones temporales (no disponibles en la base de datos al momento) sobre el tráfico en Erlangs para las URAs y lo referente a Destinos del tráfico, se hará constar siempre y cuando este disponible la información a la fecha en que se comiencen a realizar las extracciones. De no estar disponible en la base de Datos Oracle dichas observaciones, no se incluirán en el *Data Warehouse*.

7.1.12. Procesos funcionales: análisis tráfico telefónico.

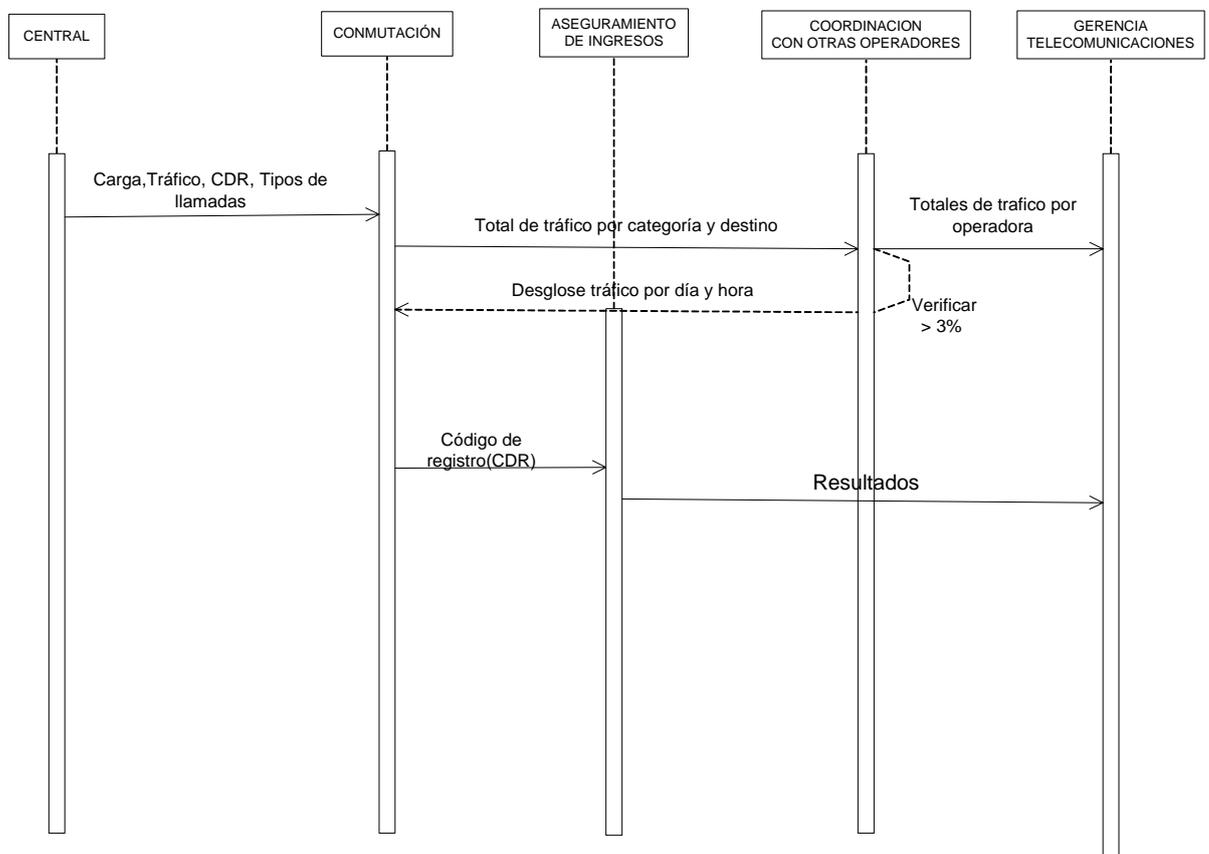


Gráfico 7.1 Proceso funcional Tráfico Telefónico

7.1.13. Procesos de conmutación

La Central es donde se origina la información a través de los procesos llevados a cabo por las observaciones temporales, permanentes y diversas. A partir de dichas

observaciones efectuadas durante la hora cargada (se realiza un análisis una vez al año durante 5 días laborables consecutivos, estableciéndose una hora durante la cual el tráfico medio es el máximo del día), es factible obtener información en “bruto” sobre: **Uras, Haces, Auxiliares e información general a nivel de Central.**

Esta información es almacenada en Contadores Brutos de Carga y Totalizadores, los mismos que posteriormente son tratados por el OM (Software de Explotación y Mantenimiento) y son “repartidos” a los **Contadores Observados:**

Para Observaciones Permanentes: CHO Contador Horario Observado.
CTO Contador Totalizador Observado.
CRO Contador de Resultados Observado.
CCO Contador de Carga Observado.

Para Observaciones Temporales: CPO Contador de Observaciones Temporales.

A más de la información en los contadores brutos, se cuenta con los Mensajes Espontáneos.

La Red de Conexión esta conformada por Máquinas Lógicas las mismas que desempeñan diferentes funciones entre las cuales se puede mencionar como principales: Tasación, Control de Comunicación, Traducción, Distribución de Mensajes, Gestión de Equipos y tonos de Auxiliares. Estas máquinas también se encargan de enviar Mensajes Espontáneos al OM, el mismo que elabora los resultados de las Observaciones Temporales.

La información de los Contadores Brutos y los Contadores Observados están repartidos en 10 clases de acuerdo al dominio de observación. Mediante la utilización de comandos la información almacenada en los Contadores Observados es editada de dos formas:

- En formato Terminal
- En formato Artículo

- En formato Terminal los resultados se muestran en un Terminal tras una interrogación o bien en un Terminal asignado por un operador a un fichero lógico.
- En formato Artículo se permite transmitir los resultados en un enlace de datos y salvaguardar los resultados en una cinta magnética.

Luego de la edición, se puede obtener información de Ficheros Lógicos los mismos que se decodifican de acuerdo a estándares y son almacenados en una base de datos ORACLE, para su posterior utilización y análisis en el área de conmutación. La información proporcionada en las áreas de Aseguramiento de Ingresos, Coordinación con otras operadoras y Planificación, se obtiene de un archivo generado en la central, dicho archivo está en formato E.B.C.D.I.C. (Extended Binary Coded Decimal Interchange Code) y es independiente de las observaciones y los contadores. Este archivo es decodificado y tratado por un operador y luego es pasado a una Base de Datos Oracle en el área de Conmutación. Dicho archivo contiene los Registros de las llamadas realizadas o llamadas eficaces. Estos Registros reciben el nombre de **CDR (código de registros)** y son utilizados para el análisis de información de las áreas ya mencionadas y posteriormente son emitidos a niveles altos de la empresa para la toma de decisiones.

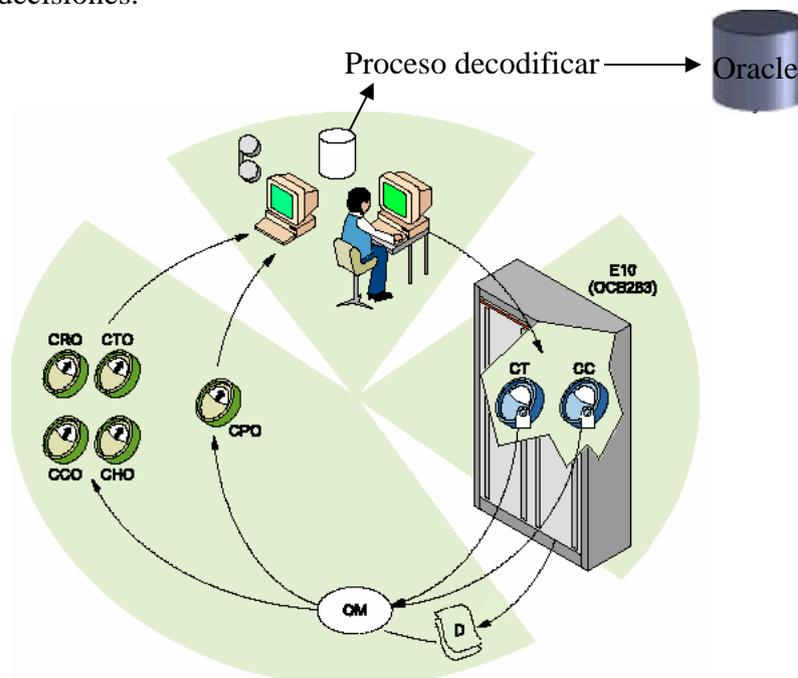


Grafico 7.2 Proceso de Conmutación ¹

7.1.14. Variables necesarias

7.1.14.1. Telefonía

- Densidad Telefónica
- Duración promedio llamada (Derivado)
- N° contactos diferentes
- N° de líneas instaladas
- N° de llamadas
- N° líneas de servicio
- N° minutos consumidos
- N° minutos redondeados(segmento 6)

7.1.14.2. Conmutación

AUXILIARES

- Intento de tomas de auxiliares
- N° de tomas de auxiliares
- N° de tomas eficaces de auxiliares
- Rechazos de tomas de auxiliares
- N° medio de auxiliares ocupados
- N° medio de auxiliares en servicio
- N° de auxiliares equipados

TRAFICO ERLANGS

- Volumen tráfico entrante
- Volumen tráfico saliente
- Volumen tráfico interno
- Volumen tráfico en tránsito
- Volumen tráfico diverso

HACES

- N° de tomas de llegada en el haz
- N° de tomas de salida en el haz
- N° de tomas rechazadas en la entrada
- N° de tomas rechazadas a la salida
- N° Llamadas eficaces en entrada
- N° Llamadas eficaces en salida
- Tráfico de llegada
- Tráfico de salida
- N° Circuitos equipados del haz
- N° medio de circuitos en servicio del haz
- Doble toma

URA

- N° llamadas presentadas
- N° llamadas rechazadas

7.1.15. Reportes

Los reportes e indicadores que se requieren son los siguientes:

7.1.15.1. Telefonía:

- Tráfico entrante por: Operadora
Destino
Categoría
- Tráfico saliente por: Operadora
Destino (País /Ciudad)
Categoría
Tipo de Acceso
Llamadas al 1-800
- Altos consumos:
 - Quién realiza más llamadas a números diferentes (contactos), esta consulta debe estar parametrizada por ejemplo a más de 100 números diferentes (es como filtrar la información).
 - Qué contacto externo llama más a un abonado de E.T.A.P.A.
 - A qué abonado de E.T.A.P.A. se llama más.

7.1.15.2. Conmutación

Reportes para verificar tráfico en el haz:

- Mantenimiento del haz.
- Saturación de haces: unidireccionales y bidireccionales.
- Dimensionamiento del haz.
- Eficacia del haz: unidireccionales y bidireccionales.
- Índice de pérdida del haz: unidireccionales y bidireccionales.
- Rendimiento del haz: unidireccionales y bidireccionales.

Reportes para verificar índices de ineficacia

- Índice de fracaso interno del conmutador.
- Índice de fracaso debido al abonado llamado.
- Índice de fracaso debido al abonado llamador.

Otros reportes

- Volumen del Tráfico en Erlangs de la Central.

Nota: En la segunda fase del proyecto se podrá verificar si los reportes mencionados se presentarán o no. De ser posible se incluirán en el *Data Warehouse* otros reportes no mencionados.

7.1.16. Alcance del proyecto.

El proyecto está enfocado a construir un repositorio único, en donde se deposite toda la información de diferentes fuentes, para el área del tráfico telefónico, así como a proveer información tanto a usuarios, así como a otras aplicaciones.

Es importante resaltar que los alcances están dados de acuerdo a las fases respectivas. Con esto se quiere enfatizar que el alcance de la primera fase, motivo de este documento, es obtener un módulo para el área del tráfico telefónico y las variables primarias que constan en las bases que interactúan en el proceso.

7.1.17. Lista de funciones / características

Diseño e Implementación del módulo de Tráfico Telefónico:

- *Datamarts* correspondientes con sus propias variables:
 - Telefonía
 - Conmutación

7.1.18. Criterio de aceptación

Para la aceptación del producto, dentro de cada una de las fases que se incluyen en la metodología *MSF* se tendrá un documento o entregable, el cual será proporcionado al líder del proyecto (Ing. Fernando Uyaguari), a las áreas de Conmutación, Aseguramiento de Ingresos y Coordinación con otras operadoras: Ing. Luís Espinoza, Ing. Juan Pablo Bermeo e Ing. Fabiola Redrován respectivamente. Luego de sus análisis y revisión se realizarán los cambios necesarios para su aprobación. Una vez aprobado se continuará con la siguiente fase. Durante el desarrollo de todas y cada una de las fases se realizarán reuniones de seguimiento, de acuerdo al cronograma. Estas se realizarán cada 30 días para un seguimiento continuo. En estas reuniones se tendrá informado a todo el equipo de trabajo del desarrollo de la fase y su estado de manera que se haga más fácil el entendimiento y comprensión de la fase del proyecto. Luego de cada una de las reuniones, el líder del proyecto debe enviar actas con los puntos y acuerdos tratados en dicha reunión.

7.1.19. Criterio operacional

La aceptación del producto tiene que estar dada por la aceptación de los diferentes roles participantes en el equipo de trabajo. Mediante la aceptación formal en el documento respectivo, se confirmará que las personas indicadas realizaron las pruebas necesarias y de esta manera confirmar que la solución está lista para liberarse y ponerse en producción.

7.1.20. Matriz de riesgos del proyecto

Recopilación Metadata Sistemas				
Consecuencia	Si este proceso no es confiable entonces la información que brinde la solución no será confiable. No se podría implementar correctamente el proyecto.			
Plan Mitigación	1) Compromiso del usuario para confirmar los campos de la base. 2) Compromiso de desarrollo de sistemas para comprobar los datos.			
Plan Contingencia	Ninguna			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Alta	Alta	80%	Desconocido	
Responsable	E.T.A.P.A.- Desarrolladores(tesis)			

Incumplimiento de Cronograma				
Consecuencia	Retrasos en la finalización del proyecto			
Plan Mitigación	Compromiso de ambas partes durante el desarrollo y seguimiento del proyecto			
Plan Contingencia	Ninguna			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Media	Alta	30%	Desconocido	
Responsable	E.T.A.P.A. – Desarrolladores(tesis)			

Cambio del Personal				
Consecuencia	Atraso en el cronograma por demora en conocer las tareas que se realizan en el área por parte del nuevo personal, así como lo referente al proyecto.			
Plan Mitigación	Solicitar se asigne por lo menos a dos personas para brindar información.			
Plan Contingencia	Soportar alargues en el proyecto.			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Media	Alta	50%	Desconocido	
Responsable	E.T.A.P.A.			

Confiabilidad de la Información				
Consecuencia	El usuario no tendrá disponible información adecuada para el análisis y la toma de decisiones.			
Plan Mitigación	El modelo proporcionará información que no tiene una asignación o codificación eficiente. Estos datos deberán ser reasignados o modificados por el usuario en el sistema fuente.			
Plan Contingencia	Ninguna			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Alta	Alta	35%	Desconocido	
Responsable	E.T.A.P.A.- Desarrolladores(tesis)			

Desconocimiento de Herramientas				
Consecuencia	Atrasos en tareas asignadas			
Plan Mitigación	Investigar			
Plan Contingencia	Investigar en manuales ó consultar a expertos			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Media	Media	50%	Desconocido	
Responsable	E.T.A.P.A. – Desarrolladores(tesis)			

Desconocimiento de las Tablas				
Consecuencia	Retraso en la identificación de fuentes			
Plan Mitigación	Solicitar documentación con especificación de campos			
Plan Contingencia	Solicitar información a los desarrolladores o a quienes manipulan la información.			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Media	Alta	30%	Desconocido	
Responsable	E.T.A.P.A.- Desarrolladores(tesis)			

Procesos de Carga de Datos				
Consecuencia	Usuarios no analizarán la información completa.			
Plan Mitigación	Confirmar que toda la información requerida en la visión sea cargada.			
Plan Contingencia	Definir una o varias personas que analicen los resultados.			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Media	Media	10%	Desconocido	
Responsable	E.T.A.P.A.- Desarrolladores(tesis)			

Utilización de Servidores				
Consecuencia	No poder realizar los procesos de extracción, transformación, etc.			
Plan Mitigación	Solicitar la autorización de otro servidor.			
Plan Contingencia	Utilizar un servidor local.			
Probabilidad	Importancia	Peso	Estado	Fecha de Revisión
Media	Alta	20%	Desconocido	
Responsable	E.T.A.P.A.- Desarrolladores(tesis)			

7.2. Fase II: Especificación y diseño

7.2.1. Diseño de la base dimensional (Modelo Lógico, Físico)

7.2.1.1. Introducción

El presente documento corresponde al entregable “Diseño de la Base Multidimensional”, de la fase de Especificación y Diseño, de la Propuesta de Valor para el desarrollo del Módulo de Tráfico Telefónico de E.T.A.P.A.

En este documento se define el diseño lógico y físico de la base de *Datos Warehouse*, así como también la estructura de la base multidimensional.

7.2.1.2. Diseño lógico

Para representar el esquema lógico se utilizó el modelo dimensional “Esquema Estrella”, ya que aumenta la velocidad de ejecución de consultas debido a que la información se encuentra desnormalizada y no necesario los joins (uniones) entre varias tablas. Es de fácil interpretación para las personas de negocios no técnicas y el modelo es sencillo a pesar de que el área de negocio es compleja.

Es una política de la empresa desarrollar el *Data Warehouse* en base a la metodología y recursos ya establecidos: el modelo estrella y las herramientas de inteligencia de negocios de *SQL Server 2005*. Es fundamental mencionar que el repositorio único de información se establecerá sobre la base *SQL Server 2005*.

7.2.1.2.1. Tráfico telefónico

La información del tráfico telefónico para ser correctamente analizada se ha dividido en dos partes: la primera corresponde al tráfico puro es decir la generación de las llamadas y la otra parte es referente a la calidad de servicio provista a los abonados y la calidad de la circulación del tráfico. Debido a esto, lo que corresponde a las llamadas (CDR's) se ha subdividido en tráfico entrante y tráfico saliente.

Algo importante que se debe considerar es la retrospcción de cada uno de los objetos que conforman el *DWH*, pero de aquellos que agreguen valor al negocio por está razón se considera la retrospcción verdadera en los casos de la categoría y el nombre del cliente (Tráfico Telefónico), por lo que se incluye en la tabla de hechos como llaves foráneas y pertenecen a la dimensión cliente (abonado), ya que es necesario contestar a las siguientes preguntas: ¿Cuántos clientes hay en una categoría en el mes de Enero del 2005?, ¿Quién realizó las llamadas en un determinado período?, ¿Cuántos minutos se han consumidos por categorías en el mes de Febrero del 2005 y Febrero del 2006? etc. Además se debe considerar la disponibilidad de información y cuando la información está lista para ser poblada al *Data Warehouse*. Esto es importante porque se debe hacer procedimientos *ETL* diferentes, aunque esto varía de acuerdo a las necesidades de la empresa.

7.2.1.2.1.1. Tráfico entrante

Tráfico entrante corresponde a todas las llamadas que provienen de otras redes (operadoras) hacia la red de E.T.A.P.A. La información que se desea analizar es el volumen del tráfico desde los diferentes destinos: nacional, regional, internacional, celular así como las operadoras que se utilizaron para la comunicación y en que fecha se realizaron. Adicionalmente se necesita conocer quien realiza las llamadas y quien las recibe, etc. Para obtener esto se ha planteado el siguiente diseño:

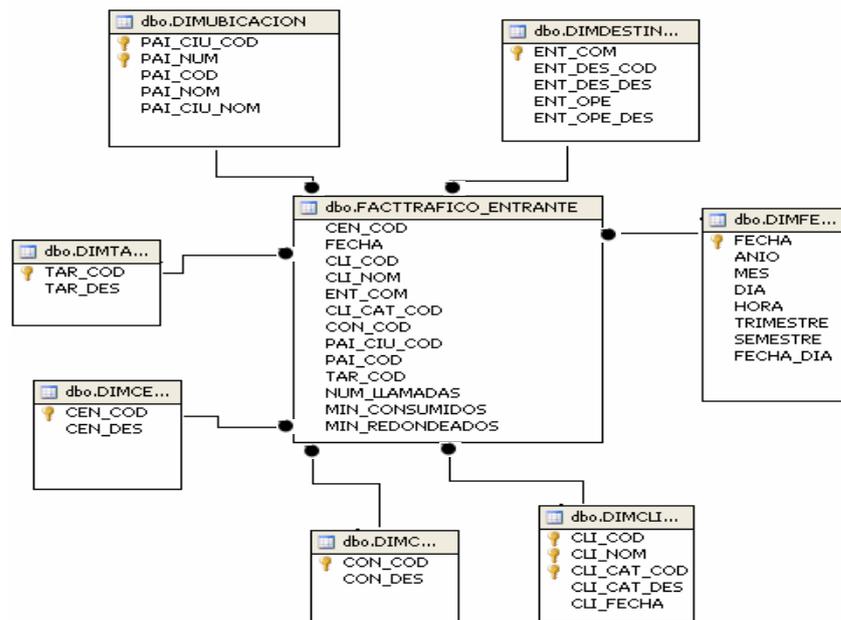


Gráfico 7.3. Esquema Estrella Tráfico Entrante

DIMUBICACION: Hace referencia al lugar de donde proviene la llamada, que es recibida por el abonado de E.T.A.P.A.

DIMTARIFA: Hace referencia al tipo de tarifa Normal/Reducida con que se recibe la llamada, generalmente desde los celulares.

DIMCENTRAL: Se refiere al lugar físico donde se genera el tráfico.

DIMDESTINO_ENTRADA: Hace referencia a la cobertura de la llamada esto es de donde se recibe la llamada (nacional, regional) y de la operadora que se utilizó para la interconexión con E.T.A.P.A.

DIMFECHA: Se refiere a la fecha en la que recibe la llamada el abonado de E.T.A.P.A.

DIMCLIENTE: Se refiere al abonado de E.T.A.P.A. que recibe la llamada.

DIMCONTACTO_EXTERNO: Hace referencia al número telefónico quien realiza la llamada a un abonado de etapa.

FACTENTRANTE: Contiene las llaves foráneas que relaciona la tabla de hecho con las dimensiones y las medidas en este caso son:

Num_llamadas: Son el número de llamadas que se recibe (Agregación Suma).

Min_consumidos: Volumen del tráfico expresado en minutos (Agregación Suma).

Min_redondeados: Volumen del tráfico expresado en minutos. Se utiliza sólo para llamadas recibidas internacionales (Agregación Suma).

En la fase I se establecieron más medidas pero no se incluyen porque algunas se pueden calcular, es decir son atributos derivados como por ejemplo: número de contactos diferentes, duración promedio de la llamada. Otras medidas no se pueden obtener porque no existe la fuente como son la densidad telefónica, líneas instaladas. Con respecto al número de llamadas ineficaces esto no es posible obtener en el *datamart* de tráfico entrante, si no en el *datamart* de conmutación.

7.2.1.2.1.2. Tráfico saliente

El tráfico saliente corresponde a todas las llamadas que salen de la red de E.T.A.P.A. y las que se encuentran en la misma red por esta razón se ha considerado dos niveles de análisis: Las llamadas que utilizan operadora para establecer la comunicación (destino internacional, regional, 1800, etc.), y las que no utilizan operadora como es el tráfico local, el cual tiene dos niveles de detalle: tráfico local y local detallado.

7.2.1.2.1.2.1. Tráfico saliente con operadora (destinos)

Al igual que el tráfico entrante este comparte las mismas dimensiones excepto la DIM_DESTINO_ENTRADA que es reemplazada por la DIMDESTINO_SALIDA debido a que en el tráfico saliente es necesario conocer el tipo de acceso, es decir si la llamada que se va a realizar utilizó discado directo, tarjeta prepago o hizo uso de una operadora. Con respecto a las medidas son las mismas y tienen las mismas descripciones ya antes mencionadas. La diferencia en el tráfico entrante y saliente es que para el tráfico saliente quien realiza la llamada es el abonado de E.T.A.P.A., y quién recibe es el contacto externo, a diferencia del tráfico entrante quien realiza la

llamada es el contacto externo y quien recibe la llamada es el abonado de E.T.A.P.A., una vez aclarada esta situación en la identificación de fuentes es más fácil establecer la metadata.

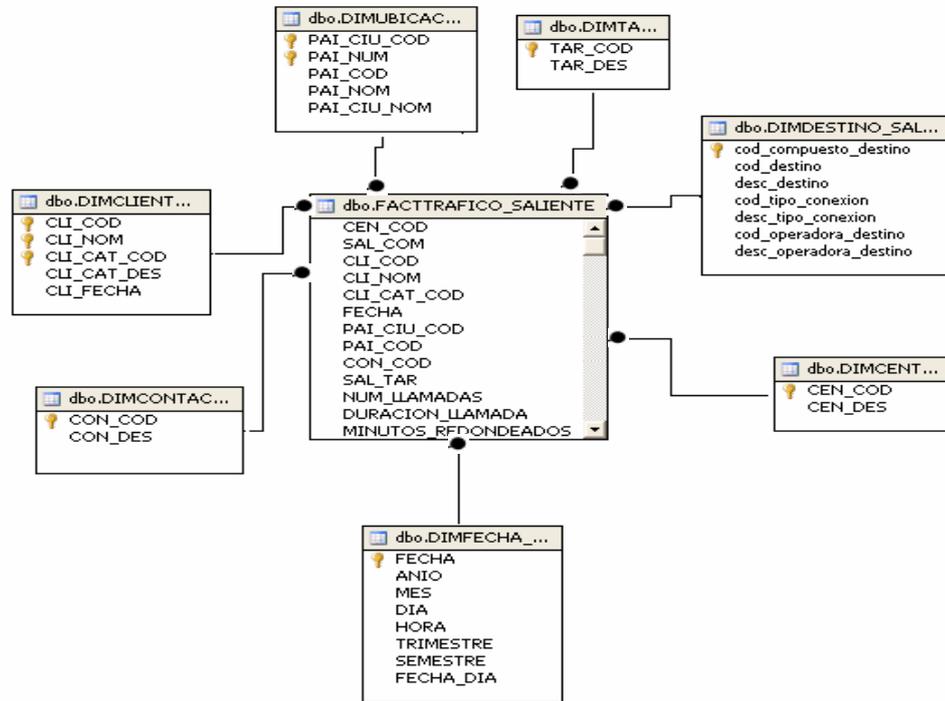


Grafico 7.4. Esquema Estrella Tráfico Saliente

7.2.1.2.1.2.2. Tráfico local

Al ser parte del tráfico saliente comparte las mismas dimensiones pero por obvias razones va a prescindir de algunas dimensiones, como por ejemplo ubicación, porque que el tráfico se realiza sólo a nivel del Cantón Cuenca y por esto también se elimina la cobertura o destino (nacional, regional) y el uso de operadora de interconexión debido a que no sale de la red de E.T.A.P.A. Sin embargo se aumenta una dimensión DIMTIPO_LOCAL, la misma que ayudará a clasificar la información de los dos tipos de llamadas locales es decir llamada entre abonado-abonado y abonado-servicio gratuito tres dígitos. Se hace esta aclaración porque para ser un abonado de E.T.A.P.A., su línea telefónica debe estar registrada y activa. Con respecto a las medidas que van en la FACTTRAFICO_LOCAL(tabla de hecho) se debe considerar que no hay como establecer el número de llamadas cuando la llamada sea de abonado – abonado porque no existe el transaccional a detalle, pero para la llamada abonado – servicio gratuito si existe detalle. En el tráfico local se desea conocer el consumo mensual de cada uno de los abonados y de los números registrados en la

central. Por otra parte el nivel de detalle para el tráfico local a servicios gratuitos, es por hora, por lo que el esquema estrella se plantea de la siguiente manera:

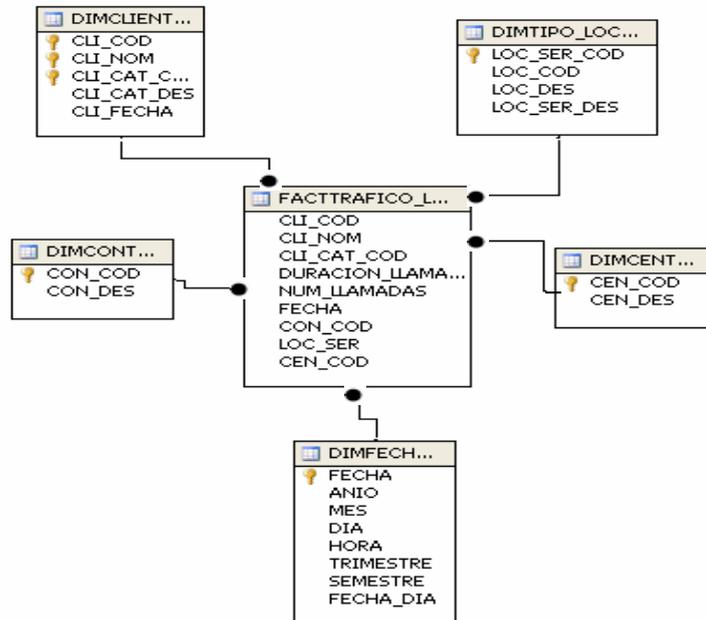


Gráfico 7.5. Esquema Estrella Tráfico local

7.2.1.2.1.2.3. Tráfico local detallado (abonados)

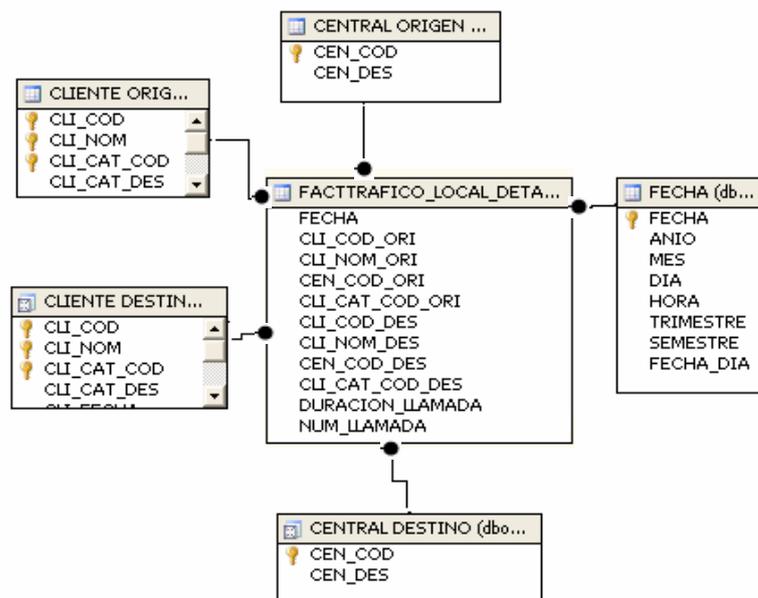


Gráfico 7.6. Esquema Estrella Tráfico local detallado

Se planteó este esquema estrella porque es necesario conocer el consumo local detallado por hora y para cumplir con los objetivos planteados en el entregable de visión. Comparte las mismas dimensiones pero se crean vistas de la central y clientes

porque quien realiza la llamada y quien recibe son abonados de etapa, pero este detalle no existe para todos los abonados sólo para algunos casos especiales. Las medidas tienen el mismo significado de las anteriores debido a que, tiene similitud todo el concepto de tráfico telefónico.

Se ha creado diferentes tipos de esquemas estrella por el contenido y la forma en la que se requiere analizar la información. Con todos estos esquemas estrella del tráfico telefónico (generación de llamadas) se puede consolidar la información en uno sólo, al cual se denomina cubo físico (SQL Server 2005) ó cubo virtual (SQL Server 2000).

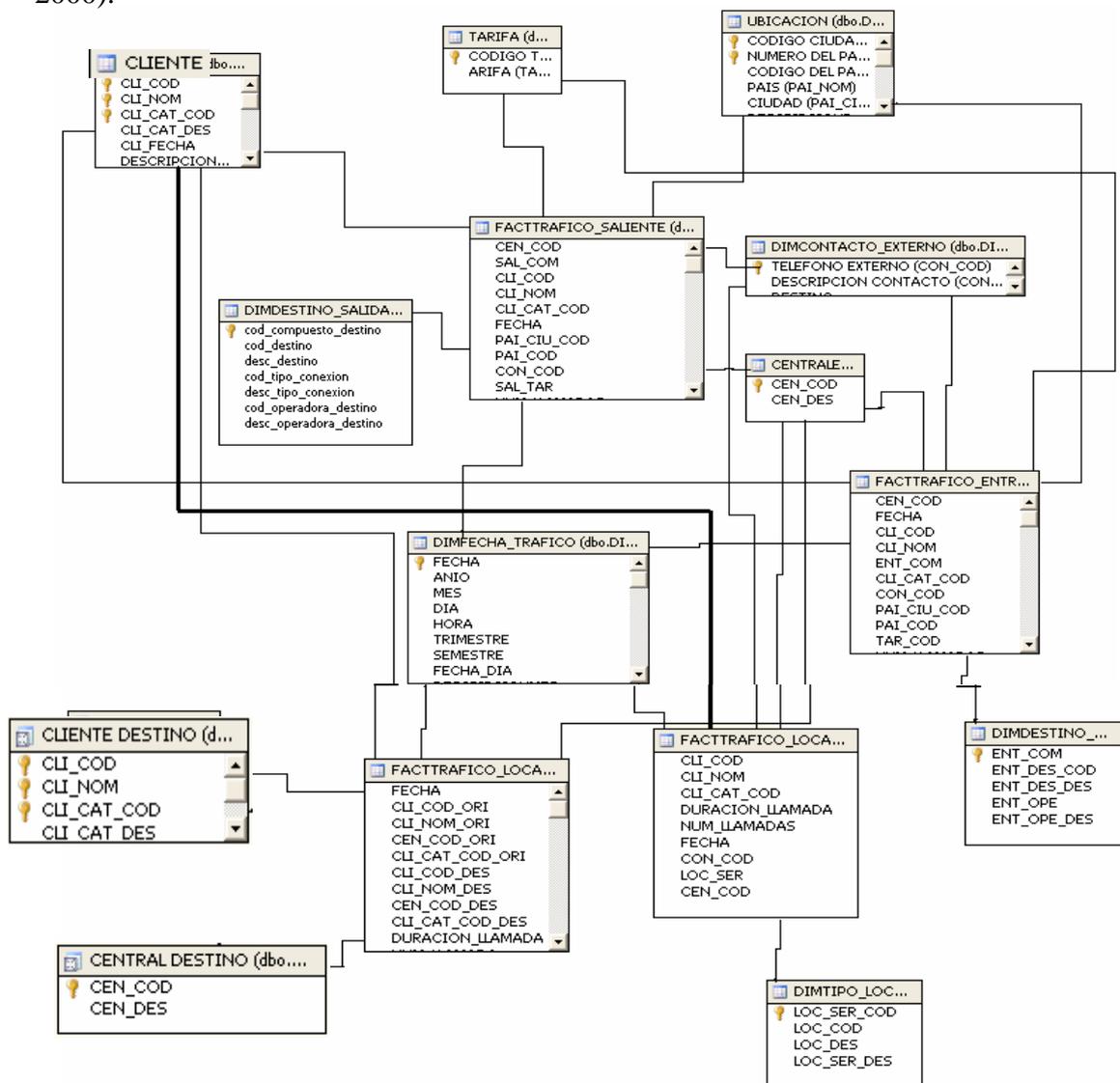


Gráfico 7.7. Esquema Estrella Tráfico Telefónico

El inconveniente de realizar esto es que cuando los usuarios finales no realicen correctamente los filtros, la información a generar tardará varios minutos.

Para evitar este problema es necesario hacer uso de las perspectivas que permiten analizar parte de la información del cubo que se basan en cada uno de los esquemas estrella antes mencionados.

7.2.1.2.2. Control de calidad de circulación del tráfico

El propósito de analizar la información generada por el sistema de conmutación ALCATEL es ayudar al operador a optimizar la rentabilidad de los equipos telefónicos, para evitar la saturación de la red y satisfacer al máximo al cliente y al nivel táctico ayuda a tomar decisiones en base a justificativos del comportamiento del tráfico telefónico, por esta razón se ha diseñado los siguientes esquemas:

Nota: Las dimensiones que son compartidas por todos los esquemas estrella son la dimensión central y la dimensión fecha, que a la vez son compartidas por el *datamart* tráfico telefónico

7.2.1.2.2.1. Auxiliares (Clase A)

Es un recurso que se utiliza para establecer la conexión y permite analizar el tráfico mediante tomas y rechazos de toma. Es el primer recurso a ser utilizado cuando se realiza una llamada. Para realizar su análisis se ha planteado el siguiente esquema estrella:

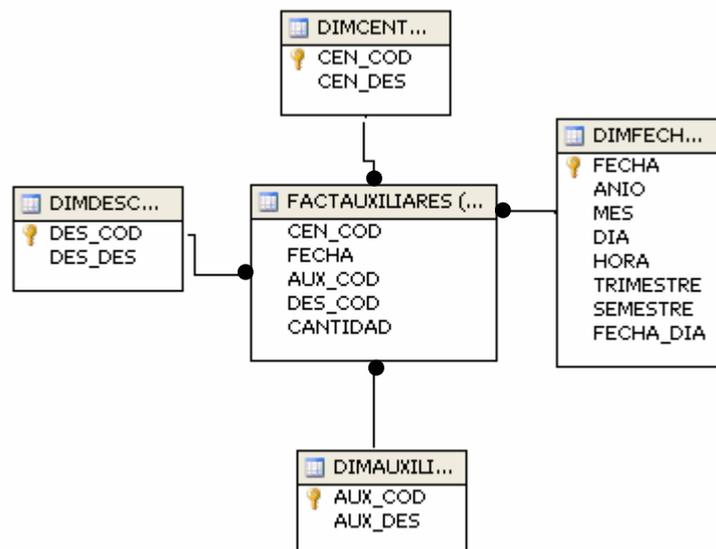


Gráfico 7.8. Esquema Estrella Auxiliares (Conmutación)

DIMDESCRIPCION_AUXILIARES. Contiene la descripción de cada una de las medidas que se presentan en los auxiliares. Por ejemplo: intento de tomas.

DIMAUXILIARES. Contiene los nombres de los auxiliares existentes en las centrales telefónicas.

FACTAUXILIARES. Contiene la medida cantidad. Se redujo a un sólo campo debido a que se hace un cruce con la dimensión que contiene la descripción de los auxiliares (DIMDESCRIPCION_AUXILIARES) para obtener los valores de las demás medidas, por esta razón es erróneo analizar la información si no está presente la dimensión dimdescripción_auxiliares, porque la agregación de la medida obtendrá resultados equivocados ya que diferentes medidas no pueden agregarse.

7.2.1.2.2.2. Tráfico erlangs (Clase C)

Erlangs es una unidad de medida del tráfico. Es importante para definir la hora cargada en las diferentes centrales y realizar otros análisis ya especificados en los entregables de visión y alcance. Se establece el siguiente esquema estrella:

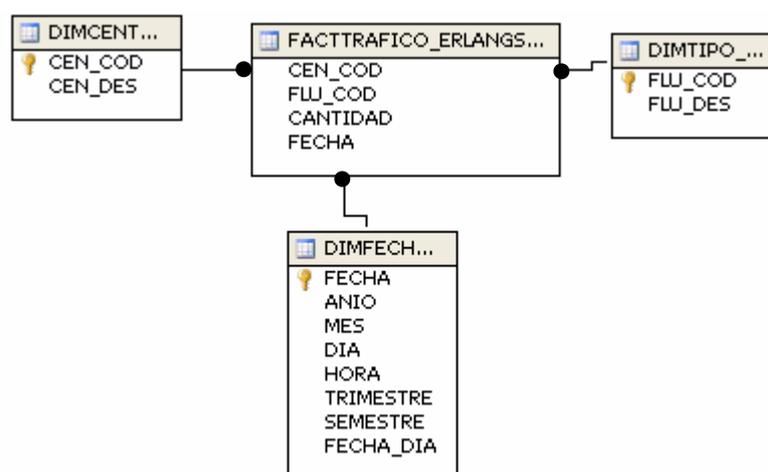


Gráfico 7.9. Esquema Estrella Tráfico Erlangs (Conmutación)

DIMTIPO_FLUJO. Contiene la información de los tipos de flujo que existen en una central. Sirve para detallar más la información del tráfico.

FACTTRAFICO_ERLANGS. Contiene una sola medida llamada cantidad que representa el tráfico cursado en la central. Igualmente se reduce a una sola medida ya que en los tipos de flujo está la descripción de las medidas, pero en este caso no es necesario que la dimensión tipo flujo de tráfico este siempre presente ya que la

medida cantidad puede agregarse sin necesidad de esta dimensión, sin causar resultados errados. Por ejemplo si se necesita saber el tráfico total cursado en la central o en una determinada fecha.

7.2.1.2.2.3. Haces (Clase FB)

Los haces son un recurso que se utiliza para establecer la comunicación entre operadoras o centrales. Se desea analizar el tráfico cursado y además cumplir con los requisitos establecidos (reportes). Se plantea el siguiente esquema estrella:

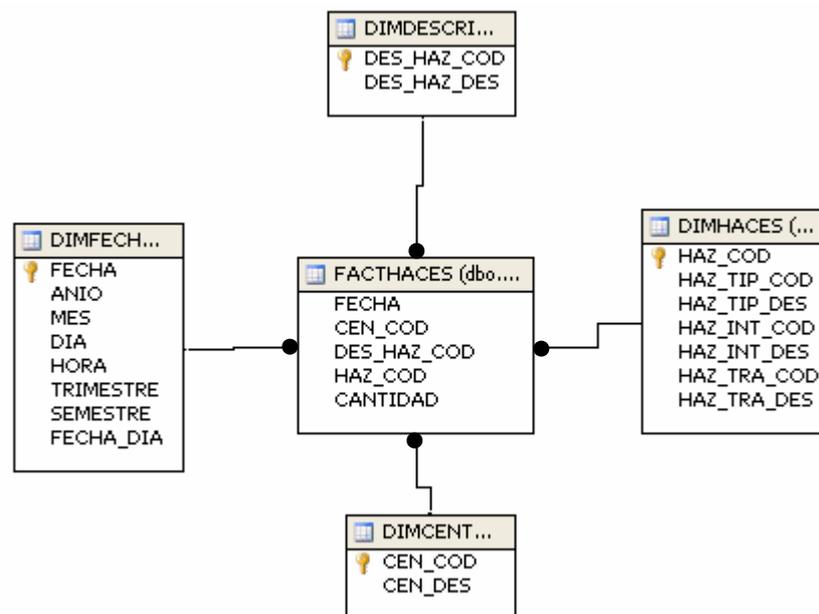


Gráfico 7.10. Esquema Estrella Haces (Conmutación)

DIMHACES. Contiene la información de los tipos de haces, tipo de tráfico que cursa y la interconexión entre centrales u operadoras.

DIMDESCRIPCION_HACES. Contiene la información de las descripciones de las medidas de los haces para poder analizar su comportamiento, tiene que estar presente esta dimensión en el análisis para evitar resultados erróneos.

FACTHACES. Contiene una sola medida llamada cantidad que representará: tráfico, ocupación en los haces, etc. Por esta razón debe estar presente la dimensión dimdescripcion_haces para evitar resultados errados.

7.2.1.2.2.4. URA (Clase R)

La Unidad de Conexión del Abonado (URA) es un concentrador que puede ser local o distante. Varias URAs forman una central y los abonados están conectados a una URA. Para cumplir con la visión de la solución se ha planteado el siguiente esquema estrella:

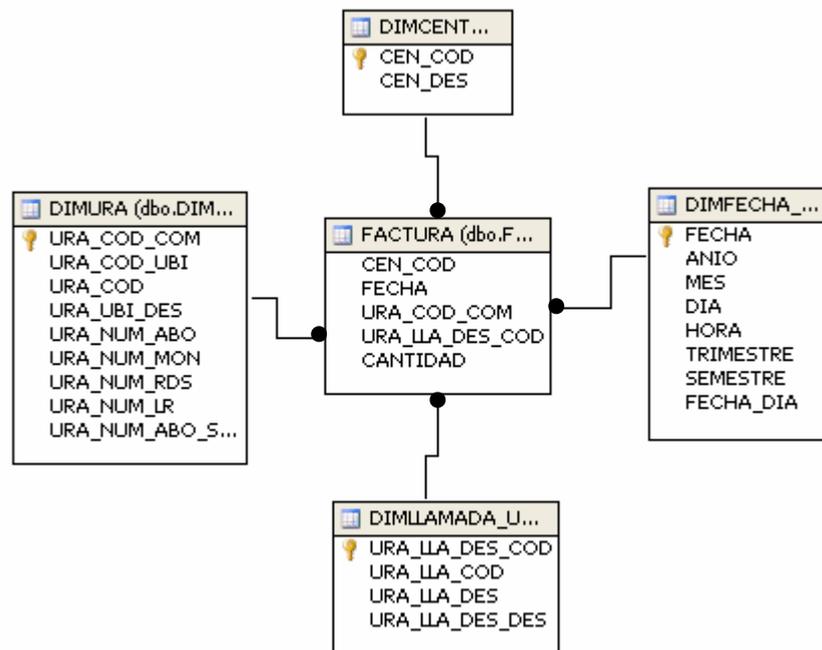


Gráfico 7.11. Esquema Estrella Unidad de Conexión de Abonado (Conmutación)

DIMMURA. Contiene los nombres, ubicación, número de los diferentes abonados y características necesarias de las URAs, para realizar un correcto análisis.

DIMLLAMADA_URA. Contiene los tipos de llamadas o descripciones de las medidas que se necesitan analizar.

FACTURA Contiene la medida cantidad la cual dependerá de la dimensión llamada_URA para poder analizar la información. Para analizar la información siempre tiene que estar presente dicha dimensión para evitar resultados errados.

7.2.1.2.2.5. Tipos de llamadas (Clase T)

Este esquema estrella nos permite conocer los tipos, subtipos y causas de llamada que se pueden presentar en una central en sus diferentes flujos de tráfico:

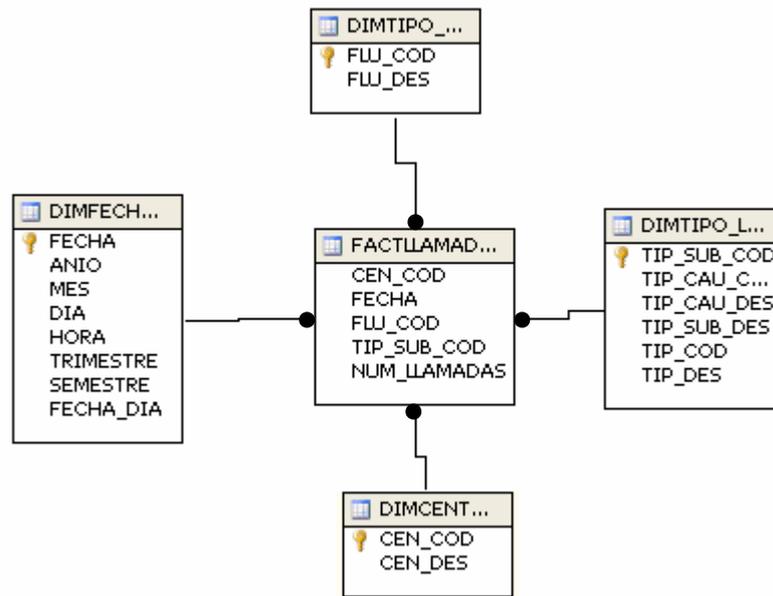


Gráfico 7.12. Esquema Estrella Tipos de llamadas (Conmutación)

DIMTIPO_LLAMADAS. Contiene la información referente al tipo de llamadas, su subclasificación y causas de ineficacia en las llamadas.

DIMTIPO_FLUJO. Es compartida con el esquema estrella de tráfico en erlangs.

FACTLLAMADAS. Contiene la medida num_llamadas al igual que los otros esquemas debido a que existe una dimensión en la cual se describen las medidas. Debe considerarse en el momento del análisis que podría verse afectado el resultado si no está presente la dimensión tipo_llamadas para el nivel de desglose. Queda a consideración del analista el mezclar todas las llamadas para obtener el total de las mismas.

Es posible consolidar la información uniando los esquemas estrella en uno sólo cubo, para poder establecer reportes a nivel de central, flujos de tráfico con los cuales se obtenga un resumen global de lo que realmente esta pasando en las diferentes centrales y puedan tomar mejores decisiones en base a indicadores, quedando de la siguiente manera:

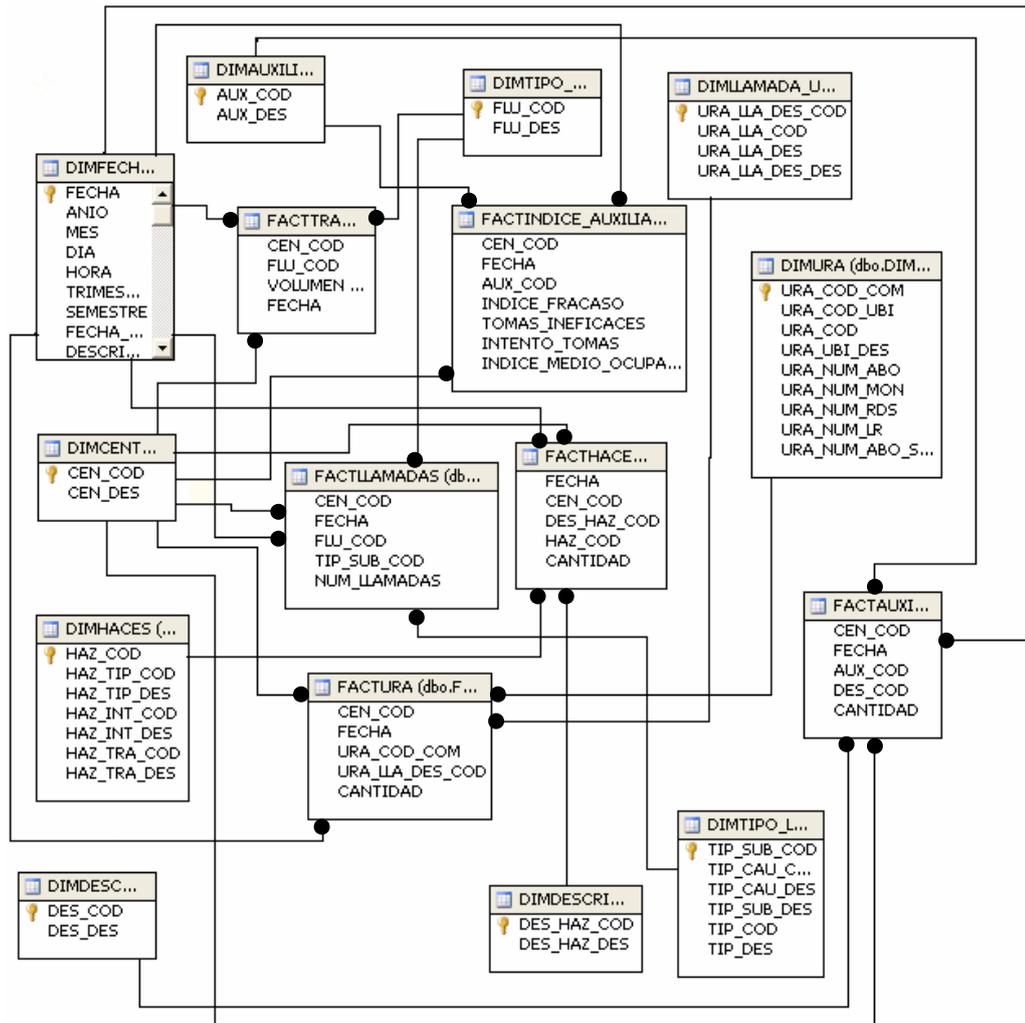


Gráfico 7.13. Esquema Estrella Control de Calidad Circulación Tráfico (Conmutación)

7.2.1.3. Estructura de la base multidimensional

Para construir la estructura de la base dimensional se utilizó *Analysis Services* que es una herramienta de Inteligencia de Negocios de SQL Server 2005. Esta herramienta nos da la facilidad de crear perspectivas (cubos virtuales) que permiten seleccionar una parte del cubo general. El mismo que se construye en base a varias tablas de hecho y varias dimensiones el objetivo de hacer esto es consolidar la información en un sólo cubo, compartiendo dimensiones.

Para saber que divisiones o partes del cubo general se escoge se debe basar en cada uno de los modelos lógicos (esquema estrella) establecidos.

7.2.1.3.1. Tráfico telefónico

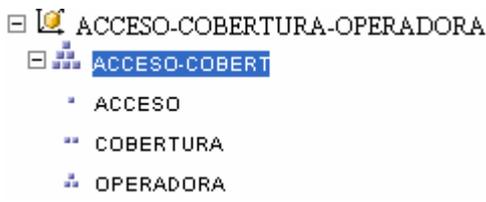
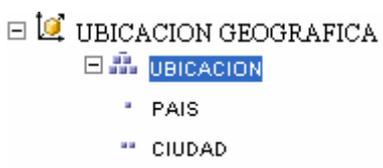
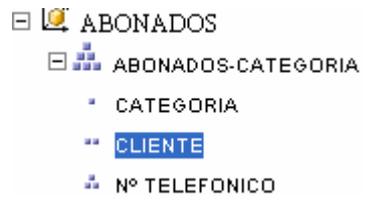
7.2.1.3.1.1. Cubo físico

Se puede consolidar la información de los 4 esquemas estrella establecidos para obtener diferentes análisis ya que se cuenta con diferentes niveles de detalle. El nombre del cubo es *DWH-CUBTRAFICO-TELEFONICO*

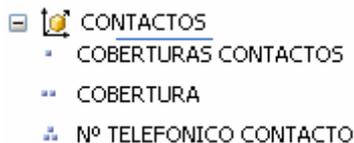
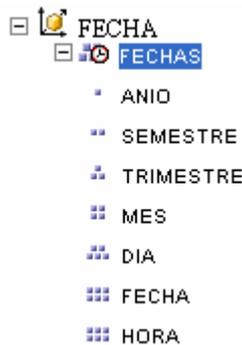
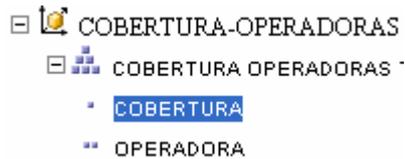
7.2.1.3.1.2. Cubos virtuales

- TRAFICO-ENTRANTE
- TRAFICO-SALIENTE
- TRAFICO-LOCAL
- TRAFICO-LOCAL-DETALLADO

7.2.1.3.1.3. Dimensiones

- **Centrales**
 - 1. Centrales
- **Destino Salida**
 - 1. TIPO CONEXIÓN
 - 2. DESTINO
 - 3. OPERADORA
- **Tarifa**
 - 1. TARIFA
- **Ubicación**
 - 1. PAIS
 - 2. CIUDAD
- **Cliente (Abonado)**
 - 1. CATEGORIA
 - 2. CLIENTE
 - 3. Numero Telefónico

- **Destino Entrada**
 1. DESTINO
 2. OPERADORA
- **Local**
 1. LOCAL
 2. TIPO
- **Fecha**
 1. AÑO
 2. SEMESTRE
 3. TRIMESTRE
 4. MES
 5. DIA
 6. FECHA
 7. HORA
- **Contacto Externo**
 1. DESTINO
 2. DESCRIPCION
 3. CONTACTO (Teléfono)



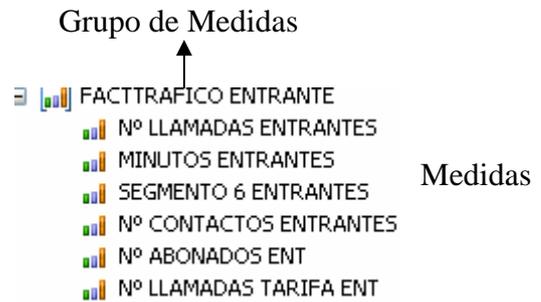
Para el tráfico local detallado se crean vistas de las dimensiones cliente (abonado destino) y central (central destino) ya que en la tabla de hechos dos campos apuntan a la misma tabla.

7.2.1.3.1.4. Medidas

El nombre de las medidas son las mismas para el tráfico saliente, entrante, local y local detallado. Por motivos de espacio sólo se detallarán las medidas del tráfico entrante

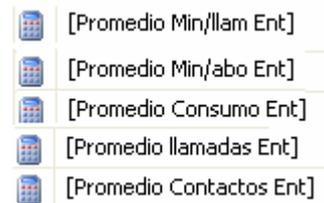
- Número de llamadas
- Minutos Redondeados

- Duración en segundos
- Número Contactos diferentes
- Número de abonados
- Número de llamadas tarifa



Las tres últimas medidas no son especificadas en el esquema estrella porque son el resultado de una función de agregación como es el Count o el Count Distinct de las llaves foráneas que se encuentran en la tabla de hechos. Las siguientes medidas tampoco están especificadas en los modelos debido a que son cálculos o valores derivados que no necesitan ser almacenados y se los conoce como miembros calculados, la única diferencia entre los tráficos es que al final del nombre de la medida termina con entrante o saliente, etc.

- Duración Promedio por abonado
- Duración Promedio de llamada
- Promedio Contactos



7.2.1.3.2. Control de calidad circulación tráfico (conmutación)

7.2.1.3.2.1. Cubos

Se ha diseñado un cubo físico para consolidar la información y realizar reportes generales para la toma de decisiones llamado *DWH-CUBCONMUTACION*.

7.2.1.3.2.2. Cubos virtuales

- AUXILIARES
- TRAFICO-ERLANGS
- HACES
- URA
- LLAMADAS CENTRAL

7.2.1.3.2.3. Dimensiones

- **Auxiliares**
 1. Auxiliares
 - **Descripción Medias Auxiliares**
 1. Descripción
 - **Tipo Flujo**
 1. Flujo
 - **Haces**
 1. Tráfico (destino)
 2. Tipo
 3. Interconexión
 4. Haz
 - **Descripción Medidas Haces**
 1. Descripción Medida
 - **URA**
 1. Ubicación
 2. URA
 - **Descripción Llamada URA**
 1. Descripción Llamada (Medida)
 - **Tipos Llamada**
 1. Tipo llamada
 2. Subtipo (causa)
 3. Llamada
-
- ```
graph TD; AUXILIARES --> TIPOS_AUXILIARES[TIPOS AUXILIARES]; MEDIDAS_AUXILIARES[MEDIDAS AUXILIARES] --> MEDIDAS_AUXILIARES_SUB[MEDIDAS-AUXILIARES]; FLUJO_TRAFICO[FLUJO TRAFICO] --> FLUJOS_DE_TRAFICO[FLUJOS DE TRAFICO]; HACES --> TRAFICO_HACES[TRAFICO HACES]; TRAFICO_HACES --> TIPO_HAZ[TIPO HAZ]; TRAFICO_HACES --> TRAFICO_HAZ[TRAFICO HAZ]; HACES --> CONEXION_HAZ[CONEXION HAZ]; HACES --> Haz; MEDIDAS_HACES[MEDIDAS HACES] --> MEDIDAS_HACES_SUB[MEDIDAS-HACES]; UNIDAD_CONEXION_ABONADO[UNIDAD CONEXION ABONADO] --> URA; URA --> UBICACION[UBICACION]; URA --> URA_SUB[URA]; TIPO_LLAMADA_URA[TIPO LLAMADA URA] --> TIPO_LLAMADAS_URA[TIPO LLAMADAS URA]; TIPO_LLAMADAS_URA --> LLAMADA_URA[LLAMADA URA]; TIPO_LLAMADAS_URA --> SUBTIPO_LLAMADA_URA[SUBTIPO LLAMADA URA]; TIPOS_CAUSAS_DE_LLAMADAS[TIPOS-CAUSAS DE LLAMADAS] --> TIPOS_CAUSAS_LLAMADAS_CENTRAL[TIPOS-CAUSAS LLAMADAS CENTRAL]; TIPOS_CAUSAS_LLAMADAS_CENTRAL --> TIPO_LLAMADA_SUB[TIPO LLAMADA]; TIPOS_CAUSAS_LLAMADAS_CENTRAL --> SUBTIPO; TIPOS_CAUSAS_LLAMADAS_CENTRAL --> CAUSA;
```

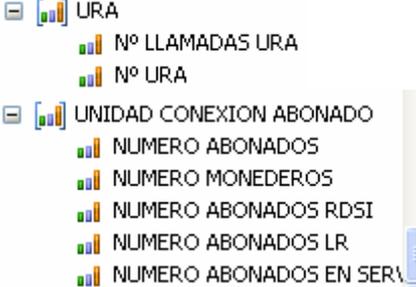
## Dimensiones Compartidas

- **Fecha** (dimensión ya definida en tráfico telefónico)
- **Centrales** (dimensión ya definida en tráfico telefónico)

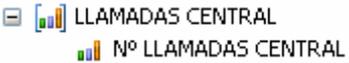
### 7.2.1.3.2.4. Medidas

- **Auxiliares**
    - Cantidad
  - **Tráfico**
    - Cantidad Tráfico
  - **Haces**
    - Cantidad
- 

No se especifica la medida N° haces en el modelo lógico porque este es una agregación (Count Distinct) de la llave Foránea Haz\_cod (código del Haz)

- **URA**
    - Cantidad
    - Número de abonados RDSI
    - Número de abonados Servicio
- 

Las últimas medidas que se puede visualizar son aquellas que pertenecen a la dimensión URA, se establecieron en la dimensión y no en la tabla de hecho porque son medidas fijas que no varían como el transaccional. En el modelo lógico (Esquema Estrella) no se detalla la medida N° URA porque esta medida se obtiene a partir de la agregación (Count Distinct) de la llave foránea (compuesta URA y Central) para obtener el número de URAs que existe en cada una de las centrales.

- **Tipo llamada**
    - Cantidad
- 

## 7.2.2. Documento de identificación de fuentes

### 7.2.2.1. Introducción

El presente documento corresponde al entregable “Identificación de Fuentes de la Información Seleccionada”, de la fase de Especificación y Diseño, de la Propuesta de Valor para el desarrollo del Módulo del Tráfico Telefónico E.T.A.P.A.

En el documento se detallan las diferentes fuentes de información que alimentarán el repositorio de datos “*Data Warehouse*”, de manera que el sistema pueda responder a las preguntas y expectativas establecidas en la fase de Visión y Alcance. De la misma forma se ha realizado un análisis de disponibilidad de la información y se han determinado datos adicionales que deben ser generados.

### 7.2.2.2. Identificación de fuentes de la información seleccionada

Para los *datamarts* de Módulo de Tráfico Telefónico y Conmutación se han identificado las siguientes fuentes de la información:

Sistemas Transaccionales:

- Sistema de Facturación de Telefonía – Servidor AS/400(DB2)
- Sistema de Generación de CDR’s – Servidor Conmutación (Oracle).
- Sistemas de Observaciones Permanentes – Servidor Conmutación (Oracle).

Diccionarios de Datos que deben generarse.

- Jerarquías y Descripciones.

A continuación se especifican las fuentes para cada *Data Mart*:

#### 7.2.2.2.1. Tráfico telefónico

| Sistema Fuente                      | Biblioteca | TABLA FUENTE | DESCRIPCION                                 |
|-------------------------------------|------------|--------------|---------------------------------------------|
| Sistema de Facturación de Telefonía | Qs36f      | Tarar7       | Categorías Clientes                         |
| Sistema de Facturación de Telefonía | Qs36f      | Telarc       | Abonados Actualizados al mes de facturación |

|                                     |       |            |                                                                                                                                |
|-------------------------------------|-------|------------|--------------------------------------------------------------------------------------------------------------------------------|
| Sistema de Facturación de Telefonía | Qs36f | Tel3AAMM   | Historial de los abonados pero del año y mes que indique el nombre de la tabla                                                 |
| Sistema de Generación de CDR's      | Etapa | ENAAMM     | Tráfico detallado de las llamadas entrantes nacional y regional al año y mes especificado en su nombre de tabla                |
| Sistema de Generación de CDR's      | Etapa | EIAAMM     | Tráfico detallado de las llamadas entrante internacional al año y mes especificado en su nombre de tabla                       |
| Sistema de Generación de CDR's      | Etapa | ECAAMM     | Tráfico detallado de las llamadas entrante celular al año y mes especificado en su nombre de tabla                             |
| Sistema de Generación de CDR's      | Etapa | Pais       | Contiene los códigos y descripciones de los Países                                                                             |
| Sistema de Generación de CDR's      | Etapa | Ciudad     | Contiene los códigos y descripciones de las ciudades y los países a que corresponden                                           |
| Sistema de Generación de CDR's      | Etapa | Contadores | Contiene el consumo mensual de los números registrados en la central.                                                          |
| Sistema de Generación de CDR's      | Etapa | SNAAMM     | Tráfico detallado de llamadas salientes nacional y regional al año y mes especificado en su nombre de tabla                    |
| Sistema de Generación de CDR's      | Etapa | SIAAMM     | Tráfico detallado de llamadas salientes internacional al año y mes especificado en su nombre de tabla                          |
| Sistema de Generación de CDR's      | Etapa | SCAAMM     | Tráfico detallado de llamadas salientes celular al año y mes especificado en su nombre de tabla                                |
| Sistema de Generación de CDR's      | Etapa | SEAAMM     | Tráfico detallado de llamadas salientes a servicio tres dígitos y números 1800 al año y mes especificado en su nombre de tabla |
| Sistema de Generación de CDR's      | Etapa | SLAAMM     | Tráfico detallado de llamadas Locales al año y mes especificado en su nombre de tabla                                          |

### Generación de archivos Excel u hojas de cálculo

| Sistema Fuente               | Biblioteca   | TABLA FUENTE            | DESCRIPCION                                                                   |
|------------------------------|--------------|-------------------------|-------------------------------------------------------------------------------|
| Generación de Archivos Excel | Diccionarios | Td_central              | Centrales existentes                                                          |
| Generación de Archivos Excel | Diccionarios | Td_Destino_Salida       | Destinos, operadoras, tipo de conexión(jerarquías)                            |
| Generación de Archivos Excel | Diccionarios | Td_Destino_Entrada      | Destinos, operadoras (jerarquías)                                             |
| Generación de Archivos Excel | Diccionarios | Directorio_1800         | Destino y descripción de los números                                          |
| Generación de Archivos Excel | Diccionarios | Servicio_3digitos       | Servicios gratuitos                                                           |
| Generación de Archivos Excel | Diccionarios | Td_Tipo_local           | Tipos Tráfico local                                                           |
| Generación de Archivos Excel | Diccionarios | Abonados_no_registrados | Series telefónicas que tiene tráfico, pero no están registrados como abonados |

### 7.2.2.2.2. Control de calidad circulación conmutación

| Sistema Fuente                       | Biblioteca | TABLA FUENTE | DESCRIPCION                                         |
|--------------------------------------|------------|--------------|-----------------------------------------------------|
| Sistema de Observaciones permanentes | Etapa      | Clase_A      | Tráfico de los Auxiliares                           |
| Sistema de Observaciones permanentes | Etapa      | Clase_C      | Tráfico tipos de flujo de la central                |
| Sistema de Observaciones permanentes | Etapa      | Clase_Fb     | Tráfico de los haces                                |
| Sistema de Observaciones permanentes | Etapa      | Clase_R      | Comportamiento de la unidad de conexión del abonado |
| Sistema de Observaciones permanentes | Etapa      | Clase_T      | Tipos de llamadas generadas en la central           |

### Generación de archivos Excel u hojas de cálculo

| Sistema Fuente               | Biblioteca   | TABLA FUENTE              | DESCRIPCION                                                             |
|------------------------------|--------------|---------------------------|-------------------------------------------------------------------------|
| Generación de Archivos Excel | Diccionarios | Td_llamada_URA            | Medidas(comportamiento) de la URA                                       |
| Generación de Archivos Excel | Diccionarios | Td_Tipo_flujo             | Tipos de flujo                                                          |
| Generación de Archivos Excel | Diccionarios | Td_Descripcion_Auxiliares | Medidas(comportamiento) de Auxiliares                                   |
| Generación de Archivos Excel | Diccionarios | Td_Descripcion_Haces      | Medidas(comportamiento) de Haces                                        |
| Generación de Archivos Excel | Diccionarios | Td_Auxiliares             | Tipos de Auxiliares                                                     |
| Generación de Archivos Excel | Diccionarios | Td_URA                    | Tipos, número abonados URA                                              |
| Generación de Archivos Excel | Diccionarios | Td_Haces                  | Tipos, conexión, tráfico Haces(jerarquía)                               |
| Generación de Archivos Excel | Diccionarios | Td_tipo_llamada           | Tipos, subtipos y causas de las llamadas que se originan en la central. |

**Nota:** El *datamart* de Tráfico Telefónico comparte fuentes de información con el *datamart* de Conmutación. (Diccionario Centrales) Para revisar a detalle las fuentes es necesario construir una metadata que no es más que la especificación de los datos que se encuentra en el anexo I.

### 7.2.2.3. Consideraciones importantes:

Debido a que se realizó el análisis de la información se debe considerar que:

- Con respecto a los destinos de las llamadas a números 1800 para la generación de reportes se debe tomar en cuenta que en el diccionario 1800 se encuentra más de una descripción o destino para un mismo número 1800, se presenta el inconveniente porque en las tablas de los registros de las llamadas no se puede distinguir a que destino fue o a quien se realizó la llamada por lo que se tiene el riesgo de que se duplique la información para conseguir este tipo de reporte.
- Para la fase de pruebas o validación de datos se debe considerar que los reportes de categoría sobre el volumen del tráfico que se generen en el *Data Warehouse* se basa en categorías actualizadas por esta razón no coincidirá los valores obtenidos por conmutación.
- El tráfico local (consumo abonados) no existe nivel de detalle porque no se registran las llamadas, pero hay algunas excepciones como el Banco del Pichincha, etc.
- No se puede incluir las observaciones temporales ya que no existe un transaccional.
- Se debe mantener un control y dar un mantenimiento adecuado a las tablas país y ciudad debido a que existen inconsistencia y errores en los datos.
- El tráfico sobre las URAs no se podrá realizar porque no existe la fuente de donde extraer la información, ni tampoco se ha recibido indicaciones para crear un diccionario en una hoja de cálculo.
- Los cálculos para las medidas e índices de eficacia están basados en los manuales de ALCATEL.
- No se ha considerado la extracción de la tabla EEAAMM (Tráfico entrante especial) debido a que las personas que manejan el sistema no conocen sobre su contenido.
- Se necesita una revisión constante de las tablas de conmutación: Clase\_A, Clase\_C, Clase\_Fb, Clase\_R, Clase\_T ya que existen fechas erróneas y

obviamente no se están considerando para la carga al *Data Warehouse*. También debe asegurarse que se estén subiendo los datos diariamente.

- La serie telefónica 848801 hasta 848999 pertenece al Banco del Pichincha se hace la aclaración porque no están registrados como abonado de etapa.
- Se ha creado un diccionario de datos con las series correspondientes al Banco del Pichincha y otras series que se utilizan en Conmutación.
- Existen números telefónicos que tienen tráfico pero no están considerados como abonados (no existen en la tabla Telarc) y no existen en el diccionario creado por lo que se estableció como abonados no definidos y categoría no definida.
- En el área de conmutación nos indicaron que no se puede establecer el país de origen de las llamadas debido a que no existen como identificar en los registros de las llamadas.
- En el tráfico local no se puede llegar a nivel de hora en el detalle de la fecha.
- En el tráfico local no se puede tener detalle de las llamadas como se lo tiene en el servicio tres dígitos.

#### **7.2.2.4. Puntos a Tomar en Cuenta:**

Con las especificaciones actuales se debe tomar en cuenta los siguientes puntos que afectan el alcance del proyecto:

- Se debe tener permiso sobre las tablas ENAAMM, EIAAMM, ECAAMM, SNAAMM, SIAAMM, SCAAMM, SEAAMM, SLAAMM para poder realizar las extracciones debido a que se crean de acuerdo al mes de consumo.
- Existe la posibilidad de que realicen cambios en los nombres de tablas, campos o bibliotecas debido a que se está desarrollando un proyecto llamado “SIGECOM”.
- La información sobre tráfico telefónico que se cargará al *Data Warehouse* tiene un retraso de un mes con relación a la fecha actual debido a que el negocio no puede establecer el consumo hasta que concluya el mes.
- La actualización de los abonados se realiza cada segundo viernes del mes por esta razón se ha decidido que los procesos de extracción, transformación,

población e integración se corran dos veces al mes el primero para tener la información del mes y el segundo será para actualizar la *data*.

- La información del Control de calidad de circulación del tráfico “Conmutación” se cargará diariamente aquí no existen modificaciones a los datos.
- Se debe considerar que la información del transaccional (oracle) puede no estar disponible por problemas de *hardware* y en este caso se debe considerar procesos adicionales para mantener los datos actualizados.

### **7.3. Fase III Desarrollo**

#### **7.3.1. Documento técnico**

##### **7.3.1.1. Proceso general tráfico telefónico**

Las operaciones para construir el *Data Warehouse* se han basado en las políticas de la empresa con relación a la disponibilidad de la información.

En el proceso del módulo del tráfico telefónico existen dos procesos importantes. El primero se refiere a la generación de llamadas y se ejecutará dos veces al mes debido a que la información no es confiable hasta después de la facturación del tráfico telefónico (esto se realiza cada segundo viernes del mes después del consumo), la cual incluye cambios en los datos, por esta razón se ejecutará el primero y el 17 de cada mes a las 02H00. El procedimiento se encuentra en el directorio C:\WORKDIR\DTSS\TRAFICO\_TELEFONICO\BATS\ PROCESO\_ TRAFICO\_TELEFONICO.BAT el cual será ejecutado a través de una tarea programada.

El segundo está dirigido a controlar la calidad de la circulación del tráfico (Conmutación). Se ejecutará todos los días (tareas programadas) a las 19H00 y se encuentra en el directorio C:\WORKDIR\DTSS\TRAFICO\_TELEFONICO\BATS\ PROCESO\_CONMUTACION.BAT

Se debe respetar las horas de ejecución debido a que se comparten dimensiones y el orden afectaría la creación correcta de la base de datos.

Los archivos .bat contienen los procesos generales de extracción, transformación, población, integridad y generación de dimensiones y cubos.

#### **7.3.1.1.1. Extracción y transformación**

La extracción consiste en obtener datos necesarios de diversas fuentes para ser almacenados temporalmente en una misma base de datos secundaria aquí denominada Tanque en la cual, nos basaremos para la posterior transformación y población del *Data Warehouse*.

La transformación se encarga de resolver las inconsistencias en los formatos de datos y la codificación, dentro de una base de datos única. También se encarga de las inconsistencias en el contenido de datos, por lo que han sido establecidas reglas de transformación en la fase de identificación de fuentes (metadata), para obtener datos conciliables y consistentes para cargarlos al *Data Warehouse*.

El proceso de extracción permite extraer los datos desde el servidor POLYLEPSIS.ETAPA.NET.EC (DB2) y TELECOM.WORLD (*Oracle*) al servidor TESISINFO-1 (local) ó Sacha.net.ec (Servidor E.T.A.P.A.). Este proceso se lleva a cabo utilizando la herramienta *Integration Services Project (ISP)* de *SQL Server 2005*.

Los *ISP* extraen los datos desde las bibliotecas QS36F del servidor POLYLEPSIS, desde ETAPA del servidor TELECOMWORLD, así como de archivos de hojas cálculo. Los datos extraídos se almacenan en las tablas de la base de datos TANQUE. Posteriormente los transforma y los coloca en la base de datos llamada *DWH* del motor de base de datos *SQL Server*.

Los procesos de extracción y transformación se basan en el entregable generado en la segunda fase denominado “Identificación de fuentes” (metadata). Es por ello que es imprescindible el uso de una metadata bien definida.

Es muy importante considerar la retrospcción en los objetos *DWH*, porque se debe crear procesos adicionales para controlar que los datos que se están transformando sean los correctos y no se pierda información que sea necesaria para la empresa, por

está razón se detalla a continuación cada uno de los procesos que se llevan a cabo para realizar bien la construcción del *DWH*.

#### **7.3.1.1.1.1. Extracción y transformación tráfico telefónico**

Se ha establecido como una norma para los nombres de los paquetes con la extensión .dtsx pertenecientes a las extracciones, la utilización de un formato que permitirá definir la fuente de la cual se extraen los datos y la base destino. El Formato incluye el nombre de la biblioteca seguida del nombre de la tabla (fuente de extracción) y separada por un guión estará el nombre de la base de datos destino el cual incluirá el nombre de la base y la tabla en la cual se depositarán los datos extraídos (base destino).

#### **EXTRACCION\_TRAFICO\_TELEFONICO.BAT**

Este archivo .bat contiene todos los procesos de extracción del tráfico telefónico, los mismos que se detallan a continuación:

#### **7.3.1.1.1.1.1. Extracción diccionarios**

#### **EXTRACCION\_DICCIONARIOS\_TRAFICO\_TELEFONICO.BAT**

Los diccionarios son archivos Excel a los cuales hay que darles mantenimiento. Contienen información necesaria para el análisis pero no se encuentran en las bases de datos transaccionales. Estos deben ser creados de acuerdo al negocio y al nivel de detalle en el que se va analizar la información con respecto a las dimensiones. El procedimiento se encuentra en el directorio C:\WORKDIR\DTSS\TRAFICO\_TELEFONICO\ARCHIVOS.

Los .dtsx que ejecutan la extracción de los diccionarios son:

DICCIONARIO.TD\_DESTINO\_SALIDA-DWHTANQUE.TD\_DESTINO\_SALIDA.dtsx

Extrae la información de jerarquía de destinos. Es la cobertura con la que llega la llamada por ejemplo nacional, regional, etc.. Las operadoras son utilizadas para realizar las interconexiones y tipo de acceso que se refiere a la forma como se realiza la llamada: discado directo, tarjeta prepago o uso de operadora.

DICCIONARIO.TD\_TARIFA-DWHTANQUE.TD\_TARIFA.dtsx

Extrae la información del tipo de tarifa. Es fundamental para el análisis del tráfico celular.

DICCIONARIO.TD\_DESTINO\_ENTRADA-DWHTANQUE.TD\_DESTINO\_ENTRADA.dtsx

Extrae la información de jerarquía de destinos, operadoras con relación al tráfico entrante.

DICCIONARIO.SERVICIO\_3DIGITOS-DWHTANQUE.SERVICIO\_3DIGITOS.dtsx

Extrae la descripción de los servicios de tres dígitos

DICCIONARIO.TIPO\_LOCAL-DWHTANQUE.TD\_TIPO\_LOCAL.dtsx

Extrae la información del tipo de tráfico local

DICCIONARIO.DIRECTORIO\_1800-DWHTANQUE.DIRECTORIO\_1800.dtsx

Extrae información de los nombres o descripciones de los números 1800.

DICCIONARIO.ABONADOS\_NOREGISTRADOS-DWHTANQUE.ABONADOS\_NO\_REGISTRADOS.dtsx

Extrae información de las series creadas y no registradas como abonados de E.T.A.P.A., para ayudar con las consideraciones de datos perdidos y para cumplir con la retrospección verdadera.

#### **7.3.1.1.1.2. Extracción tráfico entrante**

**EXTRACCION\_TRAFICO\_ENTRANTE.BAT.**- Este archivo .bat contiene los siguientes paquetes:

ETAPA.ENAAMM-DWHTANQUE.ENACIONA.dtsx

ETAPA.ENAAMM-DWHTANQUE.EREGIONA.dtsx

ETAPA.EIAAMM-DWHTANQUE.EINTERNA.dtsx

ETAPA.ECAAMM-DWHTANQUE.ECELULAR.dtsx

Estos paquetes extraen la información de los CDR's (código de registro), donde se encuentran detalladas las llamadas entrantes. Se realiza una transformación a nivel bajo es decir, conversión de tipos datos implícitos. En esta sección no se especifica la información que se extrae en cada paquete debido a que se puede consultar en el entregable de Identificación de Fuentes o en el anexo I (metadata).

#### **7.3.1.1.1.3. Transformación tráfico entrante**

**EXTRACCION\_TH\_Y\_TD\_ENTRANTE.BAT.**- Los nombres de los paquetes con extensión .dtsx que empiezan con TH transforma la información a la tabla de hechos llamada th\_entrada\_trafico.

TH\_ENTRANTE\_REGIONAL.dtsx

TH\_ENTRANTE\_NACIONAL.dtsx

TH\_ENTRANTE\_INTERNACIONAL.dtsx

TH\_ENTRANTE\_CELULAR.dtsx

Transforma la información detallada de las llamadas entrantes por destinos, los mismos que están especificados en el nombre de cada uno de los paquetes de transformación. Se aplican todas las reglas y cálculos necesarios definidos en la fase de planeación “Identificación de Fuentes”, para armar la tabla de hechos de tráfico entrante.

TH\_ENTRANTE\_NACIONAL\_NO\_REG.dtsx

TH\_ENTRANTE\_REGIONAL\_NO\_REG.dtsx

TH\_ENTRANTE\_INTERNACIONAL\_NO\_REG.dtsx

TH\_ENTRANTE\_CELULAR\_NO\_REG.dtsx

Se realizan estos procesos adicionales para no perder datos de importancia para la empresa como son los números telefónico que generan tráfico pero que no están registrados como abonados de E.T.A.P.A. Transforma la información detallada de las llamadas entrantes por destino y se almacena en la tabla de hechos de tráfico entrante(th\_entrada\_trafico).

TD\_CLIENTE\_TRAFICO\_ENTRANTE.dtsx

Transforma la información detallada de los clientes y la jerárquica por categorías pero considerando sólo a aquellos clientes que hayan generado tráfico(es decir que estén en las tablas de hecho). Se debe considerar que comparten la dimensión y por esta razón sólo se agregarán los clientes que no existan. Este procedimiento se correrá después del procedimiento del tráfico saliente.

TD\_FECHA\_TRAFICO\_ENTRANTE.dtsx

Transforma la información detallada y jerárquica del tiempo en el que se realiza una llamada. Se almacenarán los datos que no existen en la dimensión fecha, puesto que ésta ya es cargada previamente en el procedimiento de conmutación.

TD\_UBICACION\_ENTRANTE.dtsx

Transforma la información detallada de la ubicación geográfica de donde se recibe la llamada. Sólo se carga los datos que no existen en la dimensión ubicación, ya que primero se ejecutarán los procedimientos de tráfico saliente.

TD\_CONTACTO\_EXTERNO\_TRAFICO\_ENTRANTE.dtsx

Transforma la información detallada de quien realiza la llamada a un abonado de E.T.A.P.A., cargándose sólo los contactos externos que no se encuentran en la dimensión.

#### **7.3.1.1.1.4. Extracción tráfico saliente**

**EXTRACCION\_TRAFICO\_SALIENTE.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

**LIMPIAR\_DIRECTORIO.dtsx**

Limpia la información, completando o rellenando el número de dígitos que corresponde al número 1800 que se extrajeron anteriormente de un diccionario de datos. Dicho diccionario se obtiene desde Internet por ello el motivo de su limpieza.

**ETAPA.PAIS-DWHTANQUE.PAIS.dtsx**

Extrae la información correspondiente a los países que pueden realizar o recibir las llamadas.

**ETAPA.CIUDAD-DWHTANQUE.CIUDAD.dtsx**

Extrae la información correspondiente a las ciudades o estados que se pueden realizar o desde las cuales se puede recibir las llamadas.

**ETAPA.SREGIONA-DWHTANQUE.SREGIONA.dtsx**

**ETAPA.SNACIONA-DWHTANQUE.SNACIONA.dtsx**

**ETAPA.SCELULAR-DWHTANQUE.SCELULAR.dtsx**

**ETAPA.SINTERNA-DWHTANQUE.SINTERNA.dtsx**

**ETAPA.SESPECIA-DWHTANQUE.SESPECIA.dtsx**

Estos paquetes extraen la información de los CDR's (código de registro), donde se encuentran detalladas las llamadas salientes. Se realiza una transformación a nivel bajo es decir, conversión de tipos datos implícitos. En esta sección no se especifica la información que se extrae en cada paquete debido a que se puede consultar en el entregable de Identificación de Fuentes o en el anexo I (metadata).

**QS36F.TELARC-DWHTANQUE.TELAR\_MES.dtsx**

Extrae información de los abonados de E.T.A.P.A. con sus respectivas características.

**QS36F.TARAR7-DWHTANQUE.TARAR7.dtsx**

Extrae información de las categorías que se asigna a un número telefónico.

**ETAPA.SLOCAL-DWHTANQUE.SLOCAL.dtsx**

Extrae información de los contadores de los abonados y números registrados en la central de E.T.A.P.A.

**TRAFICO\_1800.dtsx**

Separa y extrae información detallada de las llamadas realizadas exclusivamente a números 1800 y además identifica el destino de la llamada (nacional, internacional, etc.).

TRAFICO\_1800\_NOREGISTRADO.dtsx

Separa la información detallada de las llamadas realizadas exclusivamente a números 1800, en las cuales no se puede identificar el destino de la llamada (nacional, internacional, etc.), se realiza este proceso adicional para no perder datos que son de interés para la empresa.

TRAFICO\_LOCAL\_SERVICIOS.dtsx

Separa la información detallada de las llamadas realizadas exclusivamente a números tres dígitos de las extracciones antes realizadas.

TRAFICO\_LOCAL.dtsx

Calcula el consumo mensual del tráfico local de los abonados y números registrados en la central en base a las extracciones ya realizadas de las lecturas de los contadores.

#### **7.3.1.1.1.5. Transformación tráfico saliente**

Estos archivos o paquetes con extensión .dtsx también se encuentran en el archivo: extraccion\_trafico\_saliente.bat que contiene los siguientes procesos:

TH\_TRAFICO\_REGIONAL\_REGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_NACIONAL\_REGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_INTERNACIONAL\_REGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_CELULAR\_REGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_1800\_REGISTRADO\_CLIENTE.dtsx

Estos paquetes se encargan de la transformación de la información detallada de las llamadas salientes por destinos, los mismos que están especificados en el nombre de cada uno de los paquetes de transformación. Se aplican todas las reglas y cálculos necesarios definidos en la fase de planeación “Identificación de Fuentes”, para construir la tabla de hechos de tráfico saliente llamada th\_salida\_trafico.

TH\_TRAFICO\_REGIONAL\_NOREGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_NACIONAL\_NOREGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_INTERNACIONAL\_NOREGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_CELULAR\_NOREGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_1800\_NOREGISTRADO\_CLIENTE.dtsx

Este grupo de paquetes transforman la información detallada de las llamadas salientes a los diferentes destinos con respecto a los abonados que no están registrados pero generan tráfico. Estos procesos adicionales son necesarios para no perder datos de importancia para la empresa.

TH\_TRAFICO\_LOCAL\_REGISTRADO\_CLIENTE.dtsx

TH\_TRAFICO\_SERVICIOS\_REGISTRADO\_CLIENTE.dtsx

Estos dos paquetes transforman la información del consumo mensual del tráfico local (entre la red de E.T.A.P.A.) y la información detallada de las llamadas salientes a números de servicio gratuitos (tres dígitos) respectivamente y se encargan de almacenar los datos en la tabla de hechos llamada th\_trafico\_local.

TH\_TRAFICO\_CONSUMO\_LOCAL\_NOREGISTRADO\_CLIENTE.dtsx

TRAFICO\_LOCAL\_SERVICIO\_NOREGISTRADO\_CLIENTE.dtsx

Estos dos paquetes transforman la información consumo mensual del tráfico local y el detalle de las llamadas salientes a números tres dígitos considerando los números telefónicos que generan tráfico pero no están registrados como abonados. Estos procesos adicionales son necesarios para no perder datos importantes para la empresa.

TH\_TRAFICO\_LOCAL\_DETALLADO\_REGISTRADO\_CLIENTE.dtsx

Transforman la información detallada de las llamadas salientes entre la red de E.T.A.P.A. (abonado– abonado) y almacena los datos en la tabla de hecho llamada th\_trafico\_local\_detallado.

TH\_TRAFICO\_LOCAL\_DETALLADO\_NOREGISTRO\_ORI\_DES.dtsx

Transforma la información detallada de las llamadas salientes entre la red de E.T.A.P.A. (abonado – abonado) con respecto a los números telefónicos que originan y reciben las llamadas, generando tráfico pero que no están considerados como abonados.

TH\_TRAFICO\_LOCAL\_DETALLADO\_NOREGISTRADO\_ORIGEN.dtsx

Transforma la información detallada de las llamadas salientes entre la red de E.T.A.P.A. (abonado– abonado) con respecto a los números telefónicos que originan la llamada, generando tráfico pero que no están considerados como abonados.

TH\_TRAFICO\_LOCAL\_DETALLADO\_NOREGISTRADO\_DESTINO.dtsx

Transforma la información detallada de las llamadas salientes entre la red de E.T.A.P.A. (abonado– abonado) con respecto a los números telefónicos que recibe la llamada, generando tráfico pero que no están considerados como abonados.

Estos procesos adicionales son necesarios para no perder datos de importancia para la empresa y que serán almacenados en la tabla de hechos llamada th\_trafico\_local\_detallado.

TD\_UBICACION.dtsx

Transforma la información de la tabla de hechos de “tráfico saliente” y la carga a una ubicación geográfica, es decir, a donde se realiza la llamada. Es importante considerar que en esta parte del proceso es necesario borrar la tabla para almacenar las nuevas ubicaciones.

TD\_UBICACION\_NOREG\_CIUADAD.dtsx

Transforma la información de la tabla de hechos de “tráfico saliente” y las carga a una

ubicación geográfica, es decir, a donde se realiza la llamada, pero considerando que existe el país y no existe la ciudad. Es un proceso adicional para no perder los datos.

#### TD\_CONTACTO\_EXTERNO\_NORMAL.dtsx

Transforma la información de la tabla de hechos de “tráfico saliente”. Se toma en cuenta la información de a quienes se realizó las llamadas. Se incluyen los números marcados a ciudades (regional/nacional), países (internacional) y celular.

#### TD\_CONTACTO\_EXTERNO\_3DIGITOS.dtsx

Transforma la información de la tabla de hechos “local detallado”. Se toma en cuenta la información de a quienes se realizó llamadas por servicio gratuito (bomberos, policía, etc.)

#### TD\_CONTACTO\_EXTERNO\_1800.dtsx

Transforma la información de la tabla de hechos “tráfico saliente”. Se toma en cuenta la información de a quienes se realizó llamadas por servicio 1800 y están registradas en el directorio 1800.

#### TD\_CONTACTO\_EXTERNO\_1800\_NODEFINIDO.dtsx

Transforma la información de la tabla de hechos “tráfico saliente”. Se toma en cuenta la información de a quienes se realizó llamadas por servicio 1800 pero que no se ha identificado el destino. Es un proceso necesario para no perder datos.

#### TD\_CLIENTE\_TRAFICO.dtsx

Transforma la información de la tabla de hechos “tráfico saliente”. Se toma en cuenta la información de quienes realizaron las llamadas a destinos como nacional, regional, internacional, celular, 1800.

#### TD\_CLIENTE\_TRAFICO\_LOCAL.dtsx

Transforma la información de la tabla de hechos “trafico local”. Se toma en cuenta la información de quienes realizaron llamadas locales es decir de abonado a abonado.

#### TD\_CLIENTE\_TRAFICO\_LOCAL\_DETALLADO\_ORIGEN.dtsx

Transforma la información de la tabla de hechos” trafico local detallado”. Se toma en cuenta la información de quienes realizaron llamadas locales, pero estas tienen el detalle del servicio de las llamadas.

#### TD\_CLIENTE\_TRAFICO\_LOCAL\_DETALLADO\_DESTINO.dtsx

Transforma la información de la tabla de hechos “trafico local detallado”. Se toma en cuenta la información de a quienes recibieron llamadas locales y que constan en el detalle de las llamadas.

#### TD\_FECHA\_TRAFICO.dtsx

Transforma la información de fechas que originó la llamada nacional, regional, internacional, celular existente en la tabla de hechos de tráfico telefónico saliente

#### TD\_FECHA\_TRAFICO\_LOCAL.dtsx

Transforma la información de fechas que originó la llamada local existente en la tabla de hechos de tráfico local.

TD\_FECHA\_TRAFICO\_LOCAL\_DETALLADO.dtsx

Transforma la información de fechas que originó la llamada local existente en la tabla de hechos de tráfico local detallado.

### **7.3.1.1.1.2. Extracción conmutación**

#### **7.3.1.1.1.2.1. Diccionarios control de calidad de circulación tráfico telefónico (conmutación)**

**EXTRACCION\_DICCIONARIOS\_CONMUTACION.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

DICCIONARIO.TD\_AUXILIARES-DWHTANQUE.TD\_AUXILIARES.dtsx

Extrae información de los tipos de auxiliares que se encuentran en las centrales.

DICCIONARIO.TD\_DESCRIPCION\_AUXILIARES-DWHTANQUE.TD\_DESCRIPCION\_AUXILIARES.dtsx

Extrae información de las descripciones de las medidas de los auxiliares.

DICCIONARIO.TD\_DESCRIPCION\_HACES-DWHTANQUE.TD\_DESCRIPCION\_HACES.dtsx

Extrae información de las descripciones de las medidas de los haces.

DICCIONARIO.TD\_HACES-DWHTANQUE.TD\_HACES.dtsx

Extrae información de las jerarquías de los haces (tipos), información del tipo de interconexión y el destino del tráfico.

DICCIONARIO.TD\_TIPO\_LLAMADAS-DWHTANQUE.TD\_TIPO\_LLAMADAS.dtsx

Extrae información de los tipos de llamadas que se registran en la central y causas de ineficacia.

DICCIONARIO.TD\_URA-DWHTANQUE.TD\_URA.dtsx

Extrae información característica de una URA.

DICCIONARIO.TD\_CENTRAL-DWHTANQUE.TD\_CENTRALES.dtsx

Extrae información de los nombres de las centrales.

DICCIONARIO.TD\_TIPOS\_FLUJO-DWHTANQUE.TD\_TIPOS\_FLUJO.dtsx

Extrae información de los tipos de flujo que se presenta en el tráfico telefónico.

DICCIONARIO.TD\_LLAMADA\_URA-  
DWHTANQUE.TD\_LLAMADA\_URA.dtsx

Extrae información de las descripciones de las medidas de las URAs.

**Nota:** La construcción de los diccionarios para conmutación están basados en los manuales Alcatel 1000 E10 (OCB283) “Gestión de las observaciones de carga y tráfico” y las fichas de explotación y mantenimiento “Compilación de análisis de las observaciones OBRECOB”.

#### **7.3.1.1.2.2. Extracción y transformación auxiliares**

**EXTRACCION\_CONMUTACION\_AUXILIARES.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

ETAPA.CLASE\_A-DWHTANQUE.AUXILIARES.dtsx

Extrae información detallada del comportamiento de los auxiliares.

TH\_AUXILIARES\_RGF\_AQ1.dtsx

Transforma la información detallada del comportamiento del auxiliar RGF y la descripción de su respectiva medida AQ1.

TH\_AUXILIARES\_RGF\_CCF\_V23.dtsx

Transforma la información detallada del comportamiento de los auxiliares RGF, CCF y V23 y la descripción de sus respectivas medidas

TD\_FECHA\_TRAFICO.dtsx

Transforma la información detallada de las fechas que se encuentran en la tabla de hechos de auxiliares.

#### **7.3.1.1.2.3. Extracción y transformación tráfico erlangs**

**EXTRACCION\_CONMUTACION\_TRAFICO\_ERLANGS.BAT.-** Los archivos .bat contienen los siguientes paquetes con extensión .dtsx

ETAPA.CLASE\_C-DWH.CLASE\_C.dtsx

Extrae información detallada de los flujos de tráfico que se han generado en las diferentes centrales.

TH\_TRAFICO\_ERLANGS.dtsx

Transforma la información de los flujos de tráfico y calcula las medidas necesarias asignando un código para su distinción.

TD\_FECHA\_TRAFICO.dtsx

Extrae las fechas de la tabla de hechos de trafico erlangs, para transformarlas en una jerarquía del tiempo.

#### **7.3.1.1.1.2.4. Extracción y transformación tráfico haces**

**EXTRACCION\_CONMUTACION\_HACES.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

ETAPA.CLASE\_FB-DWHTANQUE.CLASE\_FB.dtsx

Extrae información detallada y necesaria de los haces.

TH\_HACES\_ENTRANTES\_FPA.dtsx

TH\_HACES\_SALIENTES\_FPD.dtsx

TH\_HACES\_MIXTOS\_FPA.dtsx

TH\_HACES\_MIXTOS\_FPD.dtsx

TH\_HACES\_ENTRANTES\_FRA.dtsx

TH\_HACES\_ENTRANTES\_FEA.dtsx

TH\_HACES\_ENTRANTES\_FOA.dtsx

TH\_HACES\_SALIENTES\_FRD.dtsx

TH\_HACES\_SALIENTES\_FED.dtsx

TH\_HACES\_SALIENTES\_FOD.dtsx

TH\_HACES\_MIXTOS\_FRA.dtsx

TH\_HACES\_MIXTOS\_FRD.dtsx

TH\_HACES\_MIXTOS\_FEA.dtsx

TH\_HACES\_MIXTOS\_FED.dtsx

TH\_HACES\_MIXTOS\_FOA.dtsx

TH\_HACES\_MIXTOS\_FOD.dtsx

TH\_HACES\_FQ1.dtsx

TH\_HACES\_FS1.dtsx

TH\_HACES\_PRS1.dtsx

Este grupo de paquetes transforman la información de los haces, clasificada de acuerdo al tipo. Se calcula las medidas: FPA (tomas de llegada), FPD (tomas de salida), FPA (tomas de

llegada), FPD (tomas de salida), FRA (rechazos en las tomas de llegada), FEA (llamadas eficientes en la entrada), FOA (Tráfico en la llegada),FRD (Tráfico a la salida), FED (llamadas eficientes a la salida), FOD (Tráfico en la salida), FRA (rechazos en las tomas de llegada), FRD (Tráfico a la salida), FEA (llamadas eficientes en la entrada), FED (llamadas eficientes a la salida), FOA (Tráfico en la llegada), FOD (Tráfico en la salida), FS1 (numero de haces en servicio), PRS1 (tomas simultáneas de circuitos), respectivamente. Cada uno de los paquetes lleva como parte de su nombre las medidas que calcula.

TD\_FECHA\_HACES.dtsx

Extrae las fechas de la tabla de hechos haces para transformarlas en una jerarquía del tiempo.

#### **7.3.1.1.1.2.5. Extracción y transformación URA**

**EXTRACCION\_CONMUTACION\_URA.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

ETAPA.CLASE\_R-DWH.CLASE\_R.dtsx

Extrae información detallada sobre las URAs.

TH\_URA.dtsx

Transforma la información de las URAs formando jerarquías de acuerdo a la ubicación y calcula sus medidas.

TD\_FECHA\_TRAFICO.dtsx

Extrae las fechas de la tabla de hechos de haces para transformarlas en una jerarquía del tiempo.

#### **7.3.1.1.1.2.6. Extracción y transformación tipos de llamadas**

**EXTRACCION\_CONMUTACION\_CLASE\_T.BAT.-** Este archivo .bat contiene los paquetes con extensión .dtsx que extraen y transforman los datos. Cada uno de estos paquetes llevan como parte del nombre las medidas que calcula.

ETAPA.CLASE\_T-DWH.TANQUE.CLASE\_T.dtsx

Extrae información detallada de los tipos de llamadas que se generan en la central.

TH\_LLAMADAS\_PRESENTADAS\_I.dtsx

TH\_LLAMADAS\_PRESENTADAS\_A.dtsx

TH\_LLAMADAS\_PRESENTADAS\_D.dtsx

TH\_LLAMADAS\_PRESENTADAS\_T.dtsx

Estos paquetes transforman la información de acuerdo al tipo de flujo de tráfico (interno, entrante, saliente y en tránsito) y calcula el tipo de llamadas eficaces, en este caso las llamadas presentadas.

TH\_LLAMADAS\_COMPLETADAS\_I.dtsx

TH\_LLAMADAS\_COMPLETADAS\_D.dtsx

Transforma la información de acuerdo al tipo de flujo de tráfico (interno y de salida) y calcula el tipo de llamadas eficaces, en este caso las llamadas completadas.

TH\_LLAMADAS\_TRATADAS\_I.dtsx

TH\_LLAMADAS\_TRATADAS\_A.dtsx

TH\_LLAMADAS\_TRATADAS\_D.dtsx

TH\_LLAMADAS\_TRATADAS\_T.dtsx

Transforma la información de acuerdo al tipo de flujo de tráfico (interno, entrante, saliente y en tránsito) y calcula el tipo de llamadas eficaces, en este caso las llamadas tratadas

TH\_LLAMADAS\_TOMA\_CIRCUITO\_D.dtsx

TH\_LLAMADAS\_TOMA\_CIRCUITO\_T.dtsx

Transforma la información de acuerdo al tipo de flujo de tráfico (saliente y en tránsito) y calcula el tipo de llamadas eficaces, en este caso las que tomaron un circuito.

TH\_LLAMADAS\_PRESENTADAS\_RDSI\_I.dtsx

TH\_LLAMADAS\_PRESENTADAS\_RDSI\_A.dtsx

TH\_LLAMADAS\_PRESENTADAS\_RDSI\_D.dtsx

TH\_LLAMADAS\_PRESENTADAS\_RDSI\_T.dtsx

Transforma la información de acuerdo al tipo de flujo de tráfico (interno, entrante, saliente y en tránsito) y calcula el tipo de llamadas eficaces, en este caso las llamadas presentadas por abonados RDSI.

TH\_LLAMADAS\_MARCACION\_INCOMPLETA\_I.dtsx

TH\_LLAMADAS\_MARCACION\_INCOMPLETA\_A\_D\_T.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, entrante, saliente y en tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas con marcación incompleta.

TH\_LLAMADAS\_PREFIJOS\_INCORRECTOS\_I\_A.

Transforma la información de acuerdo al tipo de flujo (interno y entrante) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas con prefijos incorrectos.

TH\_LLAMADAS\_MANIOBRAS\_PROHIBIDAS.dtsx

Transforma la información de acuerdo al tipo de flujo (llegada, salida, tránsito) de tráfico y

calcula el tipo de llamadas ineficaces, en este caso las llamadas con marcación incompleta.

**TH\_LLAMADAS\_COLGADO\_ENTRE\_MARCACION\_Y\_CONEXION.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que colgaron entre la marcación y conexión.

**TH\_LLAMADAS\_ABONADOS\_FUERA\_SERVICIO.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas a abonados fuera de servicio.

**TH\_LLAMADAS\_INDICATIVOS\_NO\_AUTORIZADOS.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas a indicativos no autorizados.

**TH\_LLAMADAS\_DURANTE\_TIMBRE.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que colgaron durante el timbre.

**TH\_LLAMADAS\_ERROR\_MARCACION.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas con error de marcación.

**TH\_LLAMADAS\_ABONADOS\_OCUPADOS.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso llamadas a abonados ocupados.

**TH\_LLAMADAS\_AUSENCIA\_LLAMADO.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que no fueron contestadas.

**TH\_LLAMADAS\_OTRAS\_CAUSAS.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas por causas diversas.

**TH\_LLAMADAS\_BLOQUEO\_URA.dtsx**

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasaron por bloqueo en la URA.

#### TH\_LLAMADAS\_FALTA\_AUXILIARES.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por falta de auxiliares.

#### TH\_LLAMADAS\_FALLOS\_HARDWARE.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por falta de auxiliares.

#### TH\_LLAMADAS\_BLOQUEO\_URA\_FALTA\_ITINERARIOS.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por bloqueo de URA falta de itinerarios.

#### TH\_LLAMADAS\_BLOQUEO\_RED.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por bloqueo de red.

#### TH\_LLAMADAS\_REGULACION\_COMANDO.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por regulación de comando.

#### TH\_LLAMADAS\_URA\_LLDOR.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por la URA del abonado llamador.

#### TH\_LLAMADAS\_OTRAS\_CAUSAS\_CONMUTADOR.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por diversas causas del conmutador.

#### TH\_LLAMADAS\_RECHAZO\_XSEPARACION.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por rechazo al hacer la separación.

#### TH\_LLAMADAS\_DIRECCION\_SATURADA.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por dirección saturada.

#### TH\_LLAMADAS\_AUSENCIA\_INVITACION.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por ausencia de invitación a transmitir.

#### TH\_LLAMADAS\_ERROR\_SANALIZACION.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por error de señalización de registrador.

#### TH\_LLAMADAS\_RECEPCION\_SENAL\_CONGESTION.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por recepción de señal de congestión de CDC distante.

#### TH\_LLAMADAS\_ERROR\_LINEA.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por error de señalización de línea.

#### TH\_LLAMADAS\_SENAL\_INCOHERENTE.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por recepción de señal incoherente.

#### TH\_LLAMADAS\_DOBLE\_TOMA.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por doble toma.

#### TH\_LLAMADAS\_FILTRADO.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por filtrado.

#### TH\_LLAMADAS\_REDUCCION\_CONGESTION.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por reducción de congestión de señalización.

#### TH\_LLAMADAS\_RABASAMIENTO\_TEMPORIZACION.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por rebasamiento de temporización.

#### TH\_LLAMADAS\_FALLOS\_SENALIZACION\_LINEA.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por fallos de señalización de línea.

#### TH\_LLAMADAS\_CAUSA\_SERVIDOR.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por causa servidor.

#### TH\_LLAMADAS\_CAUSA\_NO\_ESPECIFICADA.dtsx

Transforma la información de acuerdo al tipo de flujo (interno, llegada, salida, tránsito) de tráfico y calcula el tipo de llamadas ineficaces, en este caso las llamadas que fracasan por causas no especificadas.

Los paquetes mencionados anteriormente almacenan los datos en la tabla de hechos th\_llamadas.

#### TD\_FECHA\_LLAMADAS.dtsx

Extrae las fechas de la tabla de hechos tipo de llamadas para transformarlas en una jerarquía del tiempo.

### 7.3.1.1.2. Población de datos

La población consiste en cargar la información ya transformada, lista para ser almacenada en la base definitiva del *Data Warehouse* en este caso a la base DWH.

Para esto se consideró:

- La frecuencia con la que se cargan los datos
- Confiabilidad
- Especificación de que dimensiones deben mantener un historial (retrospección), para truncar o añadir datos que se estén poblando ese momento.
- Data Errónea.

#### 7.3.1.1.2.1. Población hechos tráfico telefónico

Como regla general se estable que debido a que el procedimiento corre dos veces al mes se borre la información que se pobló en la primera corrida y cargue los datos confiables en la segunda ejecución en la que los datos han sido modificados.

**POBLACION\_HECHOS\_TRAFICO.BAT.-** Este archivo .bat llama a todos los paquetes .dtsx de población de las tablas de hechos del tráfico telefónico. Contiene los siguientes paquetes:

FACTTRAFICO\_SALIENTE.dtsx

FACTTRAFICO\_LOCAL.dtsx

FACTRAFICO\_LOCAL\_DETALLADO.dtsx

FACTRAFICO\_ENTRANTE.dtsx

Estos paquetes se encargan de poblar la tabla de hechos de tráfico saliente, local, local detallado y entrante de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### **7.3.1.1.2.2. Población hechos conmutación**

**POBLACION\_HECHOS\_CONMUTACION.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

FACTAUXILIARES.dtsx

FACTHACES.dtsx

FACTTRAFICO\_ERLANGS

FACTURA.dtsx

FACTLLAMADAS.dtsx

Estos paquetes se encargan de poblar la tabla de hechos de Auxiliares, Haces, Tráfico en Erlangs, URAs y llamadas de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### **7.3.1.1.2.3. Población dimensiones tráfico telefónico**

**POBLACION\_DIMENSIONES\_TRAFICO.BAT.-** Este archivo .bat contiene los siguientes paquetes con extensión .dtsx:

DIMCLIENTE\_TRAFICO.dtsx

Se encarga de poblar la tabla de dimensión cliente tráfico de la base de datos DWH con la información ya transformada en el TANQUE, pero considerando que se debe guardar el historial de cambios de nombre o categoría y subir los datos que no existan.

#### DIMUBICACION\_NO\_REGISTRADOS.dtsx

Se encarga de poblar la tabla de dimensión ubicación de la base de datos DWH con la información ya transformada en el TANQUE, considerando que se debe mantener el historial de los lugares a donde se realizó la llamada.

#### DIMUBICACION\_MODIFICAR.dtsx

Se encarga de poblar la tabla de dimensión ubicación de la base de datos DWH con la información ya transformada en el TANQUE, considerando que pueden variar los nombres de países o ciudades a los cuales hay que actualizar debido a que no se lleva un historial de cambios porque no es necesario.

#### DIMDESTINO\_SALIDA.dtsx

Se encarga de poblar la tabla de dimensión destino del tráfico de salida de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMCONTACTO\_EXTERNO.dtsx

Se encarga de poblar la tabla de dimensión contacto externo de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMTIPO\_LOCAL.dtsx

Se encarga de poblar la tabla de dimensión local de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMDESTINO\_ENTRADA.dtsx

Se encarga de poblar la tabla de dimensión destino del tráfico de entrada de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMFECHA\_TRAFICO.dtsx

Se encarga de poblar la tabla de dimensión fecha de la base de datos DWH con la información ya transformada en el TANQUE, considerando subir los datos que no existan ya en el *Data Warehouse*.

### **7.3.1.1.2.4. Población dimensiones conmutación**

**POBLACION\_DIMENSIONES\_CONMUTACION.BAT.-** Este archivo contiene los siguientes paquetes con extensión .dtsx:

#### DIMCENTRALES.dtsx

Se encarga de poblar la tabla de dimensión centrales de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMFECHA\_TRAFICO.dtsx

Se encarga de poblar la tabla de dimensión fecha de la base de datos DWH con la información ya transformada en el TANQUE, pero considerando que debe subir sólo las fechas que no existan.

#### DIMAUXILIARES.dtsx

Se encarga de poblar la tabla dimensión auxiliares de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMDESCRIPCION\_AUXILIARES.dtsx

Se encarga de poblar la tabla de dimensión descripción auxiliares de la base de datos DWH con la información ya transformada en el TANQUE.

#### DIMDESCRIPCION\_HACES.dtsx

Se encarga de poblar la tabla de dimensión descripción haces de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMHACES.dtsx

Se encarga de poblar la tabla de dimensión haces de la base de datos DWH con la información ya transformada en la Base de Datos TANQUE.

#### DIMLLAMADA\_URA.dtsx

Se encarga de poblar la tabla de dimensión llamada URA de la base de datos DWH con la información ya transformada en el TANQUE.

#### DIMURA.dtsx

Se encarga de poblar la tabla de dimensión URA de la base de datos DWH con la información ya transformada en el TANQUE.

#### DIMTIPO\_FLUJO.dtsx

Se encarga de poblar la tabla de dimensión tipo flujo de la base de datos DWH con la información ya transformada en el TANQUE.

#### DIMTIPO\_LLAMADA.dtsx

Se encarga de poblar la tabla de dimensión tipo llamada de la base de datos DWH con la información ya transformada en el TANQUE.

### **7.3.1.1.3. Control de integridad**

Este control permite mantener la integridad entre las tablas de hechos y las tablas de dimensiones (es decir que todos los registros de la tabla de hechos existan en las dimensiones), la integridad se hace en los datos que se encuentran en la base de datos DWH.

Existen dos archivos .bat para controlar la integridad y se encuentran en el directorio C:\WORKDIR\DTSS\TRAFICO\_TELEFONICO\BATS.

El primero es INTEGRACION\_CONMUTACION.BAT y el segundo INTEGRACION\_TRAFICO.BAT

#### **7.3.1.1.3.1. Integración conmutación**

**INTEGRACION\_CONMUTACION.BAT.-** Contiene la integridad de las siguientes dimensiones:

- DIMCENTRALES
- DIMAUXILIARES
- DIMDESCRIPCION\_AUXILIARES
- DIMDESCRIPCION\_HACES
- DIMHACES
- DIMLLAMADA\_URA
- DIMITIPO\_FLUJO
- DIMITIPO\_LLAMADA
- DIMURA

#### **7.3.1.1.3.2. Integración tráfico telefónico**

**INTEGRACION\_TRAFICO.BAT.-** Contiene la integridad de las siguientes dimensiones:

- DIMCLIENTE\_TRAFICO
- DIMUBICACION
- DIMCENTRALES
- DIMCONTACTO\_EXTERNO
- DIMDESTINO\_SALIDA
- DIMITARIFA
- DIMDESTINO\_ENTRADA.

#### **7.3.1.1.4. Secuencia de ejecución**

La ejecución de estos procesos se los realiza cronológicamente siguiendo un orden específico, debido a que existen dependencias y dimensiones compartidas en las cuales los procesos para la carga de estas dimensiones (Dtsx) está diseñado para

truncar la tabla y cargar los datos ó sólo para añadirlos. Por estas razones para cumplir con el objetivo, se deberá seguir el siguiente orden:

#### **7.3.1.1.4.1. Proceso de conmutación**

##### **EXTRACCION\_DICCIONARIOS\_CONMUTACION.BAT**

- Diccionario.Td\_Auxiliares-Dwhtanque.Td\_Auxiliares.dtsx
- Diccionario.Td\_Descripcion\_Auxiliares-Dwhtanque.Td\_Descripcion\_Auxiliares.dtsx
- Diccionario.Td\_Descripcion\_HACES-Dwhtanque.Td\_Descripcion\_HACES.dtsx
- Diccionario.Td\_Haces-Dwhtanque.Td\_Haces.dtsx
- Diccionario.Td\_Tipo\_Llamadas-Dwhtanque.Td\_Tipo\_Llamadas.dtsx
- Diccionario.Td\_URA-Dwhtanque.Td\_URA.dtsx
- Diccionario.Td\_Central-Dwhtanque.Td\_Centrales.dtsx
- Diccionario.Td\_Tipos\_Flujo-Dwhtanque.Td\_Tipos\_Flujo.dtsx
- Diccionario.Td\_Llamada\_URA-DWHTANQUE.TD\_LLAMADA\_URA.dtsx

##### **EXTRACCION\_CONMUTACION\_AUXILIARES.BAT**

- Etapa.Clase\_A-Dwhtanque.Auxiliares.dtsx
- Th\_Auxiliares\_Rgf\_Aq1.dtsx
- Th\_Auxiliares\_Rgf\_Ccf\_V23.dtsx
- Td\_Fecha\_Trafico.dtsx

##### **EXTRACCION\_CONMUTACION\_TRAFICO\_ERLANGS.BAT**

- Etapa.Clase\_C-Dwh.Clase\_C.dtsx
- Th\_Trafico\_Erlangs.dtsx
- Td\_Fecha\_Trafico.dtsx

##### **EXTRACCION\_CONMUTACION\_HACES.BAT**

- Etapa.Clase\_Fb-DWHTANQUE.CLASE\_FB.dtsx
- Th\_Haces\_ENTRANTES\_FPA.dtsx
- Th\_Haces\_SALIENTES\_FPD.dtsx
- Th\_Haces\_Mixtos\_FPA.dtsx
- Th\_Haces\_Mixtos\_FPD.dtsx
- Th\_Haces\_Entrantes\_FRA.dtsx
- Th\_Haces\_Entrantes\_FEA.dtsx
- Th\_Haces\_Entrantes\_FOA.dtsx
- Th\_Haces\_Salientes\_FRD.dtsx
- Th\_Haces\_Salientes\_FED.dtsx
- Th\_Haces\_Salientes\_FOD.dtsx
- Th\_Haces\_Mixtos\_FRA.dtsx
- Th\_Haces\_Mixtos\_FRD.dtsx
- Th\_Haces\_Mixtos\_FEA.dtsx
- Th\_Haces\_Mixtos\_FED.dtsx
- Th\_Haces\_Mixtos\_FOA.dtsx

- Th\_Haces\_Mixtos\_FOD.dtsx
- Th\_Haces\_FQ1.dtsx
- Th\_Haces\_FS1.dtsx
- Th\_Haces\_PRS1.dtsx
- Td\_Fecha\_Haces.dtsx

## **EXTRACCION\_CONMUTACION\_URA.BAT**

- Etapa.clase\_r-Dwh.Clase\_R.dtsx
- Th\_Ura.dtsx
- Td\_Fecha\_Trafico.dtsx

## **EXTRACCION\_CONMUTACION\_CLASE\_T.BAT**

- Etapa.Clase\_T-Dwh.Tanque.Clase\_T.dtsx
- Th\_Llamadas\_Presentadas\_I.dtsx
- Th\_Llamadas\_Presentadas\_A.dtsx
- Th\_Llamadas\_Presentadas\_D.dtsx
- Th\_Llamadas\_Presentadas\_T.dtsx
- Th\_Llamadas\_Completadas\_I.dtsx
- Th\_Llamadas\_Completadas\_D.dtsx
- Th\_Llamadas\_Tratadas\_I.dtsx
- Th\_Llamadas\_Tratadas\_A.dtsx
- Th\_Llamadas\_Tratadas\_D.dtsx
- Th\_Llamadas\_Tratadas\_T.dtsx
- Th\_Llamadas\_Toma\_Circuito\_D.dtsx
- Th\_Llamadas\_Toma\_Circuito\_T.dtsx
- Th\_Llamadas\_Presentadas\_Rdsi\_I.dtsx
- Th\_Llamadas\_Presentadas\_Rdsi\_A.dtsx
- Th\_Llamadas\_Presentadas\_Rdsi\_D.dtsx
- Th\_Llamadas\_Presentadas\_Rdsi\_T.dtsx
- Th\_Llamadas\_Marcacion\_Incompleta\_I.dtsx
- Th\_Llamadas\_Marcacion\_Incompleta\_A\_D\_T.dtsx
- Th\_Llamadas\_Maniobras\_Prohibidas.dtsx
- Th\_Llamadas\_Prefijos\_Incorrectos\_I\_A.
- Th\_Llamadas\_Colgado\_Entre\_Marcacion\_Y\_Conexion.dtsx
- Th\_Llamadas\_Abonados\_Fuera\_Servicio.dtsx
- Th\_Llamadas\_Indicativos\_No\_Autorizados.dtsx
- Th\_Llamadas\_Durante\_Timbre.dtsx
- Th\_Llamadas\_Error\_Marcacion.dtsx
- Th\_Llamadas\_Abonados\_Ocupados.dtsx
- Th\_Llamadas\_Ausencia\_Llamado.dtsx
- Th\_Llamadas\_Otras\_Causas.dtsx
- Th\_Llamadas\_Bloqueo\_URA.dtsx
- Th\_Llamadas\_Falta\_Auxiliares.dtsx
- Th\_Llamadas\_Fallos\_Hardware.dtsx
- Th\_Llamadas\_Bloqueo\_URA\_Falta\_Itinerarios.dtsx
- Th\_Llamadas\_Bloqueo\_Red.dtsx
- Th\_Llamadas\_Regulacion\_Comando.dtsx
- Th\_Llamadas\_URA\_Lldor.dtsx
- Th\_Llamadas\_Otras\_Causas\_Conmutador.dtsx
- Th\_Llamadas\_Rechazo\_Xseparacion.dtsx

- Th\_Llamadas\_Direccion\_Saturada.dtsx
- TH\_LLAMADAS\_Ausencia\_Invitacion.dtsx
- Th\_Llamadas\_Error\_Senalizacion.dtsx
- Th\_Llamadas\_Recepcion\_Senal\_Congestion.dtsx
- Th\_Llamadas\_Error\_Linea.dtsx
- Th\_Llamadas\_Senal\_Incoherente.dtsx
- Th\_Llamadas\_Doble\_Toma.dtsx
- Th\_Llamadas\_Filtrado.dtsx
- Th\_Llamadas\_Reducccion\_Congestion.dtsx
- Th\_Llamadas\_Rabasamiento\_Temporizacion.dtsx
- Th\_Llamadas\_Fallos\_Senalizacion\_Linea.dtsx
- Th\_Llamadas\_Causa\_Servidor.dtsx
- Th\_Llamadas\_Causa\_No\_Especificada.dtsx
- Td\_Fecha\_Llamadas.dtsx

### **POBLACION\_HECHOS\_CONMUTACION.BAT**

- FactAuxiliares.dtsx
- FactHaces.dtsx
- FactTrafico\_erlang.dtsx
- FactUra.dtsx
- FactLlamadas.dtsx

### **POBLACION\_DIMENSIONES\_CONMUTACION.BAT**

- Dimcentrales.dtsx
- Dimfecha\_Trafico.dtsx
- Dimauxiliares.dtsx
- Dimdescripcion\_Auxiliares.dtsx
- Dimdescripcion\_Haces.dtsx
- Dimhaces.dtsx
- Dimllamada\_Ura.dtsx
- Dimura.dtsx
- Dimtipo\_Flujo.dtsx
- Dimtipo\_Llamada.dtsx

### **INTEGRACION\_CONMUTACION.BAT**

- Integracion\_Centrales.dtsx
- Integracion\_Dimaxiliares.dtsx
- Integracion\_Dimdescripcion\_Auxiliares.dtsx
- Integracion\_Dimdescripcion\_Haces.dtsx
- Integracion\_Dimhaces.dtsx
- Integridad\_Dimllamada\_Ura.dtsx
- Integracion\_Dimtipo\_Flujo.dtsx
- Integracion\_Dimtipo\_Llamada.dtsx
- Integracion\_Dim\_Ura.dtsx

### 7.3.1.1.4.2. Proceso trafico telefónico

#### EXTRACCION\_DICCIONARIOS\_TRAFICO\_TELEFONICO.BAT

- Diccionario.Td\_Destino\_Salida-Dwhtanque.Td\_Destino\_Salida.dtsx
- Diccionario.Td\_Tarifa-Dwhtanque.Td\_Tarifa.dtsx
- Diccionario.Td\_Destino\_Entrada-Dwhtanque.Td\_Destino\_Entrada.dtsx
- Diccionario.Servicio\_3digitos-Dwhtanque.Servicio\_3digitos.dtsx
- Dicionario.Tipo\_Local-Dwhtanque.Td\_Tipo\_Local.dtsx
- Diccionario.Directorio\_1800.Dwhtanque.Directorio\_1800.dtsx

#### EXTRACCION\_TRAFICO\_SALIENTE.BAT

- Limpiar\_Directorio.dtsx
- Etapa.Pais-Dwhtanque.Pais.dtsx
- Etapa.Ciudad-Dwhtanque.Ciudad.dtsx
- Etapa.Sregiona-Dwhtanque.Sregiona.dtsx
- Etapa.Snaciona-Dwhtanque.Snaciona.dtsx
- Etapa.Sinterna-Dwhtanque.Sinterna.dtsx
- Etapa.Scelular-Dwhtanque.Scelular.dtsx
- Etapa.Sespecia-Dwhtanque.Sespecia.dtsx
- Qs36f.Telarc-Dwhtanque.Telar\_Mes.
- Qs36f.Tarar7-Dwhtanque.Tarar7.dtsx
- Etapa.Slocal-Dwhtanque.Slocal.dtsx
- Trafico\_1800.dtsx
- Trafico\_1800\_Noregistrado.dtsx
- Trafico\_Local\_Servicios.dtsx
- Trafico\_Local.dtsx
- Th\_Trafico\_Regional\_Registrado\_Cliente.dtsx
- Th\_Trafico\_Regional\_Noregistrado\_Cliente.dtsx
- Th\_Trafico\_Nacional\_Registrado\_Cliente.dtsx
- Th\_Trafico\_Nacional\_Noregistrado\_Cliente.dtsx
- Th\_Trafico\_Internacional\_Registrado\_Cliente.dtsx
- Th\_Trafico\_Internacional\_Noregistrado\_Cliente.dtsx
- Th\_Trafico\_Celular\_Registrado\_Cliente.dtsx
- Th\_Trafico\_Celular\_Noregistrado\_Cliente.dtsx
- Th\_Trafico\_1800\_Registrado\_Cliente.dtsx
- Th\_Trafico\_1800\_Noregistrado\_Cliente.dtsx
- Th\_Trafico\_Servicios\_Registrado\_Cliente.dtsx
- Trafico\_Local\_Servicio\_Noregistrado\_Cliente.dtsx
- Th\_Trafico\_Local\_Detallado\_Registrado\_Cliente.dtsx
- Th\_Trafico\_Local\_Detallado\_Noregistro\_Ori\_Des.dtsx
- Th\_Trafico\_Local\_Detallado\_Noregistrado\_Origen.dtsx
- Th\_Trafico\_Local\_Detallado\_Noregistrado\_Destino.dtsx
- Th\_Trafico\_Local\_Registrado\_Cliente.dtsx
- Th\_Trafico\_Consumo\_Local\_Noregistrado\_Cliente.dtsx
- Td\_Ubicacion.dtsx
- Td\_Ubicacion\_Noreg\_Ciudad.dtsx
- Td\_Contacto\_Externo\_Normal.dtsx
- Td\_Contacto\_Externo\_3digitos.dtsx

- Td\_Contacto\_Externo\_1800.dtsx
- Td\_Contacto\_Externo\_1800\_Nodefinido.dtsx
- Td\_Cliente\_Trafico.dtsx
- Td\_Cliente\_Trafico\_Local.dtsx
- Td\_Cliente\_Trafico\_Local\_Detallado\_Origen.dtsx
- Td\_Cliente\_Trafico\_Local\_Detallado\_Destino.dtsx
- Td\_Fecha\_Trafico.dtsx
- Td\_Fecha\_Trafico\_Local.dtsx
- Td\_Fecha\_Trafico\_Local\_Detallado.dtsx

### **EXTRACCION\_TRAFICO\_ENTRANTE.BAT**

- Etapa.Enaamm-Dwhtanque.Enaciona. dtsx
- Etapa.Enaamm-Dwhtanque.Eregiona. dtsx
- Etapa.Eiaamm-Dwhtanque.Einterna. dtsx
- Etapa.Ecaamm-Dwhtanque.Ecelular. dtsx

### **EXTRACCION\_TH\_Y\_TD\_ENTRANTE.BAT**

- Th\_Entrante\_Nacional.dtsx
- Th\_Entrante\_Nacional\_No\_Reg.dtsx
- Th\_Entrante\_Regional.dtsx
- Th\_Entrante\_Regional\_No\_Reg.dtsx
- Th\_Entrante\_Internacional.dtsx
- Th\_Entrante\_Internacional\_No\_Reg.dtsx
- Th\_Entrante\_Celular.dtsx
- Th\_Entrante\_Celular\_No\_Reg.dtsx
- Td\_Cliente\_Trafico\_Entrante.dtsx
- Td\_Fecha\_Trafico\_Entrante.dtsx
- Td\_Ubicacion\_Entrante.dtsx
- Td\_Contacto\_Externo\_Trafico\_Entrante.dtsx

### **POBLACION\_HECHOS\_TRAFICO.BAT**

- Factrafico\_Saliente.dtsx
- Factrafico\_Local.dtsx
- Factrafico\_Local\_Detallado.dtsx
- Factrafico\_Entrante.dtsx

### **POBLACION\_DIMENSIONES\_TRAFICO.BAT**

- Dimcliente\_Trafico.dtsx
- Dimubicacion\_No\_Registrados.dtsx
- Dimubicacion\_Modificar.dtsx
- Dimdestino\_Salida.dtsx
- Dimcontacto\_Externo.dtsx
- Dimtipo\_Local.dtsx
- Dimdestino\_Entrada.dtsx
- Dimfecha\_Trafico.dtsx

## **INTEGRACION\_TRAFICO.BAT**

- Integridad\_Td\_Cliente\_Trafico.dtsx
- Integridad\_Dimubicacion.dtsx
- Integridad\_Dimcentrales.dtsx
- Integridad\_Dimcontacto\_Externo.dtsx
- Integridad\_Dimdestino\_Salida.dtsx
- Integridad\_Dimcliente\_Trafico\_Entrada.dtsx
- Integridad\_Dimubicacion\_Entrante.dtsx
- Integridad\_Dimcentrales\_Entrada.dtsx
- Integridad\_Dimtarifa\_Entrada.dtsx
- Integridad\_Dimtarifa\_Salida.dtsx
- Integridad\_Dimdestino\_Entrada.dtsx
- Integridad\_Dimcontacto\_Externo\_Entrada.dtsx

Todas las operaciones necesarias para construir el *Data Warehouse* que se especificaron son procesos automáticos que no necesitan intervención de los desarrolladores a excepción de que tenga que realizarse cambios en la estructura.

### **7.3.1.1.5. Cubos**

Un cubo de información es una estructura multidimensional que está basado en la analogía de que sus caras son las distintas dimensiones del negocio (fecha, ubicación, cliente) y sus medidas (duración llamada, número contactos diferentes, etc.). Los cubos están especialmente diseñados para responder consultas analíticas de manera rápida, eficiente y flexible.

#### **7.3.1.1.5.1. Dimensiones**

Las dimensiones de un cubo representan las distintas categorías para analizar los datos del negocio como: fechas, ubicación, cliente, contacto externo, etc. Estas categorías permiten obtener la información en diferentes combinaciones, seleccionando las variables que interesan en una consulta específica.

##### **7.3.1.1.5.1.1. Jerarquías asociadas a una dimensión**

Las dimensiones pueden contener una jerarquía de niveles, permitiendo obtener información con distintos grados de detalle.

Drill Down: bajar a un nivel detallado de la información.

Drill up: subir a un nivel agregado de la información

#### **7.3.1.1.5.2. Medidas**

Las Medidas, son datos numéricos de interés para los usuarios. Algunas medidas comunes son número de llamadas, duración, etc.

#### **7.3.1.1.5.3. Modelo de cubos**

Los cubos están desarrollados en un modelo basado en el esquema estrella.

#### **7.3.1.1.5.4. Diseño de almacenamiento de cubos**

El almacenamiento de los cubos está desarrollado con el Método MOLAP.

#### **7.3.1.1.5.5. Generación de cubos y dimensiones**

Los cubos se generan de acuerdo a su prioridad. Esto se lleva a cabo con el siguiente paquete con extensión .dtsx:

Existe un archivo .bat PROCESO\_CUBOS\_DIMENSIONES.BAT y se encuentra en el directorio C:\WORKDIR\cubos\PROCESAMIENTO\_CUBOS\PROCESAMIENTO\_CUBOS.dtsx que ejecuta el siguiente dtsx::

PROCESAMIENTO\_CUBOS.dtsx

Permite procesar las dimensiones y los cubos de Conmutación y Tráfico Telefónico.

Para poder realizar cambios en los cubos tenemos que revisar las vistas de origen de datos en base a la conexión previamente realizada en *Analysis Services*, las conexiones se realiza a la base de datos DWH. Se utiliza la misma conexión para todas las vistas para que nos permita compartir las dimensiones.

DWH-TRAFICO-TELEFONICO.dsv

En esta vista fuente del origen de datos se establece las relaciones de las tablas de hechos

“FACTTRAFICO\_ENTRANTE, FACTTRAFICO\_SALIENTE, FACTTRAFICO\_LOCAL, FACTTRAFICO\_LOCAL\_DETALLADO” a cada una de las dimensiones, de acuerdo al modelo lógico establecido anteriormente.

DWH-CONMUTACION.dsv

En esta vista fuente del origen de datos se establece las relaciones de las tablas de hechos “FACTAUXILIARES, FACTHACES, FACTTRAFICO\_ERLANGS, FACTURA, FACTLLAMADAS” a cada una de las dimensiones, de acuerdo al modelo lógico establecido anteriormente.

### **7.3.1.2. Proceso para administrar la base de datos**

Es importante considerar el espacio físico que va a utilizar los datos por eso es recomendable dedicar tiempo a la organización y administración de la base de datos para evitar futuros inconvenientes. Uno de los problemas que se puede presentar es que el fichero Log crezca indefinidamente hasta hacerse tan grande que consuma todo el disco duro, llegando en el peor de los casos a llenarlo con el siguiente error:

*Error: 1105, Severity: 17, State: 2: Could not allocate space for the transaction log for database 'DWH' because the device is full. Dump the log or enlarge the device to create more space.*

Debido a que en el *Data Warehouse* no transaccional no es necesario llevar un registro de todas las transacciones que se realiza en la base de datos. Para evitar los inconvenientes antes mencionados se crearon procesos para borrar el contenido de los log, pero antes de esto es necesario hacer un respaldo de la base de datos si es la primera vez que va a truncar los log, caso contrario se borra directamente.

#### **7.3.1.2.1. Proceso para crear respaldo y truncar log.**

##### **RESPALDO\_DWH.BAT**

Primero realiza un *checkpoint*, es decir, pasa toda la paginación que se encuentra en la memoria virtual al disco físico, luego indica el directorio donde se va a guardar el respaldo (G:\respaldos), hace un respaldo de la base y del log, al final trunca el log a 1MB, reduciendo el espacio en disco.

## **BORRAR\_LOG\_DWH.BAT**

Trunca o borra el contenido del fichero `dwh_log.lmf` de la base de datos DWH. Ya no es necesario hacer el respaldo de la base de datos, se lo hace de esta manera por el tiempo y además el objetivo es liberar espacio

## **BORRAR\_LOG\_TANQUE.BAT**

Trunca o borra el contenido del fichero `Tanque_log.lmf` de la base de datos TANQUE.

### **7.3.1.2.2. Restaurar base de datos**

Para restaurar una copia de seguridad seleccionamos la base de datos DWH con el botón derecho pulsamos en la opción “Todas las tareas” -> Restaurar base de datos. En esta ventana, seleccionamos “desde dispositivo”, pulsamos el botón “dispositivo” y en la siguiente pantalla pulsamos “agregar” y seleccionamos el fichero que se encuentra en el directorio `G:\respaldo\logdwh.bak`. En opciones de restauración escogemos Completa. En la siguiente ficha de esta pantalla “opciones”, seleccionamos las opciones de restauración de la copia:

- Forzar restauración sobre la base de datos existente, forzamos a que se restaure la copia sobrescribiendo el fichero existente.

### **7.3.1.3. Visualización**

Se pueden visualizar los cubos mediante el uso de: Excel ó *XLCUBED* usando conexiones *ODBC*, y por medio de *Accelerator Report Portal* en donde tenemos reportes definidos.

## **REFERENCIAS**

---

<sup>1</sup> ALCATEL, “Gestión de las Observaciones de Carga y Tráfico”, 2002.

## CONCLUSIONES

- Para cualquier proyecto que se pretenda implementar es importante definir la metodología que se va utilizar en un inicio para poder convertir los requerimientos en *software* como primer paso al desarrollo de una bodega de datos consistente, histórica e integrada.
- Las metodología se adaptará a las necesidades cambiantes de la empresa y describirá actividades para construir las soluciones de la tecnología de la información, la misma que se encargará de mantener un control de datos dentro de la empresa en diferentes niveles empresariales, a través del uso de la Inteligencia de Negocios, Sistemas de Información Ejecutiva y Sistemas de Soporte de decisiones, con el fin de tomar mejores decisiones.
- Previo a la implementación del *Data Warehouse* se hará necesario el desarrollo del modelo conceptual, lógico y físico para poder permitir a la gente de negocios describir sus requerimientos a través de esquemas utilizados para el efecto, basados en conceptos de dimensiones, hechos, agregación, jerarquía y niveles.
- La correcta especificación de requerimientos será el paso previo al desarrollo del modelo lógico y físico. En la implementación será necesaria y vital la integración de los datos, esto es su depuración, así como la estandarización de los datos. Todo esto para cumplir con los objetivos *Data Warehouse*.
- Los *Data marts* serán considerados como una solución que permitirá el ahorro de recursos económicos al momento de implementar una solución, que nos proporcione un conocimiento en primera instancia de los resultados positivos ó no de un proyecto *Data Warehouse*. Los resultados serán positivos en caso de que dicho proyecto haya sido bien llevado. Sin embargo se debe tener cuidado al integrar todos los módulos, pues puede existir *data* repetida que consuma espacio de almacenamiento innecesario. Por otra parte, se tomarán en cuenta aspectos como espacio en disco y la velocidad con la

que se ejecuten las consultas, para elegir cualquiera de los modelos de datos *OLAP*, *MOLAP*, *ROLAP*. Esto dependerá de la empresa y los recursos que posea, así como sus prioridades y objetivos.

- La forma como se presenten los reportes y la interacción con el usuario definirá de gran forma la validez de la información. Esto reflejará si el *Data Warehouse* ha cumplido o no con su cometido, ya que el usuario final y la gente de negocios, es a quienes irá dirigido el proyecto y quienes decidirán si el mismo tuvo éxito o no.

## RECOMENDACIONES

- Establecer reuniones previas con la gente de negocios para que conozcan los objetivos que se logran con el desarrollo de un *Data Warehouse* en sus empresas y puedan ayudar a establecer los requerimientos de información.
- Concienciar a la gente de negocios sobre la disponibilidad de tiempo que se debe tener para analizar la información que se debe cargar al *DWH*. Se deberá especificar sólo información necesaria debido a que su estructura ocupa mucho espacio y más aún lo haría si es que existiera información no útil.
- La empresa en donde se desarrolle un *Data Warehouse* debe enfocarse en hacer conocer a los desarrolladores cuales son los indicadores, reglas de negocios que influyen en la toma de decisiones.
- Que se defina con claridad los niveles de granularidad necesarios.
- Crear tabla de hechos con información más sumariada para aumentar la rapidez en la ejecución de una consulta y sólo para casos especiales tener las tablas de hecho con la información mucho más detallada.
- Fomentar un compromiso del equipo de trabajo es decir desarrolladores y empleados de la empresa donde se va a desarrollar para poder establecer la descripción y las fuentes de los datos.
- Los desarrolladores deben de realizar un análisis del espacio en disco que utilizara la información que se va almacenar para evitar inconvenientes en las extracciones.
- Una vez implementado la bodega de datos se debe realizar constantemente mantenimiento a la base de datos (liberar espacio).
- Tener un conocimiento previo de la herramienta que se va a utilizar independientemente de cual se seleccione para poder hacer un mejor uso de la misma, dicha selección estará basada en las necesidades de la empresa, recursos económicos con los que se cuente ó simplemente en el cumplimiento de los estándares ya establecidos.
- Se debe considerar la retrosección de los objetos *DWH* que agreguen valor al negocio.
- Empezar la construcción del *DWH* cuando la metadata esté aprobada.

## ANEXO I METADATA

### 1.1. Tráfico Telefónico

#### 1.1.1 Tráfico Entrante

Las tablas de los registros detallados de las llamadas tienen los mismos nombres de campos, por esa razón en la fuente se especifica primero todas las tablas de las cuales se van a extraer y al final el nombre del campo.

Cuando se construye la tabla de hecho es una regla el agrupar todos los campos que se extraen es decir hacer lo siguiente:

#### **TH\_TRAFICO\_ENTRANTE**

Group by ENTDES, ENTGEN, TELNOM, TARDE4, ENT TIP, ENTOPE,  
ENTTAR,ENTFEC,ENTHOR

Join: qs36f.telarc.tel tel=trafico.entdes

| CAMPO        | TIPO            | FUENTE                                                                | TRANSFORMACION                                                                                           | DESCRIPCION                                                 |
|--------------|-----------------|-----------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
| CEN_COD      | SMALLINT        | ETAPA.ENAAMM,<br>ECAAMM,EIAAMM.<br>ENTCEN                             | Igual (conversión implícita del dato)                                                                    | Código de la central que recibe la llamada                  |
| CLI_COD      | INTEGER         | ETAPA.ENAAMM,<br>ECAAMM,EIAAMM.<br>ENTDES                             | Substr(entdes,<br>length(entdes)-<br>7,6),                                                               | Cliente que recibe la llamada                               |
| CLI_NOM      | VARCHAR<br>(30) | QS36F.TELARC.<br>TELNOM                                               | Ltrim(rtrim(telno<br>m))                                                                                 | Nombre del abonado que recibe llamada                       |
| CLI_CAT_COD  | SMALLINT        | QS36F.TELARC.<br>TELTAR                                               | Igual                                                                                                    | Código de la categoría                                      |
| ENT_COM      | VARCHAR<br>(5)  | ETAPA.ENAAMM,<br>ENTTIP,<br>ETAPA.ENAAMM,<br>ECAAMM,EIAAMM.<br>ENTOPE | Si es Regional ó Nacional ENTIP '+' ENTOPE; Si Es Celular ó Internacional se concatena la C ó I y SALOPE | Código compuesto para formar la jerarquía destino operadora |
| CON_COD      | VARCHAR<br>(15) | ETAPA.ENAAMM,<br>ECAAMM,<br>EIAAMM.ENTORI                             | Rtrim(ltrim(entori<br>)                                                                                  | Contacto externo, es quién realiza llamada                  |
| TAR_COD      | VARCHAR<br>(1)  | ETAPA.ECAAMM.ENTTAR                                                   | Igual, y para el resto de tráfico se coloca una 'N'                                                      | Código tarifa                                               |
| NUM_LLAMADAS | FLOAT           | ETAPA.ENAAMM,<br>ECAAMM,<br>EIAAMM.ENTORI                             | Count(entori)                                                                                            | Número de llamadas realizadas por contacto externo          |

|                 |              |                                                       |                                                                                                           |                                                        |
|-----------------|--------------|-------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|--------------------------------------------------------|
| MIN_CONSUMIDOS  | FLOAT        | ETAPA.ENAAMM,<br>ECAAMM,<br>EIAAMM.ENTDUR             | Round(sum(entdur)/60,2)                                                                                   | Minutos Consumidos                                     |
| MIN_REDONDEADOS | FLOAT        | ETAPA.EIAAMM<br>ENTDUI                                | SUM(ENTDUI)<br>en los otros casos<br>0                                                                    | Minutos redondeados segmento 6                         |
| FECHA           | DATE<br>TIME | ETAPA.ENAAMM,<br>ECAAMM,<br>EIAAMM.ENTFECH,<br>ENTHOR | Cast(convert(char,entfec, 101) +<br>cast(enthoras varchar(2))+ ':' +<br>'00' + ':' + '00' as<br>datetime) | Fecha y hora de consumo formato<br>aaaa/mm/dd:hh:mm:ss |
| PAI_COD         | INT          | No existe fuente no se puede identificar              | 58 regi.,nac, -1 el resto del destino                                                                     | Número del país                                        |
| CUI_COD         | INT          | ETAPA.ENAAMM.<br>ENTZON                               | Igual, 9 celu. Para el resto -1                                                                           | Número de zona                                         |

**Importante:** Es necesario que el nombre del cliente y la categoría estén en la tabla de hechos porque se necesita saber quién realizó la llamada y con qué categoría para que no altere los resultados (no perder datos) ya que cada mes puede existir cambios de nombre o categoría por esta razón cuando el abonado no exista en el telarc se debería realizar lo siguiente:

```
case when entdes>848800 and entdes<=848999 then 'BANCO DEL PICHINCHA'
'CENTRALILLA DE PLANEACION' when entdes=849950 then 'INTERNET DE ONNET'
when entdes=875755 then 'REPARTIDOR. 15 PAR100' else 'ABONADO NO DEFINIDO'
```

Para evitar los inconvenientes de la administración se realizó un archivo de Excel donde se encuentran los números telefónicos que no están registrados todavía como abonados pero si al realizar los joins (uniones), tampoco se encuentra el número telefónico el abonado quedará con nombre y categoría no definidos. Se ha definido esto por entrevistas con el Ing. Vintimilla e Ing. Marcel Vidal encargados de facturación y atención con el cliente respectivamente, quienes han manifestado que por motivos internos no se registran estos números como abonados.

### TD\_CENTRALES

| CAMPO   | TIPO            | FUENTE                         | TRANSFORMACION | DESCRIPCION               |
|---------|-----------------|--------------------------------|----------------|---------------------------|
| CEN_COD | SMALLINT        | DICCIONARIO.TD_CENTRAL.CEN_COD | Igual          | Código de la central      |
| CEN_DES | VARCHAR<br>(15) | DICCIONARIO.TD_CENTRAL.CEN_DES | Igual          | Descripción de la central |

**Importante:** Esta dimensión es compartida para todos los cubos mencionados en el entregable diseño de la base Multidimensional, pero se realiza una sola vez la extracción debido a que es generada mediante un diccionario de datos.

### TD\_DESTINO\_ENTRADA

| CAMPO       | TIPO         | FUENTE                         | TRANSFORMACION | DESCRIPCION                                                       |
|-------------|--------------|--------------------------------|----------------|-------------------------------------------------------------------|
| ENT_COM     | VARCHAR (5)  | DICCIONARIO.TD_DESTINO_ENTRADA | Igual          | Código compuesto del destino( destino, tipo conexión , operadora) |
| ENT_DES_COD | VARCHAR (1)  | DICCIONARIO.TD_DESTINO_ENTRADA | Igual          | Código del destino (N,R,I,C,I)                                    |
| ENT_DES_DES | VARCHAR (13) | DICCIONARIO.TD_DESTINO_ENTRADA | Igual          | Descripción del destino nacional, Regional, internacional, etc.   |
| ENT_OPE     | VARCHAR (1)  | DICCIONARIO.TD_DESTINO_ENTRADA | Igual          | Código operadora                                                  |
| ENT_OPE_DES | VARCHAR (20) | DICCIONARIO.TD_DESTINO_ENTRADA | Igual          | Descripción operadora                                             |

### TD\_FECHA\_TRAFICO

| CAMPO     | TIPO     | FUENTE                   | TRANSFORMACION                                                                            | DESCRIPCION                                   |
|-----------|----------|--------------------------|-------------------------------------------------------------------------------------------|-----------------------------------------------|
| FECHA     | DATETIME | TH.ENTRADA_TRAFICO.FECHA | Distinct fecha                                                                            | Fecha y hora en que se realiza la llamada     |
| SEMESTRE  | SMALLINT | TH.ENTRADA_TRAFICO.FECHA | Case datepart("qq", fecha)<br>When 1 then 1 when 2 then 1 when 3 then 2 when 4 then 2 end | Convierte en semestre a partir del trimestre. |
| TRIMESTRE | SMALLINT | TH.ENTRADA_TRAFICO.FECHA | DATEPART("qq",FEC HA)                                                                     | Convierte trimestre a partir de la fecha dada |
| ANIO      | SMALLINT | TH.ENTRADA_TRAFICO.FECHA | Year(fecha)                                                                               | Convierte la fecha en año.                    |
| MES       | SMALLINT | TH.ENTRADA_TRAFICO.FECHA | MONTH(FECHA)                                                                              | Convierte la fecha en mes.                    |
| DIA       | SMALLINT | TH.ENTRADA_TRAFICO.FECHA | DATEPART("dw", H.FECHA) AS DIA,                                                           | Convierte la fecha en día.                    |
| HORA      | SMALLINT | TH.ENTRADA_TRAFICO.FECHA | DATEPART("Hh",FEC HA) AS HORA                                                             | Obtiene la hora.                              |
| FECHA_DIA | DATETIME | TH.ENTRADA_TRAFICO.FECHA | CONVERT(CHAR,FEC HA ,103)                                                                 | Fecha sin hora (aaaa/mm/dd)                   |

**Importante:** La fecha o la dimensión tiempo está presente en todos los cubos por lo tanto se comparte la dimensión, lo único que cambia es la fuente pero el resto es exactamente igual, por ello hay que tener muy claro en donde se cargan primero las

fechas para que en la transformación de fecha (.dtsx) se trunque la dimensión y luego simplemente añadir las fechas que no existen. La palabra tráfico en la columna fuente es reemplazada por cada uno de las siguientes tablas:

- TH\_SALIDA\_TRAFICO
- TH\_TRAFICO\_LOCAL
- TH\_TRAFICO\_LOCAL\_DETALLADO
- TH\_AUXILIARES
- TH\_TRAFICO\_ERLANGS
- TH\_HACES
- TH\_URA
- TH\_TIPOS\_LLAMADA

### TD\_CONTACTO\_EXTERNO

| CAMPO   | TIPO            | FUENTE                         | TRANSFORMACION                                                                                                                                | DESCRIPCION                                                  |
|---------|-----------------|--------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------|
| CON_COD | VARCHAR<br>(15) | TH_ENTRADA_TRAFICO.<br>CON_COD | Distinct con_cod                                                                                                                              | Código del contacto externo                                  |
| CON_DES | VARCHAR<br>(60) | TH_ENTRADA_TRAFICO.<br>ENT_COM | substring(t.ent_com,1,1)<br>When 'n' then 'nacional'<br>When 'r' then 'regional'<br>When 'i' then 'internacionall'<br>When 'c' then 'celular' | Descripción del contacto de acuerdo al destino de la llamada |

**Importante:** Esta dimensión de contacto externo es compartida con otros cubos pero se especifica en cada uno porque cambia la transformación y la fuente pero el almacenamiento físico es el mismo. Además se debe considerar en donde primero se extrae para truncar la dimensión (TANQUE) y luego añadir los no existentes para que no se duplique la información, así se ahorra espacio en disco.

### TD\_CLIENTE\_TRAFICO

Join qs36f.tarcod=th\_entrada\_trafico.clicat\_cod

| CAMPO       | TIPO            | FUENTE                              | TRANSFORMACION          | DESCRIPCION              |
|-------------|-----------------|-------------------------------------|-------------------------|--------------------------|
| CLI_CAT_COD | SMALLINT        | TH_ENTRADA_TRAFICO<br>.CLI_CAT_COD  | DISTINCT<br>CLI_CAT_COD | Código de categoría      |
| CLI_CAT_DES | VARCHAR<br>(30) | QS36F.TARAR7.<br>TARDE4             | UPPER(TARDE4)           | Descripción de categoría |
| CLI_COD     | INTEGER         | TH_ENTRADA<br>_TRAFFICO.<br>CLI_COD | DISTINCT CLI_COD        | Código de cliente        |

|         |                 |                                                                             |                                                                                                                          |                                                                                         |
|---------|-----------------|-----------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------|
| CLI_NOM | VARCHAR<br>(30) | TH_ENTRADA_ TRAFICO.<br>CLI_COD                                             | DISTINCT CLI_NOM                                                                                                         | Nombre de cliente                                                                       |
| CLI_FEC | VARCHAR<br>(6)  | No tiene fuente Ayuda a poder controlar el historial de cambios del cliente | cast(month(dateadd("month",-1,getdate())) as varchar(2))<br>+<br>cast(year(dateadd("month",-1,getdate())) as varchar(4)) | Indica en que año y mes se pobló a un cliente para controlar la retrospectiva verdadera |

**Importante:** Esta dimensión es compartida con los cubos TRAFICO SALIENTE, TRAFICO LOCAL, TRAFICO LOCAL DETALLADO por este motivo no se repetirá la dimensión ya que sólo la fuente cambia y la transformación es exactamente igual, pero es indispensable establecer un orden de los bats que se deben ejecutar porque sólo a uno se pueda truncar y en el resto sólo se añade los que no existan. Por esa razón en el manual técnico se especifica el orden de los procesos que obligadamente se debe seguir. Las fuentes para esta dimensión son las siguientes:

- TH\_SALIDA\_TRAFICO
- TH\_TRAFICO\_LOCAL
- TH\_TRAFICO\_LOCAL\_DETALLADO

#### TD\_TARIFA

| CAMPO   | TIPO           | FUENTE                | TRANSFORMACION | DESCRIPCION        |
|---------|----------------|-----------------------|----------------|--------------------|
| TAR_COD | VARCHAR<br>(1) | DICCIONARIO.TD_TARIFA | Igual          | Código Tarifa      |
| TAR_DES | VARCHAR<br>(8) | DICCIONARIO.TD_TARIFA | Igual          | Descripción Tarifa |

**Importancia:** Esta dimensión se comparte con tráfico saliente pero no es necesario especificarlo de nuevo porque es exactamente igual, además se extrae una sola vez ya que es a través de la generación de un diccionario.

#### TD\_UBICACION

```
Join etapa.pais.painum = etapa.ciudad.painum and
th_entrada_trafico.pai_cod=etapa.pais.painum and th_entrada_trafico.ciu_cod
= etapa.ciudad.ciu_cod
```

| CAMPO       | TIPO         | FUENTE              | TRANSFORMACION       | DESCRIPCION                 |
|-------------|--------------|---------------------|----------------------|-----------------------------|
| PAI_NUM     | INTEGER      | TH_ENTRADA_TRAFICO  | Distinct pai_cod     | Número del país             |
| PAI_COD     | INTEGER      | ETAPA.PAIS.PAICOD   | Igual                | Código de marcación al país |
| PAI_NOM     | VARCHAR (30) | ETAPA.PAIS.PAINOM   | Igual                | Nombre país                 |
| PAI_CIU_COD | INTEGER      | TH_ENTRADA_TRAFICO  | Distinct pai_ciu_cod | Código de la ciudad         |
| PAI_CIU_NOM | VARCHAR (50) | ETAPA.CIUDAD.CIUNOM | Igual                | Nombre ciudad               |

**Importante:** Esta dimensión es compartida lo único que cambia es la fuente por esa razón ya no se vuelve a especificar. Las fuentes son las siguientes

- TH\_SALIDA\_TRAFICO.PAI\_COD (Número del País)
- TH\_SALIDA\_TRAFICO.PAI\_CIU\_COD (Código de la ciudad)

### 1.1.2. Tráfico Saliente

#### TH\_SALIDA\_TRAFICO

Filtro: substring(saldes,1,4)='1800' ( este filtro se aplica solo cuando se va a subir el trafico 1800 de la tabla SEAAMM, ya que en esta tabla también se encuentra el tráfico a servicio tres dígitos)

Join substring(Etapa.Trafico.salori,len(etapa.Trafico.salori)-5,6) =

Qs36f.telarc.teltel

Tráfico: es reemplazado por cada una de las tablas SNAAMM, ETAPA.SNAAMM, SCAAMM,SEAAMM, SIAAMM.

Como antes ya se indico se debe agrupar y en este caso es:

group by salori, salfec, salhor, salope, saldir, telnom, salcen, salzon, saldes, teltar, saltip, salpai, salcui

| CAMPO   | TIPO        | FUENTE                                                     | TRANSFORMACION                                                                                          | DESCRIPCION                                          |
|---------|-------------|------------------------------------------------------------|---------------------------------------------------------------------------------------------------------|------------------------------------------------------|
| CEN_COD | SMALLINT    | ETAPA.SNAAMM, SCAAMM,SEAAMM ,SIAAMM. SALCEN                | Igual                                                                                                   | Código de la central                                 |
| SAL_COM | VARCHAR (5) | ETAPA.SNAAMM, SCAAMM,SEAAMM ,SIAAMM. SALTIP, SALDIR,SALOPE | Si el tráfico es Reg.-Nac SALTIP+' '+SALDIR+' '+SALOPE, CEL 'C'+ '+SALDIR+SALOPE, INT 'T'+SALDIR+SALOPE | Código compuesto (destino tipo de acceso, operadora) |

|                         |                 |                                                               |                                                                                                           |                                                    |
|-------------------------|-----------------|---------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|----------------------------------------------------|
| CLI_COD                 | INTEGER         | ETAPA.SNAAMM,<br>SCAAMM,SEAAMM<br>,SIAAMM. SALORI             | Substring(SALORI,LEN(SA<br>LORI)-5,6)                                                                     | Código del<br>cliente (#<br>teléfono 6<br>dígitos) |
| CLI_NOM                 | VARCHAR<br>(30) | QS36F.TELARC.<br>TELNOM                                       | Igual                                                                                                     | Nombre del<br>cliente                              |
| CLI_CAT_<br>COD         | SMALLINT        | QS36F.TELARC.<br>TELTAR                                       | Igual                                                                                                     | Categoría del<br>cliente                           |
| PAI_CIU_<br>COD         | INTEGER         | ETAPA.SIAAMM.<br>SALCUI,<br>ETAPA.SNAAMM.<br>SALZON           | IGUAL, Si es celular<br>Va 9,1800 va -1,no definido                                                       | Código de la<br>ciudad                             |
| PAI_COD                 | INTEGER         | ETAPA.SNAAMM.<br>SALPAI                                       | = 58 Cel,nac,reg -1 1800.                                                                                 | Código del<br>país                                 |
| CON_COD                 | VARCHAR<br>(15) | ETAPA.SNAAMM,<br>SCAAMM,SEAAMM<br>,SIAAMM.<br>SALDES          | Igual                                                                                                     | Número<br>quien se<br>llama                        |
| SAL_TAR                 | VARCHAR<br>(1)  | SCAAMM.<br>SALTAR                                             | LTRIM(RTRIM(SALTAR)),<br>'N' para el nac,reg,cel,inter.                                                   | Código de la<br>tarifa                             |
| FECHA                   | DATETIME        | ETAPA.SNAAMM,<br>SCAAMM,SEAAMM<br>,SIAAMM..SALFEC ,<br>SALHOR | Cast(convert(char ,salfec,<br>101) + cast(salhor as<br>varchar(2))+ ':' + '0' + ':' + '0'<br>as datetime) | Fecha y hora<br>formato<br>aaaa/mm/dd:<br>hh:mm:ss |
| Num_llamadas            | INTEGER         | ETAPA.SNAAMM,<br>SCAAMM,SEAAMM<br>,SIAAMM..SALDES             | Count (saldes)                                                                                            | Número de<br>llamadas                              |
| Duracion_<br>llamada    | FLOAT           | ETAPA.SNAAMM,<br>SCAAMM,SEAAMM<br>,SIAAMM..SALDUR             | Round(sum(saldur)/60,2)                                                                                   | Duración en<br>minutos                             |
| Minutos_<br>redondeados | FLOAT           | SINTERNA.DUI                                                  | Sum(saldui)                                                                                               | Minutos<br>redondeados<br>al segmento 6            |

**Importante:** Como se explicó anteriormente existen números que generan tráfico pero no están registrados en la tabla TELARC como abonados por esta razón también se debe hacer lo mismo que en el tráfico entrante con respecto al nombre del cliente y categoría.

TD\_CLIENTE\_TRAFICO, ya fue detalla en tráfico entrante.

#### TD\_SALIDA\_DESTINO

| CAMPO           | TIPO            | FUENTE                            | TRANSFOR<br>MACION | DESCRIPCION                                                 |
|-----------------|-----------------|-----------------------------------|--------------------|-------------------------------------------------------------|
| SAL_COM         | VARCHAR<br>(5)  | DICCIONARIO.TD_DE<br>STINO_SALIDA | Igual              | Código compuesto<br>destino, tipo de<br>conexión, operadora |
| SAL_OPE_<br>COD | VARCHAR<br>(1)  | DICCIONARIO.TD_DE<br>STINO_SALIDA | Igual              | Código de operadora                                         |
| SAL_OPE_<br>DES | VARCHAR<br>(20) | DICCIONARIO.TD_DE<br>STINO_SALIDA | Igual              | Descripción de la<br>operadora                              |
| SAL_TIP_<br>ACC | VARCHAR<br>(1)  | DICCIONARIO.TD_DE<br>STINO_SALIDA | Igual              | Código de acceso                                            |

|             |             |                               |       |                         |
|-------------|-------------|-------------------------------|-------|-------------------------|
| SAL_DES_ACC | VARCHAR(20) | DICCIONARIO.TD_DESTINO_SALIDA | Igual | Descripción del acceso  |
| SAL_COD_DES | VARCHAR(1)  | DICCIONARIO.TD_DESTINO_SALIDA | Igual | Código del destino      |
| SAL_DES_DES | VARCHAR(20) | DICCIONARIO.TD_DESTINO_SALIDA | Igual | Descripción del destino |

TD\_UBICACION, ya fue detallada en tráfico entrante.

### TD\_CONTACTO\_EXTERNO

Filtro len (Con\_cod) = 3 (Servicio tres dígitos)

Join th\_salida\_trafico.con\_cod=servicio\_3digitos.con\_cod

| CAMPO   | TIPO        | FUENTE                    | TRANSFORMACION   | DESCRIPCION                        |
|---------|-------------|---------------------------|------------------|------------------------------------|
| CON_COD | VARCHAR(15) | TH_SALIDA_TRAFICO.CON_COD | Distinct con_cod | Código del contacto externo        |
| CON_DES | VARCHAR(60) | SERVICIO_3DIGITOS.CONDES  | Igual            | Descripción servicio tres dígitos. |

Filtro: substring(con\_cod,1,4) = '1800'

Join Directorio\_1800.num\_180=Th\_salida\_trafico.con\_cod

| CAMPO   | TIPO        | FUENTE                    | TRANSFORMACION   | DESCRIPCION                 |
|---------|-------------|---------------------------|------------------|-----------------------------|
| CON_COD | VARCHAR(15) | TH_SALIDA_TRAFICO.CON_COD | Distinct con_cod | Código del contacto externo |
| CON_DES | VARCHAR(60) | DIRECTORIO_1800.DES_180   | Igual            | Descripción 1800            |

| CAMPO   | TIPO        | FUENTE                    | TRANSFORMACION                                                                                                                   | DESCRIPCION                                    |
|---------|-------------|---------------------------|----------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------|
| CON_COD | VARCHAR(15) | TH_SALIDA_TRAFICO.CON_COD | Distinct con_cod                                                                                                                 | Código del contacto externo                    |
| CON_DES | VARCHAR(60) | TH_SALIDA_TRAFICO.SAL_COM | Substring(t.ent_com,1,1) When 'n' then 'nacional' When 'r' then 'regional' When 'i' then 'internacional' When 'c' then 'celular' | Descripción del contacto de acuerdo al destino |

**Importante:** Cuando se realice los procesos de extracción y transformación (paquetes con extensión .dtsx) se debe considerar el orden en el que se va a construir la dimensión por que debe tener en cuenta en donde se trunca la dimensión y en el resto de procesos debe añadir la información.

TD\_TARIFA y TD\_CENTRALES, ya está detallado anteriormente.

### 1.1.2.1. Tráfico local

#### TH\_TRAFICO\_LOCAL

Filtro: len(rtrim(ltrim(saldes)))=3(para el tráfico de servicio tres dígitos)

Join cast(substring(cast(connumero as varchar(7)),2,6)as int) = teltel (local)

Join teltel = cast(substring(salori,len(salori)-5,6) as int) (servicio 3\_dígitos)

Group by salori, saldes, salfec,telnom,teltar,salcen (solo para tráfico 3\_dígitos)

| CAMPO                | TIPO            | FUENTE                                                              | TRANSFORMACION                                                                                                             | DESCRIPCION                                   |
|----------------------|-----------------|---------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------|
| FECHA                | DATETIME        | ETAPA.CONTADORES.CO<br>NANIO, CONMES<br><br>ETAPA.SEAAMM.<br>SALFEC | cast(cast(conanio as<br>char(4)) + cast(conmes as<br>varchar(2)) + '01' as<br>datetime) ;<br>CONVERT(CHAR,<br>SALFEC, 103) | Fecha del<br>consumo<br>formato<br>aaaa/mm/dd |
| CLI_COD              | INT             | ETAPA.CONTADORES.CO<br>NNUMERO<br>ETAPA.SEAAMM.<br>SALORI           | cast(substring(cast(connum<br>ero) as varchar(7)),2,6)as<br>int),<br>substring(salori,len(salori)-<br>5,6)                 | Código del<br>cliente                         |
| CLI_NOM              | VARCHAR<br>(30) | QS36F.TELARC.<br>TELNOM                                             | Igual                                                                                                                      | Nombre del<br>cliente                         |
| CLI_CAT_<br>COD      | SMALLINT        | QS36F.TELARC.<br>TELTAR                                             | Igual                                                                                                                      | Categoría del<br>cliente                      |
| CON_COD              | VARCHAR<br>(15) | ETAPA.SEAAMM.<br>SALDES                                             | Igual; pero '0' para el<br>trafico local.                                                                                  | Número del<br>contacto                        |
| LOC_SER              | SMALLINT        | DICCIONARIO.TIPO_<br>LOCAL                                          | '2' local;'1' servicios                                                                                                    | Tipo tráfico<br>local                         |
| NUM_<br>LLAMADAS     | INT             | ETAPA.SEAAMM.<br>SALDES                                             | En local 0 por que no hay<br>como identificar; 3_dígitos<br>count(saldes)                                                  | Número de<br>llamadas (tres<br>dígitos)       |
| DURACION_<br>LLAMADA | FLOAT           | ETAPA.CONTADORES.<br>CONCONTA,<br>ETAPA.SLAAMM.<br>SALDUR           | round(consumo/60,2);<br>round(sum(saldur/60),2)                                                                            | Minutos<br>consumido<br>local,<br>3_dígitos   |
| CEN_COD              | SMALLINT        | ETAPA.SEAAMM.<br>SALCEN                                             | -1 central no definida<br>trafico local, igual<br>3_dígitos                                                                | Central que<br>origina la<br>llamada          |

**Importante:** Al igual que el resto de tablas de hechos en las que se trabaje con los clientes se tendrán en cuenta los números que generen tráfico pero no están registrados como abonados. Con relación al consumo mensual del tráfico local se debe obtener de la resta del mes presente con el mes anterior al cual en la metadata se le asignó el nombre de variable Consumo.

## TD\_TIPO\_LOCAL

| CAMPO           | TIPO            | FUENTE                        | TRANSFORMACION | DESCRIPCION                 |
|-----------------|-----------------|-------------------------------|----------------|-----------------------------|
| LOC_COD         | SMALLINT        | DICCIONARIO.TD<br>_TIPO_LOCAL | Igual          | Código del tráfico          |
| LOC_DES         | VARCHAR<br>(30) | DICCIONARIO.TD<br>_TIPO_LOCAL | Igual          | Descripción tráfico         |
| LOC_SER<br>_COD | SMALLINT        | DICCIONARIO.TD<br>_TIPO_LOCAL | Igual          | Código del servicio         |
| LOC_SER<br>_DES | VARCHAR<br>(30) | DICCIONARIO.TD<br>_TIPO_LOCAL | Igual          | Descripción del<br>servicio |

TD\_CENTRALES, TD\_FECHA\_TRAFICO, dimensiones ya detalladas

### 1.1.2.2. Tráfico local detallado

Cuando se haga los procesos de transformación se debe considerar que en este caso el origen como el destino son abonados o números de E.T.A.P.A. por lo que se debe crear un alias a las tablas para que se pueda obtener los nombres y categorías, además deben subir los datos así no existan en la tabla “telarc”, por lo que se debe hacer procesos adicionales considerando que:

- Exista el abonado origen pero no el abonado destino
- Exista el abonado destino pero no el abonado de origen
- No exista abonado origen ni abonado destino.

## TH\_TRAFICO\_LOCAL\_DETALLADO

Join **d.telcel** = saldes and **o.telcel** = salori

group by salfec, salhor,salori,salceo,teltar,saldes,salced, d.telnom, .o.telnom

| CAMPO               | TIPO            | FUENTE                  | TRANSFORMACION                                                                                              | DESCRIPCION                                         |
|---------------------|-----------------|-------------------------|-------------------------------------------------------------------------------------------------------------|-----------------------------------------------------|
| FECHA               | DATETIME        | ETAPA.SLAAMM.<br>SALFEC | cast(convert(char ,a.salfec,<br>101) + cast(a.salhor as<br>varchar(2))+ ':'+ '0' + ':' +<br>'0' as datetime | Fecha consumo<br>formato<br>aaaa/mm/dd:hh:mm:<br>ss |
| CLI_COD_<br>ORI     | INT             | ETAPA.SLAAMM.<br>SALORI | substr(ltrim(rtrim(salori)),l<br>ength(salori)-6,6)                                                         | Código del cliente                                  |
| CLI_NOM_<br>ORI     | VARCHAR<br>(30) | QS36F.TELARC.<br>TELNOM | Igual                                                                                                       | Nombre del cliente<br>origen                        |
| CLI_CAT_<br>COD_ORI | SMALLINT        | QS36F.TARAR7.<br>TARCOD | Igual                                                                                                       | Categoría del cliente<br>origen                     |
| CEN_COD_<br>ORI     | SMALLINT        | ETAPA.SLAAMM.<br>SALCEO | Igual                                                                                                       | Central donde se<br>origina la llamada              |
| CLI_COD_<br>DES     | VARCHAR<br>(15) | ETAPA.SLAAMM.<br>SALDES | substr(ltrim(rtrim(saldes)),l<br>ength(ltrim(rtrim(saldes)))-<br>5,6)                                       | Código del cliente<br>destino                       |

|                  |                 |                         |                         |                                    |
|------------------|-----------------|-------------------------|-------------------------|------------------------------------|
| CLI_NOM_DES      | VARCHAR<br>(30) | QS36F.TELARC.<br>TELNOM | Igual                   | Nombre del cliente destino         |
| CLI_CAT_COD_DES  | SMALLINT        | QS36F.TELARC.<br>TELTAR | Igual                   | Categoría del cliente destino      |
| CEN_COD_DES      | SMALLINT        | ETAPA.SEAAMM.<br>SALCED | Igual                   | Central donde se recibe la llamada |
| NUM_LLAMADAS     | INT             | ETAPA.SEAAMM.<br>SALDES | count(saldes)           | Número de llamadas                 |
| DURACION_LLAMADA | FLOAT           | ETAPA.SEAAMM.<br>SALDES | round(sum(saldur)/60,2) | Minutos consumido                  |

**Importante:** Como se explicó anteriormente existen números que generan tráfico pero no están registrados en la tabla TELARC como abonados por esta razón también se debe hacer lo mismo que en el tráfico entrante con respecto al nombre del cliente y categoría. Las dimensiones que ocupa este cubo ya están detalladas anteriormente.

## 1.2. Control de conmutación

### 1.2.1. Tráfico Auxiliares

#### TH\_AUXILIARES

Filtro: fecha<to\_char(sysdate,'yy-mm-dd')

Group by fecha, hora, central

| CAMPO    | TIPO     | FUENTE                                        | TRANSFORMACION                                                                                                                                                                                                     | DESCRIPCION                                                      |
|----------|----------|-----------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------|
| AUX_COD  | SMALLINT | DICCIONARIO.TD<br>_AUXILIARES                 | '1' RGF, '2' CCf, '3' V23                                                                                                                                                                                          | Código del auxiliar (rgf/ccf/v23)                                |
| CEN_COD  | SMALLINT | ETAPA.<br>CLASE_A<br>.CENTRAL                 | when 'centro' then 0 when 'ejido' then 1 when 'totora' then 2 else -1                                                                                                                                              | Código de la central                                             |
| FECHA    | DATETIME | ETAPA.<br>CLASE_A.<br>FECHA,<br>HORA          | length(rtrim(ltrim(fecha))) when 8 then concat('20', rtrim(ltrim(fecha))) else concat('2',rtrim(ltrim(fecha))) ; cast(convert(char,[fecha], 101) + cast([hora] as varchar(2)) + ':' + '0' + ':' + '0' as datetime) | Fecha en la que se captura los datos formato aaaa/mm/dd:hh:mm:ss |
| DES_COD  | SMALLINT | DICCIONARIO.TD<br>_DESCRIPCION_<br>AUXILIARES | Esto va a depender del tipo de auxiliar que se suba y la descripción de la medida                                                                                                                                  | Código de la descripción de la medida por ej. # de tomas         |
| CANTIDAD | FLOAT    | ETAPA.<br>CLASE_A.<br>A02,A04,AP4<br>...ETC   | Va a depender del tipo de auxiliar y descripción de medida. indicados en las tablas siguientes                                                                                                                     | Cantidad de acuerdo al auxiliar y descripción medida             |

Para entender mejor la transformación de los auxiliares se realizó una tabla estándar con la cual se debe transformar la tabla de hechos de auxiliares

#### AUXILIARES RGF

| AUX_<br>COD | DES_<br>COD | Canti<br>dad |
|-------------|-------------|--------------|
| 1           | 1           | AQ1          |
| 1           | 2           | AP1          |
| 1           | 4           | AR1          |
| 1           | 5           | AE1          |
| 1           | 6           | A01          |
| 1           | 7           | AS1          |

#### AUXILIARES CCF

| AUX_<br>COD | DES_<br>COD | Canti<br>dad |
|-------------|-------------|--------------|
| 2           | 1           | AQ2          |
| 2           | 2           | AP2          |
| 2           | 4           | AR2          |
| 2           | 5           | AE2          |
| 2           | 6           | A02          |
| 2           | 7           | AS2          |

| AUX_<br>COD | DES_<br>COD | Canti<br>dad |
|-------------|-------------|--------------|
| 3           | 1           | AQ4          |
| 3           | 3           | AP4          |
| 3           | 4           | AR4          |
| 3           | 5           | AE4          |
| 3           | 6           | A04          |
| 3           | 7           | AS4          |

#### AUXILIARES V23

#### TD\_AUXILIARES

| CAMPO   | TIPO            | FUENTE                    | TRANSFORMACION | DESCRIPCION              |
|---------|-----------------|---------------------------|----------------|--------------------------|
| AUX_COD | SMALLINT        | DICCIONARIO.TD_AUXILIARES | Igual          | Código del auxiliar      |
| AUX_DES | VARCHAR<br>(45) | DICCIONARIO.TD_AUXILIARES | Igual          | Descripción del auxiliar |

#### TD\_DESCRIPCION\_AUXILIARES

| CAMPO   | TIPO            | FUENTE                                | TRANSFORMACION | DESCRIPCION           |
|---------|-----------------|---------------------------------------|----------------|-----------------------|
| DES_COD | SMALLINT        | DICCIONARIO.TD_DESCRIPCION_AUXILIARES | Igual          | Código de descripción |
| DES_DES | VARCHAR<br>(40) | DICCIONARIO.TD_DESCRIPCION_AUXILIARES | Igual          | Descripción           |

TD\_CENTRALES y TD\_FECHA\_TRAFICO, están descritas anteriormente

## 1.2.2. Tráfico Erlangs

### TH\_TRAFICO\_ERLANGS

Filtro: fecha<to\_char(sysdate,'yy-mm-dd')

Group by central,fecha,hora

| CAMPO    | TIPO     | FUENTE                           | TRANSFORMACION                                                                                               | DESCRIPCION                        |
|----------|----------|----------------------------------|--------------------------------------------------------------------------------------------------------------|------------------------------------|
| CEN_COD  | SMALLINT | ETAPA.CLASE_C.CENTRAL            | when central = 'centro' then 0<br>when central = 'ejido' then 1<br>when central = 'totorá' then 2<br>else -1 | Código de la central               |
| FECHA    | DATETIME | ETAPA.CLASE_C.FECHA, HORA        | cast(convert(char,fecha,101) +<br>cast(hora as varchar(2))<br>+':'+'0'+':'+'0' as datetime                   | Fecha: formato<br>aa/mm/dd h:mm:ss |
| FLU_COD  | SMALLINT | DICCIONARIO.TD_TIPO_FLUJO        | depende de la extracción de la<br>cantidad                                                                   | Código del flujo de<br>tráfico     |
| CANTIDAD | FLOAT    | ETAPA.CLASE_C.CI, CD, CA, CC, CV | Sumar el dato que correcto de<br>acuerdo al código que se<br>registre, indicados en la tabla<br>siguiente    | Cantidad de tráfico<br>porcentaje  |

Al igual que los auxiliares, en el tráfico de erlangs es necesario considerar la siguiente tabla para construir el contenido de la tabla de hecho:

| FLU_COD | CANTIDAD |
|---------|----------|
| 1       | CI       |
| 2       | CA       |
| 3       | CD       |
| 4       | CT       |
| 5       | CV       |

### TD\_TIPO\_FLUJO

| CAMPO   | TIPO            | FUENTE                    | TRANSFORMACION | DESCRIPCION                            |
|---------|-----------------|---------------------------|----------------|----------------------------------------|
| FLU_COD | SMALLINT        | DICCIONARIO.TD_TIPO_FLUJO | Igual          | Código del tipo de<br>flujo de tráfico |
| FLU_DES | VARCHAR<br>(17) | DICCIONARIO.TD_TIPO_FLUJO | Igual          | Descripción del flujo<br>de tráfico    |

TD\_CENTRALES, TD\_FECHA\_TRAFICO, ya están detallados.

### 1.2.3. Tráfico Hazes

#### TH\_HACES

Filtro: LTRIM(RTRIM(HAZ)) <> ''

Filtro: fecha<to\_char(sysdate,'yy-mm-dd')

Group by fecha, hora, central

Para distinguir el tipo de haz

Join Haz in (select haz from clase\_fb group by haz having (sum(foa) <> 0 and sum(fod) = 0) ) Haz entrante

Join Haz in (select haz from clase\_fb group by haz having (sum(foa) = 0 and sum(fod) <> 0) ) Haz saliente

Join Haz in (select haz from clase\_fb group by haz having (sum(foa) <> 0 and sum(fod) <> 0) ) Haz Mixto

Group by central, fecha, hora, haz

Filtro: fecha<to\_char(sysdate,'yy-mm-dd')

Group by fecha, hora ,central,haz

| CAMPO           | TIPO           | FUENTE                                   | TRANSFORMACION                                                                                                                                                                           | DESCRIPCION                                                                 |
|-----------------|----------------|------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------|
| HAZ_COD         | VARCHAR<br>(6) | ETAPA.CLASE_FB.<br>HAZ                   | Igual                                                                                                                                                                                    | Código del Haz                                                              |
| FECHA           | DATETIME       | ETAPA.CLASE_FB.<br>FECHA, HORA           | len(fecha) when 8 then<br>cast('20'+fecha+cast(hora as<br>varchar(2)) + ':'+0+':'+'0' as<br>datetime) else<br>cast('2'+fecha+cast(hora as<br>varchar(2)) + ':'+0+':'+'0' as<br>datetime) | Fecha que se<br>genera el tráfico<br>formato<br>aaaa/mm/dd:hh:mm<br>:ss     |
| CEN_COD         | SMALLINT       | ETAPA.CLASE_FB.<br>CENTRAL               | Case when central = 'centro'<br>then 0 when central = 'ejido'<br>then 1 when central = 'totora'<br>then 2 else -1                                                                        | Código de la<br>central                                                     |
| DES_HAZ_C<br>OD | SMALLINT       | DICCIONARIO.TD_<br>DESCRIPCION_HA<br>CES | Depende de la extracción de<br>cantidad y del tipo del haz                                                                                                                               | Código de la<br>descripción de la<br>medida                                 |
| CANTIDAD        | FLOAT          | ETAPA.CLASE_FB.<br>FQ1,FPA.....ETC       | Sumar el dato que corresponda<br>de acuerdo al código que se<br>registre, indicados en las tablas<br>siguientes                                                                          | Cantidad de<br>acuerdo a la<br>descripción de<br>medida y al tipo de<br>haz |

Para saber que se carga a cantidad y en la descripción de la medida nos basaremos en las siguientes tablas:

**Haz Entrante**

| DES_HAZ_COD | CANTIDAD |
|-------------|----------|
| 1           | FPA      |
| 3           | FRA      |
| 5           | FEA      |
| 7           | FOA      |
| 9           | FQ1      |
| 10          | FS1      |
| 11          | PRS1     |

**Haz Saliente**

| DES_HAZ_COD | CANTIDAD |
|-------------|----------|
| 2           | FPD      |
| 4           | FRD      |
| 6           | FED      |
| 8           | FOD      |
| 9           | FQ1      |
| 10          | FS1      |
| 11          | PRS1     |

**Haz Mixto**

| DES_HAZ_COD | CANTIDAD |
|-------------|----------|
| 1           | FPA      |
| 2           | FPD      |
| 3           | FRA      |
| 4           | FRD      |
| 5           | FEA      |
| 6           | FED      |
| 7           | FOA      |
| 8           | FOD      |
| 9           | FQ1      |
| 10          | FS1      |
| 11          | PRS1     |

**TD\_HACES**

| CAMPO       | TIPO        | FUENTE               | TRANSFORMACION | DESCRIPCION                           |
|-------------|-------------|----------------------|----------------|---------------------------------------|
| HAZ_COD     | VARCHAR(6)  | DICCIONARIO.TD_HACES | Igual          | Código (nombre ) del haz              |
| HAZ_TIP_COD | VARCHAR(1)  | DICCIONARIO.TD_HACES | Igual          | Código del tipo de haz                |
| HAZ_TIP_DES | VARCHAR(15) | DICCIONARIO.TD_HACES | Igual          | Descripción del tipo de haz           |
| HAZ_INT_COD | SMALLINT    | DICCIONARIO.TD_HACES | Igual          | Código interconexión                  |
| HAZ_INT_DES | VARCHAR(51) | DICCIONARIO.TD_HACES | Igual          | Descripción de interconexión          |
| HAZ_TRA_COD | SMALLINT    | DICCIONARIO.TD_HACES | Igual          | Código del trafico cursado por el haz |
| HAZ_TRA_DES | VARCHAR(46) | DICCIONARIO.TD_HACES | Igual          | Descripción del trafico               |

## TD\_DESCRIPCION\_HACES

| CAMPO       | TIPO         | FUENTE                           | TRANSFORMACION | DESCRIPCION                          |
|-------------|--------------|----------------------------------|----------------|--------------------------------------|
| DES_HAZ_COD | SMALLINT     | DICCIONARIO.TD_DESCRIPCION_HACES | Igual          | Código descripción de la medida      |
| DES_HAZ_DES | VARCHAR (81) | DICCIONARIO.TD_DESCRIPCION_HACES | Igual          | Descripción medida ej. Tomas llegada |

TD\_CENTRALES, TD\_CLIENTE\_FECHA, ya están detalladas.

### 1.2.4. Tráfico URA

#### TH\_URA

Filtro: fecha<to\_char(sysdate,'yy-mm-dd')

Group by central,fecha,ura,hora

| CAMPO           | TIPO         | FUENTE                         | TRANSFORMACION                                                                                                                                                                                                                                        | DESCRIPCION                                                                        |
|-----------------|--------------|--------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------|
| CEN_COD         | SMALLINT     | ETAPA.CLASE_R.CENTRAL          | Case when central = 'centro' then 0 when central = 'ejido' then 1 when central = 'totora' then 2 else -1                                                                                                                                              | Código central                                                                     |
| FECHA           | DATETIME     | ETAPA.CLASE_R.FECHA,HORA       | length(rtrim(ltrim(fecha)))<br>when 8 then<br>concat('20',rtrim(ltrim(fecha)))<br>else<br>concat('2',rtrim(ltrim(fecha)));<br>cast(convert(char,[fecha],101)<br>+cast([hora]as varchar(2)) +<br>'+'0'+ '+'0' as datetime)                             | Fecha y hora de la observación<br>formato<br>aaaa/mm/dd:hh:<br>mm:ss               |
| URA_COD_COM     | VARCHAR (10) | ETAPA.CLASE_R.URA,CENTRAL      | when central = 'centro' then '0'<br>+ '-' + cast(ura as varchar(3))<br>when central = 'ejido' then '1'<br>+ '-' + cast(ura as varchar(3))<br>when central = 'totora' then '2'<br>+ '-' + cast(ura as varchar(3))<br>else -1 + cast(ura as varchar(3)) | Código compuesto para identificar la URA ubicación y la central a la que pertenece |
| URA_LLA_DES_COD | SMALLINT     | DICCIONARIO.TD                 | Si llamada entrante 1; saliente 2                                                                                                                                                                                                                     | Código medida                                                                      |
| CANTIDAD        | FLOAT        | ETAPA.CLASE_R.RRA,RPA,RAA, ETC | Sumar cada uno de los campos de acuerdo al código de descripción de medida, indicados en la tabla siguiente                                                                                                                                           | Cantidad llamadas a abonados, ej. presentadas, etc.                                |

Para facilitar el proceso de la extracción se ha identificado que se debe subir en URA\_LLA\_DES\_COD y CANTIDAD de acuerdo a la siguiente tabla:

| URA_LLA_DES_COD | CANTIDAD |
|-----------------|----------|
| 1               | RPA      |
| 2               | RRA      |
| 3               | RAA      |
| 4               | RPD      |
| 5               | RRD      |
| 6               | RAD      |
| 7               | RRX      |

### TD\_LLAMADA\_URA

| CAMPO           | TIPO            | FUENTE                                         | TRANSFORMACION | DESCRIPCION                                          |
|-----------------|-----------------|------------------------------------------------|----------------|------------------------------------------------------|
| URA_LLA_COD     | SMALLINT        | DICCIONARIO.TD_LLAMADA_URA.<br>URA_LLA_COD     | Igual          | Código del tipo de llamada                           |
| URA_LLA_DES     | VARCHAR<br>(18) | DICCIONARIO.TD_LLAMADA_URA.<br>URA_LLA_DES     | Igual          | Descripción de la llamada                            |
| URA_LLA_DES_COD | SMALLINT        | DICCIONARIO.TD_LLAMADA_URA.<br>URA_LLA_DES_COD | Igual          | Código de la descripción de los campos RPA, RRA etc. |
| URA_LLA_DES_DES | VARCHAR<br>(50) | DICCIONARIO.TD_LLAMADA_URA.<br>URA_LLA_DES_DES | Igual          | Descripción de los campos RPA, RRA etc.              |

### TD\_URA

| CAMPO       | TIPO            | FUENTE              | TRANSFORMACION | DESCRIPCION                              |
|-------------|-----------------|---------------------|----------------|------------------------------------------|
| URA_COD_COM | VARCHAR<br>(10) | DICCIONARIO.TD_URA. | Igual          | Código compuesto para identificar la ura |
| URA_COD_UBI | SMALLINT        | DICCIONARIO.TD_URA  | Igual          | Código de ubicación                      |
| URA_COD     | SMALLINT        | DICCIONARIO.TD_URA  | Igual          | Código de la URA                         |
| URA_UBI_DES | VARCHAR<br>(45) | DICCIONARIO.TD_URA  | Igual          | Descripción de la ubicación de la URA    |
| URA_NUM_AB0 | INT             | DICCIONARIO.TD_URA  | Igual          | Número de abonados                       |
| URA_NUM_MON | SMALLINT        | DICCIONARIO.TD_URA  | Igual          | Número de monederos                      |
| URA_NUM_RDS | INT             | DICCIONARIO.TD_URA  | Igual          | Número de abonados RDSI                  |

|                 |          |                    |       |                                |
|-----------------|----------|--------------------|-------|--------------------------------|
| URA_NUM_LR      | SMALLINT | DICCIONARIO.TD_URA | Igual | Número de LR                   |
| URA_NUM_ABO_SER | INT      | DICCIONARIO.TD_URA | Igual | Número de abonados en servicio |

Esta dimensión por contener valores numéricos al momento de procesar el cubo es reconocida como dimensión y tabla de hecho debido a que los campos numéricos no están asociados a ninguna tabla de hechos y estos son considerados como medidas y pueden ser agregados (Sumados).

TD\_CENTRALES, TD\_FECHA\_TRAFICO, ya están detalladas.

### 1.2.5. Tipos de Llamadas en la central

#### TH\_LLAMADA

Filtro: fecha < to\_char(sysdate,'yy-mm-dd')

| CAMPO        | TIPO     | FUENTE                      | TRANSFORMACION                                                                                                                                                                                                     | DESCRIPCION                                                                |
|--------------|----------|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------|
| CEN_COD      | SMALLINT | ETAPA.CLASE_T               | central when 'centro' then 0<br>when 'Totorá' then 2 when<br>'Ejido' then 1 else -1                                                                                                                                | Código de la central                                                       |
| FLU_COD      | SMALLINT | DICCIONARIO_TD_TIPO_FLUJO   | Depende de lo que se extrae en num_llamadas                                                                                                                                                                        | Código del tipo de flujo                                                   |
| TIP_SUB_COD  | SMALLINT | DICCIONARIO_TD_TIPO_LLAMADA | Depende de lo que se extrae en num_llamadas                                                                                                                                                                        | Código subcausa/tipo de llamada                                            |
| FECHA        | DATETIME | ETAPA.CLASE_T               | length(rtrim(ltrim(fecha)))<br>when 8 then concat('20',rtrim(ltrim(fecha))) else<br>concat('2',rtrim(ltrim(fecha)))<br>;cast(convert(char,fecha,101)<br>+ cast(hora as varchar(2)) +<br>'+'0'+':'+'0' as datetime) | Fecha en la que se genera la llamada<br>formato<br>aaaa/mm/dd:<br>hh:mm:ss |
| NUM_LLAMADAS | FLOAT    | ETAPA.CLASE_T               | Sumar los campos de la clase t de acuerdo al tipo de flujo y subcausa indicados en la tabla siguiente                                                                                                              | Número de llamadas según su tipo                                           |

Para poder transformar correctamente nos basaremos en la siguiente tabla la cual contiene en la primera columna el nombre de la medida (num\_llamadas), en la segunda columna el código del flujo de tráfico (flu\_cod) y la tercera columna el código de la subcausa de la llamada (tip\_sub\_cod).

|        |   |    |
|--------|---|----|
| T05_I  | 1 | 38 |
| T05_A  | 2 | 38 |
| T05_D  | 3 | 38 |
| T05_T  | 4 | 38 |
| T01_I+ | 1 | 36 |
| T02_I  |   |    |
| T01_D+ | 3 | 36 |
| T02_D  |   |    |
| T04_I  | 1 | 37 |
| T04_A  | 2 | 37 |
| T04_D  | 3 | 37 |
| T04_T  | 4 | 37 |
| T06_D  | 3 | 40 |
| T06_T  | 4 | 40 |
| T07_I  | 1 | 39 |
| T07_A  | 2 | 39 |
| T07_D  | 3 | 39 |
| T07_T  | 4 | 39 |
| T12_I  | 1 | 1  |
| T12_A  | 2 | 1  |
| T12_D  | 3 | 1  |
| T12_T  | 4 | 2  |
| T13_I  | 1 | 3  |
| T13_A  | 2 | 3  |
| T13_D  | 3 | 3  |
| T13_T  | 4 | 3  |
| T14_I  | 1 | 4  |
| T14_A  | 2 | 4  |
| T15_I  | 1 | 5  |
| T15_A  | 2 | 5  |
| T15_D  | 3 | 6  |
| T15_T  | 4 | 6  |
| T16_I  | 1 | 7  |
| T16_A  | 2 | 7  |
| T16_D  | 3 | 7  |
| T16_T  | 4 | 7  |
| T17_I  | 1 | 8  |
| T17_A  | 2 | 8  |
| T17_D  | 3 | 8  |
| T17_T  | 4 | 8  |

|       |   |    |
|-------|---|----|
| T18_I | 1 | 9  |
| T18_A | 2 | 9  |
| T18_D | 3 | 9  |
| T18_T | 4 | 9  |
| T19_I | 1 | 10 |
| T19_D | 3 | 10 |
| T22_I | 1 | 11 |
| T22_A | 2 | 11 |
| T22_D | 3 | 11 |
| T22_T | 4 | 11 |
| T23_I | 1 | 12 |
| T23_A | 2 | 12 |
| T23_D | 3 | 12 |
| T23_T | 4 | 12 |
| T24_I | 1 | 13 |
| T24_A | 2 | 13 |
| T24_D | 3 | 13 |
| T24_T | 4 | 13 |
| T32_I | 1 | 14 |
| T32_A | 2 | 14 |
| T33_I | 1 | 15 |
| T33_T | 4 | 15 |
| T34_I | 1 | 16 |
| T34_A | 2 | 16 |
| T34_D | 3 | 16 |
| T34_T | 4 | 16 |
| T35_I | 1 | 17 |
| T35_A | 2 | 17 |
| T36_I | 1 | 18 |
| T36_A | 2 | 18 |
| T36_D | 3 | 18 |
| T36_T | 4 | 18 |
| T37_I | 1 | 19 |
| T37_A | 2 | 19 |
| T37_D | 3 | 19 |
| T37_T | 4 | 19 |
| T38_I | 1 | 20 |
| T38_D | 3 | 20 |
| T39_I | 1 | 21 |

|       |   |    |
|-------|---|----|
| T39_A | 2 | 21 |
| T39_D | 3 | 21 |
| T39_T | 4 | 21 |
| T50_I | 1 | 22 |
| T50_A | 2 | 22 |
| T50_D | 3 | 22 |
| T50_T | 4 | 22 |
| T42_D | 3 | 23 |
| T42_T | 4 | 23 |
| T43_D | 3 | 24 |
| T43_T | 4 | 24 |
| T44_D | 3 | 25 |
| T44_T | 4 | 25 |
| T45_D | 3 | 26 |
| T45_T | 4 | 26 |
| T46_D | 3 | 27 |
| T46_T | 4 | 27 |
| T47_D | 3 | 28 |
| T47_T | 4 | 28 |
| T48_D | 3 | 29 |
| T48_T | 4 | 29 |
| T49_D | 3 | 30 |
| T49_T | 4 | 30 |
| T51_D | 3 | 31 |
| T51_T | 4 | 31 |
| T54_A | 2 | 32 |
| T54_T | 4 | 32 |
| T56_A | 2 | 33 |
| T56_T | 4 | 33 |
| T62_I | 1 | 34 |
| T62_A | 2 | 34 |
| T62_D | 3 | 34 |
| T62_T | 4 | 34 |
| T64_I | 1 | 35 |
| T64_A | 2 | 35 |
| T64_D | 3 | 35 |
| T64_T | 4 | 35 |

### TD\_TIPOS\_LLAMADAS

| CAMPO       | TIPO         | FUENTE                      | TRANSFORMACION | DESCRIPCION                                  |
|-------------|--------------|-----------------------------|----------------|----------------------------------------------|
| TIP_COD     | SMALLINT     | DICCIONARIO.TD_TIPO_LLAMADA | Igual          | Código del tipo de llamada(eficaz/ineficaz)  |
| TIP_DES     | VARCHAR (8)  | DICCIONARIO.TD_TIPO_LLAMADA | Igual          | Descripción de la llamada                    |
| TIP_CAU_COD | SMALLINT     | DICCIONARIO.TD_TIPO_LLAMADA | Igual          | Código de la causa o subtipo de llamada      |
| TIP_CAU_DES | VARCHAR (40) | DICCIONARIO.TD_TIPO_LLAMADA | Igual          | Descripción de la causa o subtipo de llamada |

|             |             |                             |       |                            |
|-------------|-------------|-----------------------------|-------|----------------------------|
| TIP_SUB_COD | SMALLINT    | DICCIONARIO.TD_TIPO_LLAMADA | Igual | Código de la subcausa      |
| TIP_SUB_DES | VARCHAR(60) | DICCIONARIO.TD_TIPO_LLAMADA | Igual | Descripción de la subcausa |

TD\_CENTRALES, TD\_FECHA\_TRAFICO, TD\_TIPO\_FLUJO, ya están detalladas lo único que cambia es la fuente como ya se indicó anteriormente en este caso es TH\_LLAMADAS. No se ha vuelto a explicar porque ventajosamente tienen el mismo nombre y tipos los campos de las dimensiones que comparten.

## ANEXO II DICCIONARIO DE DATOS

### 2.1. Descripción de las Bases de datos utilizadas

La BDD Tanque es, en parte, espejo del transaccional, un espejo que sólo contiene la información necesaria muchas de las veces realizada ya algunas conversiones implícitas de datos, además aquí se realiza las transformaciones indicadas en el entregable de la identificación de fuentes contiene la información necesaria de lo que se va a cargar al *Data Warehouse*. Por ejemplo en el caso del Tráfico telefónico contendrá solo la información del mes en cambio en el control de calidad de circulación del tráfico (Conmutación) tendrá la información que corresponda al día que se va a subir. Se utiliza esta base de datos auxiliar ya que no es posible extraer, transformar los datos de dos servidores a la vez.

La BDD DWH es la que contiene toda la información ya transformada, es decir pasó por un proceso de población, integridad y contiene el historial. DWH es una base de datos multidimensional cuyos datos se almacenan de manera que se optimice el acceso a las informaciones que se precisan para atender a peticiones no previstas al crearse la base.

Para poder establecer los tipos de datos que se van a utilizar se hizo un análisis ya que la base dimensional ocupa mucho más espacio que en los modelos relacionales debido a que están desnormalizadas las tablas y ya no existen relaciones entre ellas excepto entre la tabla de hechos y dimensiones considerando el incremento en el futuro. A pesar de eso existieron muchos datos que no utilizan completamente el espacio asignado como por ejemplo: cambiar un dato tipo integer a un smallint ó de un char a varchar etc.

El estándar que se estableció para el nombre de tablas y campos son las siguientes:

**En la base Tanque** las tablas que representes las dimensiones empezarán con TD\_NombreTabla , las tablas de hechos empezaran con TH\_NombreTabla

**En la base DWH** las tablas que representes las dimensiones empezarán con DIMNombreTabla , las tablas de hechos empezaran con FACTNombreTabla

**NombreTabla:** tiene que estar relacionada con el contenido de la información y si contiene más de una palabra debe separarse con el guión bajo por ejemplo TD\_FECHA\_TRAFICO.(TANQUE), DIMFECHA\_TRAFICO(DWH).

**Nombres de los campos** Deben empezar con los tres caracteres del NombreTabla, separados por el guión bajo seguidos de tres caracteres que indiquen el contenido del campo si para indicar el contenido del campo es necesario más de una palabra se debe formar el nombre con las tres primeras letras de cada una de las palabras separadas por un guión bajo a excepción de la tabla fecha\_trafico que puede especificarse en una sola palabra el contenido del dato. Ejemplo:

| <b>Tabla</b>     | <b>Campo</b>             |
|------------------|--------------------------|
| TD_CENTRALES     | CEN_COD (Código Central) |
| TD_FECHA_TRAFICO | MES                      |

Estas reglas se aplican a las nuevas tablas creadas para el objetivo porque para mantener el espejo de base fuente se utilizó los mismos nombres de tablas y campos ya establecidos.

### 2.1.1. Base de datos TANQUE

#### 2.1.1.1. Tablas para Tráfico Telefónico

| <b>CAMPO</b> | <b>TIPO</b>  | <b>TABLA</b> | <b>FUNCION TABLA</b>                                 | <b>DESCRIPCION</b>                |
|--------------|--------------|--------------|------------------------------------------------------|-----------------------------------|
| PAINUM       | INT          | PAIS         | Permitir construir la dimensión ubicación            | Número del País                   |
| PAICOD       | INT          | PAIS         | Permitir construir la dimensión ubicación            | Código del País                   |
| PAINOM       | VARCHAR (30) | PAIS         | Permitir construir la dimensión ubicación            | Nombre del País                   |
| SALFEC       | DATETIME     | SCELULAR     | Permitir la construcción th_salida_trafico(celeular) | Fecha de la llamada               |
| SALHOR       | SMALLINT     | SCELULAR     | Permitir la construcción th_salida_trafico(celeular) | Hora de la llamada                |
| SALORI       | VARCHAR (9)  | SCELULAR     | Permitir la construcción th_salida_trafico(celeular) | Número telefónico realiza llamada |
| SALDES       | VARCHAR (9)  | SCELULAR     | Permitir la construcción th_salida_trafico(celeular) | Número telefónico recibe llamada  |
| SALDUR       | FLOAT        | SCELULAR     | Permitir la construcción th_salida_trafico(celeular) | Duración en segundos              |
| SALOPE       | VARCHAR (1)  | SCELULAR     | Permitir la construcción th_salida_trafico(celeular) | Código de operadora               |

|         |                 |                       |                                                                                       |                                        |
|---------|-----------------|-----------------------|---------------------------------------------------------------------------------------|----------------------------------------|
| SALTAR  | VARCHAR<br>(1)  | SCELULAR              | Permitir la construcción<br>th_salida_trafico(celular)                                | Código de la<br>tarifa(n/r)            |
| SALCEN  | SMALLINT        | SCELULAR              | Permitir la construcción<br>th_salida_trafico(celular)                                | Código central                         |
| SALDIR  | VARCHAR<br>(1)  | SCELULAR              | Permitir la construcción<br>th_salida_trafico(celular)                                | Código tipo de<br>acceso               |
| CON_COD | VARCHAR<br>(15) | SERVICIO_3<br>DIGITOS | Permitir construir la<br>dimensión contacto externo                                   | Número servicio tres<br>dígitos        |
| CON_DES | VARCHAR<br>(40) | SERVICIO_3<br>DIGITOS | Permitir construir la<br>dimensión contacto externo                                   | Descripción servicio<br>3_dígitos      |
| SALFEC  | DATETIME        | SESPECIA              | Permitir la construcción<br>th_salida_trafico(1-800)<br>y th_trafico_local(3-dígitos) | Fecha de la llamada                    |
| SALHOR  | SMALLINT        | SESPECIA              | Permitir la construcción<br>th_salida_trafico(1-800)<br>y th_trafico_local(3-dígitos) | Hora de la llamada                     |
| SALORI  | VARCHAR<br>(9)  | SESPECIA              | Permitir la construcción<br>th_salida_trafico(1-800)<br>y th_trafico_local(3-dígitos) | Número Telefónico<br>realiza llamada   |
| SALDES  | VARCHAR<br>(12) | SESPECIA              | Permitir la construcción<br>th_salida_trafico(1-800)<br>y th_trafico_local(3-dígitos) | Número<br>Telefónico recibe<br>llamada |
| SALDUR  | FLOAT           | SESPECIA              | Permitir la construcción<br>th_salida_trafico(1-800)<br>y th_trafico_local(3-dígitos) | Duración de la<br>llamada en segundos  |
| SALCEN  | SMALLINT        | SESPECIA              | Permitir la construcción<br>th_salida_trafico(1-800)<br>y th_trafico_local(3-dígitos) | Código de la central                   |
| SALFEC  | DATETIME        | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Fecha de la llamada                    |
| SALHOR  | SMALLINT        | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Hora de la llamada                     |
| SALORI  | VARCHAR<br>(9)  | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Número Telefónico<br>realiza llamada   |
| SALDES  | VARCHAR<br>(12) | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Número<br>Telefónico recibe<br>llamada |
| SALDUR  | FLOAT           | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Duración llamada en<br>segundos        |
| SALOPE  | VARCHAR<br>(1)  | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Código de operadora                    |
| SALCEN  | SMALLINT        | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Código de la central                   |
| SALDIR  | VARCHAR<br>(1)  | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Código de acceso                       |
| SALDUI  | FLOAT           | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Minutos<br>redondeados<br>segmento 6   |
| SALPAI  | INT             | SINTERNA              | Permitir la construcción<br>th_salida_trafico(internacional<br>)                      | Número del País                        |

|        |             |                   |                                                           |                                    |
|--------|-------------|-------------------|-----------------------------------------------------------|------------------------------------|
| SALCUI | INT         | SINTERNA          | Permitir la construcción th_salida_trafico(internacional) | Código de la ciudad.               |
| SALFEC | DATETIME    | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Fecha de la llamada                |
| SALHOR | SMALLINT    | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Hora de la llamada                 |
| SALORI | VARCHAR(9)  | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Número Telefónico realiza llamada  |
| SALDES | VARCHAR(9)  | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Número Telefónico recibe llamada   |
| SALDUR | FLOAT       | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Duración de la llamada en segundos |
| SALOPE | VARCHAR(1)  | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Código de operadora                |
| SALZON | VARCHAR(1)  | SNACIONA          | Permitir la construcción th_salida_trafico(nacional)      | Código de la zona provincia        |
| SALFEC | DATETIME    | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Fecha de la llamada                |
| SALHOR | SMALLINT    | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Hora de la llamada                 |
| SALORI | VARCHAR(9)  | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Número Telefónico realiza llamada  |
| SALDES | VARCHAR(9)  | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Número Telefónico recibe llamada   |
| SALDUR | FLOAT       | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Duración segundos                  |
| SALOPE | VARCHAR(1)  | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Código de operadora                |
| SALZON | VARCHAR(1)  | SREGIONA          | Permitir la construcción th_salida_trafico(regional)      | Código de la zona o región.        |
| SALDES | VARCHAR(12) | TRAFICO_SERVICIOS | Permite construir la th_trafico_local                     | Num. telefónico 3 dígitos          |
| SALORI | VARCHAR(9)  | TRAFICO_SERVICIOS | Permite construir la th_trafico_local                     | Número que realiza llamada         |
| SALFEC | DATETIME    | TRAFICO_SERVICIOS | Permite construir la th_trafico_local                     | Fecha de la llamada                |
| SALDUR | FLOAT       | TRAFICO_SERVICIOS | Permite construir la th_trafico_local                     | Duración segundos                  |
| SALCEN | SMALLINT    | TRAFICO_SERVICIOS | Permite construir la th_trafico_local                     | Código central                     |
| SALFEC | DATETIME    | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Fecha llamada                      |
| SALHOR | SMALLINT    | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Hora llamada                       |
| SALORI | VARCHAR(9)  | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Número realiza llamada             |
| SALDES | VARCHAR(12) | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Número recibe llamada              |
| SALDUR | FLOAT       | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Duración minutos                   |
| SALCEN | SMALLINT    | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Código central                     |
| TIPO   | VARCHAR(1)  | TRAFICO_1800      | Permite construir la th_salida_trafico(1-800)             | Identifica el destino 1-800        |

|           |              |                     |                                                                                                                |                                                             |
|-----------|--------------|---------------------|----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------|
| CONNUMERO | INT          | CONSUMO_LOCAL       | Permite obtener el consumo mensual para luego construir th_trafico_local                                       | Número telefónico                                           |
| CONSUMO   | FLOAT        | CONSUMO_LOCAL       | Permite obtener el consumo mensual para luego construir th_trafico_local                                       | Volumen consumido en minutos                                |
| CONANIO   | SMALLINT     | CONSUMO_LOCAL       | Permite obtener el consumo mensual para luego construir th_trafico_local                                       | Año de consumo                                              |
| CONMES    | SMALLINT     | CONSUMO_LOCAL       | Permite obtener el consumo mensual para luego construir th_trafico_local                                       | Mes de consumo                                              |
| TARCOD    | INT          | TARAR7              | Permite construir la dimensión cliente                                                                         | Código de Categoría                                         |
| TARDE4    | VARCHAR (30) | TARAR7              | Permite construir la dimensión cliente                                                                         | Descripción de Categoría                                    |
| SALFEC    | DATETIME     | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Fecha                                                       |
| SALHOR    | SMALLINT     | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Hora                                                        |
| SALORI    | INT          | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Número realiza llamada                                      |
| SALCEO    | SMALLINT     | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Central de origen llamada                                   |
| SALDES    | INT          | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Número recibe llamada                                       |
| SALCED    | SMALLINT     | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Central que recibe llamada                                  |
| SALDUR    | FLOAT        | SLOCAL              | Permite construir la th_trafico_local_detallado                                                                | Duración minutos                                            |
| NUM_180   | VARCHAR (15) | AUX_DIRECTORIO_1800 | Tabla temporal que permite hacer limpieza de los datos y además rellenar con ceros hasta completar a 6 dígitos | Número telefónico 1800                                      |
| DES_180   | VARCHAR (60) | AUX_DIRECTORIO_1800 | Tabla temporal que permite hacer limpieza de los datos y además rellenar con ceros hasta completar a 6 dígitos | Descripción del número telefónico                           |
| COB_180   | VARCHAR (15) | AUX_DIRECTORIO_1800 | Tabla temporal que permite hacer limpieza de los datos y además rellenar con ceros hasta completar a 6 dígitos | Cobertura Nacional, reg. Internacional                      |
| ENTFEC    | DATETIME     | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Fecha llamada                                               |
| ENTHOR    | SMALLINT     | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Hora llamada                                                |
| ENTORI    | VARCHAR (9)  | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Número telefónico que realiza la llamada (contacto externo) |
| ENTDES    | INT          | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Número telefónico que realiza la llamada (abonado)          |
| ENTDUR    | FLOAT        | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Duración minutos llamada                                    |
| ENTTIP    | VARCHAR (1)  | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Tipo de llamada nacional/regional                           |
| ENTOPE    | VARCHAR (1)  | ENACIONA            | Permite construir th_entrada_trafico(nacional)                                                                 | Código Operadora                                            |

|         |              |          |                                                     |                                                      |
|---------|--------------|----------|-----------------------------------------------------|------------------------------------------------------|
| ENTZON  | SMALLINT     | ENACIONA | Permite construir th_entrada_trafico(nacional)      | Código de la zona                                    |
| ENTCEN  | SMALLINT     | ENACIONA | Permite construir th_entrada_trafico(nacional)      | Código central                                       |
| PAI_NUM | INT          | CIUDAD   | Permite construir la dimensión ubicación            | Número del País                                      |
| CIU_COD | INT          | CIUDAD   | Permite construir la dimensión ubicación            | Código de la ciudad                                  |
| CIU_NOM | VARCHAR (50) | CIUDAD   | Permite construir la dimensión ubicación            | Nombre Ciudad                                        |
| ENTFEC  | DATETIME     | EREGIONA | Permite construir th_entrada_trafico(regional)      | Fecha llamada                                        |
| ENTHOR  | SMALLINT     | EREGIONA | Permite construir th_entrada_trafico(regional)      | Hora llamada                                         |
| ENTORI  | VARCHAR (9)  | EREGIONA | Permite construir th_entrada_trafico(regional)      | Número telefónico realiza llamada (contacto externo) |
| ENTDES  | INT          | EREGIONA | Permite construir th_entrada_trafico(regional)      | Número telefónico que recibe llamada (abonado)       |
| ENTDUR  | FLOAT        | EREGIONA | Permite construir th_entrada_trafico(regional)      | Duración segundos llamada                            |
| ENTTIP  | VARCHAR (1)  | EREGIONA | Permite construir th_entrada_trafico(regional)      | Tipo de llamada nacional/regional                    |
| ENTOPE  | VARCHAR (1)  | EREGIONA | Permite construir th_entrada_trafico(regional)      | Código Operadora                                     |
| ENTZON  | SMALLINT     | EREGIONA | Permite construir th_entrada_trafico(regional)      | Código de la zona                                    |
| ENTCEN  | SMALLINT     | EREGIONA | Permite construir th_entrada_trafico(regional)      | Código central                                       |
| ENTFEC  | DATETIME     | EINTERNA | Permite construir th_entrada_trafico(internacional) | Fecha llamada                                        |
| ENTHOR  | SMALLINT     | EINTERNA | Permite construir th_entrada_trafico(internacional) | Hora llamada                                         |
| ENTORI  | VARCHAR (9)  | EINTERNA | Permite construir th_entrada_trafico(internacional) | Número telefónico realiza llamada (contacto externo) |
| ENTDES  | INT          | EINTERNA | Permite construir th_entrada_trafico(internacional) | Número telefónico que recibe llamada (abonado)       |
| ENTDUR  | FLOAT        | EINTERNA | Permite construir th_entrada_trafico(internacional) | Duración segundos llamada                            |
| ENTOPE  | VARCHAR (1)  | EINTERNA | Permite construir th_entrada_trafico(internacional) | Código Operadora                                     |
| ENTDUI  | FLOAT        | EINTERNA | Permite construir th_entrada_trafico(internacional) | Duración redondeo segmento 6                         |
| ENTCEN  | SMALLINT     | EINTERNA | Permite construir th_entrada_trafico(internacional) | Código central                                       |
| ENTFEC  | DATETIME     | ECELULAR | Permite construir th_entrada_trafico(celular)       | Fecha llamada                                        |
| ENTHOR  | SMALLINT     | ECELULAR | Permite construir th_entrada_trafico(celular)       | Hora llamada                                         |

|         |                 |                     |                                                                                                                                      |                                                         |
|---------|-----------------|---------------------|--------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------|
| ENTORI  | VARCHAR<br>(9)  | ECELULAR            | Permite construir<br>th_entrada_trafico(celular)                                                                                     | Número celular<br>realiza llamada<br>(contacto externo) |
| ENTDES  | INT             | ECELULAR            | Permite construir<br>th_entrada_trafico(celular)                                                                                     | Número telefónico<br>recibe llamada<br>(abonado)        |
| ENTDUR  | FLOAT           | ECELULAR            | Permite construir<br>th_entrada_trafico(celular)                                                                                     | Duración segundos<br>llamada                            |
| ENTTAR  | VARCHAR<br>(1)  | ECELULAR            | Permite construir<br>th_entrada_trafico(celular)                                                                                     | Tipo de tarifa<br>normal/reducida                       |
| ENTOPE  | VARCHAR<br>(1)  | ECELULAR            | Permite construir<br>th_entrada_trafico(celular)                                                                                     | Código Operadora                                        |
| ENTCEN  | SMALLINT        | ECELULAR            | Permite construir<br>th_entrada_trafico(celular)                                                                                     | Código central                                          |
| NUM_180 | VARCHAR<br>(15) | DIRECTORIO<br>_1800 | Permite construir la tabla de<br>hechos th_trafico_salida(1-<br>800) y la dimensión contacto<br>externo td_contacto_externo          | Número telefónico<br>1800                               |
| DES_180 | VARCHAR<br>(60) | DIRECTORIO<br>_1800 | Permite construir la tabla de<br>hechos th_trafico_salida(1-<br>800) y la dimensión contacto<br>externo td_contacto_externo          | Descripción del<br>número telefónico                    |
| COB_180 | VARCHAR<br>(15) | DIRECTORIO<br>_1800 | Permite construir la tabla de<br>hechos th_trafico_salida (1-<br>800) y la dimensión contacto<br>externo td_contacto_externo         | Cobertura Nacional,<br>reg. Internacional               |
| TELTEL  | INT             | TELAR_<br>MES       | Permite construir<br>th_salida_trafico;<br>th_entrada_trafico_th_trafico_<br>local;th_trafico_local_detallad<br>o;td_cliente_trafico | Número telefónico<br>(abonado)                          |
| TELNOM  | VARCHAR<br>(30) | TELAR_<br>MES       | Permite construir<br>th_salida_trafico;<br>th_entrada_trafico_th_trafico_<br>local;th_trafico_local_detallad<br>o;td_cliente_trafico | Nombre del abonado                                      |
| TELTAR  | SMALLINT        | TELAR_<br>MES       | Permite construir<br>th_salida_trafico;<br>th_entrada_trafico_th_trafico_<br>local;th_trafico_local_detallad<br>o;td_cliente_trafico | Código de la<br>categoría                               |

### 2.1.1.2. Tablas para el control de calidad Conmutación

|         |                 |         |                                          |                                     |
|---------|-----------------|---------|------------------------------------------|-------------------------------------|
| CENTRAL | VARCHAR<br>(10) | CLASE_A | Permite la construcción<br>th_auxiliares | Nombre de la central                |
| FECHA   | DATETIME        | CLASE_A | Permite la construcción<br>th_auxiliares | Fecha genera tráfico                |
| HORA    | FLOAT           | CLASE_A | Permite la construcción<br>th_auxiliares | Hora genera el<br>tráfico           |
| A01     | FLOAT           | CLASE_A | Permite la construcción<br>th_auxiliares | Cantidad Media RF<br>ocupados       |
| AS1     | FLOAT           | CLASE_A | Permite la construcción<br>th_auxiliares | Cantidad media de<br>RF en servicio |
| AP1     | FLOAT           | CLASE_A | Permite la construcción<br>th_auxiliares | Cantidad de tomas<br>de RF          |

|         |              |          |                                       |                                             |
|---------|--------------|----------|---------------------------------------|---------------------------------------------|
| AR1     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de rechazo de toma por falta de RF |
| AE1     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad tomas de RF                        |
| AQ1     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de RF equipados                    |
| A02     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad Media CCF ocupados                 |
| AS2     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad CCF servicio                       |
| AP2     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de tomas Eficaces CCF              |
| AR2     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad rechazo x falta CCF                |
| AE2     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de tomas Eficaces de CCF           |
| AQ2     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de CCF equipados                   |
| A04     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad Media V23 ocupados                 |
| AS4     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad media V23 en servicio              |
| AP4     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de intentos V23                    |
| AR4     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad rechazo                            |
| AE4     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de tomas Ef. V23                   |
| AQ4     | FLOAT        | CLASE_A  | Permite la construcción th_auxiliares | Cantidad de V23 equipados                   |
| CENTRAL | VARCHAR (10) | CLASE_FB | Permite la construcción th_haces      | Nombre de la central                        |
| FECHA   | VARCHAR (10) | CLASE_FB | Permite la construcción th_haces      | Fecha                                       |
| HORA    | SMALLINT     | CLASE_FB | Permite la construcción th_haces      | Hora                                        |
| HAZ     | VARCHAR (6)  | CLASE_FB | Permite la construcción th_haces      | Nombre/Código del haz                       |
| FOA     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Tráfico a la llegada                        |
| FOD     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Tráfico a la salida                         |
| FS1     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Cantidad media en servicio                  |
| FEA     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Cantidad de llamadas Efi. a la llegada      |
| FED     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Cantidad de llamadas Efi. a la salida       |
| FQ1     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Cantidad circuitos equipados                |
| FPA     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Cantidad tomas a la llegada                 |
| FPD     | FLOAT        | CLASE_FB | Permite la construcción th_haces      | Cantidad tomas a la salida                  |

|         |          |          |                                  |                                                                 |
|---------|----------|----------|----------------------------------|-----------------------------------------------------------------|
| FRA     | FLOAT    | CLASE_FB | Permite la construcción th_haces | Cantidad tomas rechazadas a la llegada                          |
| FRD     | FLOAT    | CLASE_FB | Permite la construcción th_haces | Cantidad tomas rechazadas a la salida                           |
| PRS1    | FLOAT    | CLASE_FB | Permite la construcción th_haces | Cantidad de doble toma                                          |
| CENTRAL | SMALLINT | CLASE_T  | Permite construir el th_llamadas | Nombre de la central                                            |
| FECHA   | DATETIME | CLASE_T  | Permite construir el th_llamadas | Fecha                                                           |
| HORA    | SMALLINT | CLASE_T  | Permite construir el th_llamadas | Hora                                                            |
| T01_I   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas con respuesta Tráfico interno                 |
| T01_D   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad de llamadas con respuesta Tráfico salida               |
| T02_I   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas efi. con regulación UR Tráfico interno        |
| T02_D   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas efi. con regulación UR tráfico salida         |
| T04_I   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad de llamadas tratadas tráfico Interno                   |
| T04_D   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad de llamadas tratadas tráfico salida                    |
| T04_A   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad de llamadas tratadas tráfico Llegada                   |
| T04_T   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad de llamadas tratadas tráfico tránsito                  |
| T05_I   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas presentadas Tráfico interno                   |
| T05_D   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas presentadas Tráfico salida                    |
| T05_A   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas presentadas tráfico llegada                   |
| T05_T   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas presentadas tráfico. tránsito                 |
| T06_D   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas salida con toma de circuito trafi. Salida     |
| T06_T   | FLOAT    | CLASE_T  | Permite construir el th_llamadas | Cantidad llamadas salida con toma de circuito tráfico. Tránsito |

|       |       |         |                                  |                                                                          |
|-------|-------|---------|----------------------------------|--------------------------------------------------------------------------|
| T07_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Llamadas presentadas por los abonado RDSI tráfico interno                |
| T07_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Llamadas presentadas por abonado RDSI trafico salida                     |
| T07_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Cantidad llamadas presentadas por abonado RDSI trafico llegada           |
| T07_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Cantidad llamadas presentadas por abonado RDSI trafico tránsito          |
| T12_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado antes de la recepción de marcación completa tráfico interno      |
| T12_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado antes de la recepción de marcación completa tráfico salida       |
| T12_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado antes de la recepción de la marcación completa tráfico llegada   |
| T12_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado antes de la recepción de la marcación completa tráfico tránsito  |
| T13_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Maniobras prohibidas Tráfico interno                                     |
| T13_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Maniobras prohibidas Tráfico salida                                      |
| T13_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Maniobras prohibidas Tráfico llegada                                     |
| T13_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Maniobras prohibidas. Tráfico Transito                                   |
| T14_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Error en el número tráfico interno                                       |
| T14_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Error en el número tráfico llegada                                       |
| T15_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado entre el la marcación y el final de la selección tráfico interno |
| T15_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado entre la marcación y el final de la selección tráfico salida     |

|       |       |         |                                  |                                                                  |
|-------|-------|---------|----------------------------------|------------------------------------------------------------------|
| T15_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado la marcación y el final de la selección tráfico llegada  |
| T15_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado la marcación y el final de la selección tráfico tránsito |
| T16_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Llamada a un número fuera de tráfico interno                     |
| T16_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Llamada a un número fuera de servicio tráfico salida             |
| T16_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Llamada a un número fuera de servicio tráfico llegada            |
| T16_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Llamada a un numero fuera de tráfico tránsito                    |
| T17_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Prefijo no autorizado Tráfico interno                            |
| T17_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Prefijo no autorizado Tráfico salida                             |
| T17_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Prefijo autorizado Tráfico llegada                               |
| T17_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Prefijo no autorizado Tráfico tránsito                           |
| T18_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado durante el timbre Tráfico interno                        |
| T18_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado durante el timbre Tráfico salida                         |
| T18_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado el timbre Tráfico llegada                                |
| T18_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado durante el timbre Tráfico tránsito                       |
| T19_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Error de marcación tráfico salida                                |
| T19_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Error de marcación tráfico interno                               |
| T22_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Línea ocupada Tráfico interno                                    |
| T22_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Línea ocupada Tráfico salida                                     |
| T22_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Línea ocupada Tráfico llegada                                    |
| T22_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Línea ocupada Tráfico tránsito                                   |
| T23_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Ausencia de respuesta tráfico interno                            |
| T23_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Ausencia de respuesta tráfico salida                             |

|       |       |         |                                  |                                                        |
|-------|-------|---------|----------------------------------|--------------------------------------------------------|
| T23_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Ausencia de respuesta tráfico llegada                  |
| T23_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Ausencia de respuesta tráfico tránsito                 |
| T24_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado ante la ausencia de respuesta Tráfico interno  |
| T24_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado ante la ausencia de respuesta Tráfico salida   |
| T24_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado ante la ausencia de respuesta Tráfico llegada  |
| T24_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Colgado ante la ausencia de respuesta Tráfico tránsito |
| T32_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Falta de auxiliares Tráfico interno                    |
| T32_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Falta de auxiliares Tráfico llegada                    |
| T33_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Falta de auxiliares Tráfico interno                    |
| T34_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Fallo interno Trafico interno                          |
| T34_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Fallo interno Trafico salida                           |
| T34_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Fallo interno Trafico llegada                          |
| T34_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Fallo interno Trafico Tránsito                         |
| T35_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA Tráfico interno                            |
| T35_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA Tráfico llegada                            |
| T36_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA tráfico interno                            |
| T36_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA tráfico salida                             |
| T36_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA tráfico llegada                            |
| T36_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA tráfico tránsito                           |
| T37_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo regulación del comando tráfico interno         |
| T37_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo regulación del comando tráfico salida          |
| T37_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo regulación del comando tráfico llegada         |

|       |       |         |                                  |                                                            |
|-------|-------|---------|----------------------------------|------------------------------------------------------------|
| T37_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo por regulación del comando tráfico tránsito        |
| T38_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA Tráfico interno                                |
| T38_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo URA Tráfico tránsito                               |
| T39_I | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo otra causa Tráfico interno                         |
| T39_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo otra causa Tráfico salida                          |
| T39_A | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo otra causa Tráfico llegada                         |
| T39_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Bloqueo otra causa Tráfico tránsito                        |
| T42_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Falta de circuitos tráfico salida                          |
| T42_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Falta de circuitos tráfico tránsito                        |
| T43_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Problema de señalización tráfico salida                    |
| T43_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Problema de señalización tráfico Tránsito                  |
| T44_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Desbordamiento de temporización más abajo tráfico salida   |
| T44_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Desbordamiento de temporización más abajo trafico tránsito |
| T45_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Congestión del conmutador distante tráfico salida          |
| T45_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Congestión del conmutador distante tráfico tránsito        |
| T46_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Problema de señalización tráfico salida                    |
| T46_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Problema de señalización tráfico tránsito                  |
| T47_D | FLOAT | CLASE_T | Permite construir el th_llamadas | fracaso más abajo causo no identificada tráfico salida     |
| T47_T | FLOAT | CLASE_T | Permite construir el th_llamadas | fracaso mas abajo causo no identificada tráfico tránsito   |
| T48_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Doble toma trafico salida                                  |
| T48_T | FLOAT | CLASE_T | Permite construir el th_llamadas | Doble toma trafico tránsito                                |
| T49_D | FLOAT | CLASE_T | Permite construir el th_llamadas | Filtrados trafico salida                                   |

|         |              |         |                                  |                                                             |
|---------|--------------|---------|----------------------------------|-------------------------------------------------------------|
| T49_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Filtrados trafico tránsito                                  |
| T50_I   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Rechazo por espaciamiento de llamadas Tráfico interno       |
| T50_D   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Rechazo por espaciamiento de llamadas Tráfico salida        |
| T50_A   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Rechazo por espaciamiento de llamadas Tráfico llegada       |
| T50_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Rechazo por espaciamiento de llamadas Tráfico tránsito      |
| T51_D   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Congestión de señalización tráfico salida                   |
| T51_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Congestión de señalización tráfico tránsito                 |
| T54_A   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Desbordamiento de temporización más arriba tráfico llegada  |
| T54_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Desbordamiento de temporización más arriba tráfico tránsito |
| T56_A   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Fallo de señalización más arriba tráfico llegada            |
| T56_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Fallo de señalización más arriba tráfico tránsito           |
| T62_I   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa servidor tráfico interno                              |
| T62_D   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa servidor tráfico salida                               |
| T62_A   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa servidor tráfico llegada                              |
| T62_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa servidor tráfico tránsito                             |
| T64_I   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa diversa tráfico interno                               |
| T64_D   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa diversa tráfico salida                                |
| T64_A   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa diversa tráfico llegada                               |
| T64_T   | FLOAT        | CLASE_T | Permite construir el th_llamadas | Causa diversa tráfico tránsito                              |
| CENTRAL | VARCHAR (10) | CLASE_R | Permite construir el th_URA      | Central                                                     |
| FECHA   | VARCHAR (12) | CLASE_R | Permite construir el th_URA      | Fecha                                                       |
| HORA    | SMALLINT     | CLASE_R | Permite construir el th_URA      | Hora                                                        |
| URA     | SMALLINT     | CLASE_R | Permite construir el th_URA      | Código de URA                                               |

|         |              |         |                                        |                                                     |
|---------|--------------|---------|----------------------------------------|-----------------------------------------------------|
| RPA     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad llamadas presentadas sobre abonados libres |
| RRA     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad llamadas de entrada rechazadas bloqueo URA |
| RAA     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad llamadas terminadas presentadas a las URAs |
| RPD     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad llamadas originadas por la central         |
| RRD     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad llamadas bloqueo de URA                    |
| RAD     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad llamadas originadas por los abonados       |
| RRX     | FLOAT        | CLASE_R | Permite construir el th_URA            | Cantidad conexiones                                 |
| CENTRAL | VARCHAR (10) | CLASE_C | Permite construir la th_trafico_erlang | Código central                                      |
| FECHA   | DATETIME     | CLASE_C | Permite construir la th_trafico_erlang | Fecha                                               |
| HORA    | SMALLINT     | CLASE_C | Permite construir la th_trafico_erlang | Hora                                                |
| CT      | FLOAT        | CLASE_C | Permite construir la th_trafico_erlang | Cantidad tráfico de tránsito                        |
| CV      | FLOAT        | CLASE_C | Permite construir la th_trafico_erlang | Cantidad tráfico indefinido                         |
| CI      | FLOAT        | CLASE_C | Permite construir la th_trafico_erlang | Cantidad tráfico interno                            |
| CC      | FLOAT        | CLASE_C | Permite construir la th_trafico_erlang | Cantidad tráfico de salida                          |
| CA      | FLOAT        | CLASE_C | Permite construir la th_trafico_erlang | Cantidad tráfico de llegada                         |

Existen más tablas en la base TANQUE, pero ya no es necesario detallarlas ya que se encuentran especificadas en el anexo del entregable de identificación de fuentes (metadata-anexo I)

TD\_AUXILIARES

TD\_FECHA\_TRAFICO

TD\_CENTRALES

TD\_HACES

TD\_CLIENTE\_TRAFICO

TD\_LLAMADA\_URA

TD\_CONTACTO\_EXTERNO

TD\_TARIFA

TD\_DESCRIPCION\_AUXILIARES

TD\_TIPO\_FLUJO

TD\_DESCRIPCION\_HACES

TD\_TIPO\_LOCAL

TD\_DESTINO\_ENTRADA

TD\_TIPO\_LLAMADA

TD\_DESTINO\_SALIDA

TD\_UBICACION

|                    |                            |
|--------------------|----------------------------|
| TD_URA             | TH_SALIDA_TRAFICO          |
| TH_AUXILIARES      | TH_TRAFICO_ERLANGS         |
| TH_ENTRADA_TRAFICO | TH_TRAFICO_LOCAL           |
| TH_HACES           | TH_TRAFICO_LOCAL_DETALLADO |
| TH_LLAMADAS        | TH_URA                     |

### 2.1.2. Base de datos DWH

En la base DWH se encuentran las siguientes tablas tienen los mismos nombres y tipos de datos que las tablas antes mencionadas en la base TANQUE, lo único que cambia es el nombre de la tabla las TD\_ por DIM y las TH\_ por FACT

|                            |                              |
|----------------------------|------------------------------|
| DIM_AUXILIARES             | DIM_TIPO_LLAMADA             |
| DIM_CENTRALES              | DIM_TIPO_LOCAL               |
| DIM_CLIENTE_TRAFICO        | DIM_UBICACION                |
| DIM_CONTACTO_EXTERNO       | DIM_URA                      |
| DIM_DESCRIPCION_AUXILIARES | FACT_AUXILIARES              |
| DIM_DESCRIPCION_HACES      | FACT_HACES                   |
| DIM_DESTINO_ENTRADA        | FACT_LLAMADAS                |
| DIM_DESTINO_SALIDA         | FACT_TRAFICO_SALIENTE        |
| DIM_FECHA_TRAFICO          | FACT_TRAFICO_ENTRANTE        |
| DIM_HACES                  | FACT_TRAFICO_ERLANGS         |
| DIM_LLAMADA_URA            | FACT_TRAFICO_LOCAL           |
| DIM_TARIFA                 | FACT_TRAFICO_LOCAL_DETALLADO |
| DIM_TIPO_FLUJO             | FACTURA                      |

## GLOSARIO

|                                  |                                                                                                                                       |
|----------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|
| <b>.BAT</b>                      | Es un archivo de procesamiento por lotes, contiene un conjunto de comandos <i>DOS</i> . Permite automatizar diversas tareas.          |
| <b>.dtsx</b>                     | Es un paquete de la herramienta <i>Integration Services</i> , que contienen los procedimientos de extracción, transformación y carga. |
| <b>AD</b>                        | <i>Application Development</i> . Desarrollo de aplicaciones.                                                                          |
| <b>Ad-hoc</b>                    | Consulta sobre tablas dinámicas.                                                                                                      |
| <b>Agregaciones</b>              | Proceso de calcular la <i>data</i> sumariada.                                                                                         |
| <b>ALCATEL</b>                   | Es un proveedor de <i>hardware</i> , <i>software</i> y servicios para las Telecomunicaciones.                                         |
| <b>Analysis Services</b>         | Es una herramienta que sirve para la construcción de cubos y elaboración de reportes.                                                 |
| <b>Aseguramiento de ingresos</b> | Es un área de la empresa encargada de vigilar el comportamiento del tráfico para evitar fraudes.                                      |
| <b>Auxiliares</b>                | Es un recurso de la red telefónica que sirve para dar tono cuando se descuelga el auricular.                                          |
| <b>Auxiliares en servicio</b>    | Auxiliares disponibles para cursar el tráfico.                                                                                        |
| <b>Auxiliares equipados</b>      | Auxiliares que existen en la central y pueden estar en servicio ó no.                                                                 |

|                                          |                                                                                                                                              |
|------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Auxiliares Ocupados</b>               | Auxiliares que tienen tráfico.                                                                                                               |
| <b>BI</b>                                | <i>Business Intelligence</i> . Inteligencia de Negocios                                                                                      |
| <b>Bypass</b>                            | Tráfico Telefónico ilegal.                                                                                                                   |
| <b>Miembro Calculados</b>                | Valores numéricos que no se almacenan en la base de datos multidimensional. Se derivan de datos almacenados en la base a través de cálculos. |
| <b>Carga</b>                             | Es la suma de tiempos de ocupación acumulados en los canales de conmutación ó líneas de abonados.                                            |
| <b>CDR's</b>                             | Código de registros. (Detalle de llamadas).                                                                                                  |
| <b>Checkpoint</b>                        | Pasar la información que está en la memoria virtual al disco.                                                                                |
| <b>Conmutación</b>                       | Es un área de la empresa que se encarga del control de la calidad de la circulación del tráfico.                                             |
| <b>Contadores Brutos</b>                 | Los contadores brutos registran de manera permanente los valores observados de los recursos del conmutador.                                  |
| <b>Coordinación con otras operadoras</b> | Es un área de la empresa encargada de la conciliación del tráfico con las operadoras con las que tiene un convenio.                          |
| <b>Cubo</b>                              | Es la representación multidimensional de la información.                                                                                     |
| <b>Data</b>                              | Datos que se cargan en el <i>Data Warehouse</i>                                                                                              |
| <b>Data Mart</b>                         | Mini <i>Data Warehouse</i> . Utilizado para un área de la empresa.                                                                           |

|                                |                                                                                                                                                  |
|--------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------|
| <b><i>Data Warehouse</i></b>   | Bodega de datos que almacena información para la toma de decisiones                                                                              |
| <b><i>Data Warehousing</i></b> | Proceso que lleva a la construcción de la bodega de datos.                                                                                       |
| <b><i>DBMS</i></b>             | <i>Data Base Manager System.</i> Sistema Gestor de base de datos.                                                                                |
| <b><i>DICE</i></b>             | Obtener partes de la información de un cubo.                                                                                                     |
| <b><i>Dimensiones</i></b>      | Tablas que rodean a la tabla de hecho en el esquema estrella. Permiten la granularidad y las agregaciones.                                       |
| <b><i>Dot Modeling</i></b>     | Modelado Punto.                                                                                                                                  |
| <b><i>DSS</i></b>              | <i>Decision Support System.</i> Sistema de soporte de Decisiones.                                                                                |
| <b><i>DWA</i></b>              | <i>Data Warehouse Architecture.</i> Arquitectura del <i>Data Warehouse.</i>                                                                      |
| <b><i>DWH</i></b>              | <i>Data Warehouse.</i>                                                                                                                           |
| <b><i>EASI</i></b>             | <i>Evolvability/availability, sharability, integrity.</i> Evolución, disponibilidad, compartir e integración.                                    |
| <b><i>E.B.C.D.I.C.</i></b>     | Es un código binario que representa caracteres alfanuméricos, controles y signos de puntuación. Cada caracter está compuesto por 8 bits = 1 byte |
| <b><i>EIS</i></b>              | <i>Executive Information System.</i> Sistema de Información Ejecutiva.                                                                           |
| <b><i>ER</i></b>               | <i>Entity Relationship.</i> Entidad Relación.                                                                                                    |

|                              |                                                                                                                            |
|------------------------------|----------------------------------------------------------------------------------------------------------------------------|
| <b>Erlangs</b>               | Unidad de intensidad de tráfico que corresponde a la ocupación de un sólo recurso.                                         |
| <b>ERP</b>                   | <i>Enterprise Resource Planning</i> . Sistema de planeación de recursos de la empresa.                                     |
| <b>Esquema</b>               | Colección de tablas.                                                                                                       |
| <b>Esquema Copo de Nieve</b> | Es la representación gráfica del modelo de datos que relaciona una dimensión a un hecho(tabla) a través de otra dimensión. |
| <b>Esquema Estrella</b>      | Es la representación gráfica del modelo de datos que relaciona directamente una dimensión con la tabla de hechos.          |
| <b>E.T.A.P.A.</b>            | Empresa Pública Municipal de Telecomunicaciones, Agua Potable Alcantarillado y Saneamiento de Cuenca.                      |
| <b>ETL</b>                   | <i>Extraction, Transformation Load</i> . Extracción, transformación y carga.                                               |
| <b>GCM</b>                   | <i>General Conceptual Model</i> . Modelo Conceptual General.                                                               |
| <b>Granularidad</b>          | En la jerarquía, ordenar las columnas de lo más general a lo más específico (detalle).                                     |
| <b>GUI</b>                   | <i>Graphical User Interface</i> . Interfaz gráfica de Usuario.                                                             |
| <b>HASH</b>                  | Es un método de búsqueda que genera claves.                                                                                |
| <b>Haz</b>                   | Une dos centrales telefónicas a través del cual fluye la información (llamadas).                                           |

|                             |                                                                                                                                                                    |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Hecho</b>                | Valores numéricos en la tabla principal del esquema estrella o copo de nieve (tabla de hechos).                                                                    |
| <b>HOLAP</b>                | <i>Hybrid Online Analytical Processing</i> . Proceso Analítico en Línea Híbrido.                                                                                   |
| <b>ID</b>                   | <i>Infrastructure Deployment</i> . Implementación de desarrollo.                                                                                                   |
| <b>INEC</b>                 | Instituto Nacional de Estadísticas y Censos del Ecuador.                                                                                                           |
| <b>Integración</b>          | Conjunto de reglas de procedimiento para asegurar que la data esté estandarizada en el <i>DWH</i> .                                                                |
| <b>IT</b>                   | <i>Technology Information</i> . Tecnología de Información.                                                                                                         |
| <b>JAD</b>                  | <i>Joint Application Development</i> . Desarrollo de aplicaciones colectivas.                                                                                      |
| <b>Jerarquía</b>            | Componentes de una dimensión.                                                                                                                                      |
| <b>Kimball</b>              | Precursor del <i>Data Warehouse</i> .                                                                                                                              |
| <b>Llamadas Eficaces</b>    | Llamada que da lugar al descolgar el abonado llamado ó la aceptación de un comando de servicio suplementario.                                                      |
| <b>Llamadas Presentadas</b> | Llamada que ha dado lugar a la búsqueda de un circuito libre sobre el haz, seguido de un intento de toma, sin que haya necesariamente emisión de la señal de toma. |
| <b>MDDB</b>                 | <i>Multi Dimensional Data Base</i> . Base de Datos Multi-dimensional.                                                                                              |
| <b>Medida</b>               | Hecho, valor numérico.                                                                                                                                             |

|                                 |                                                                                                     |
|---------------------------------|-----------------------------------------------------------------------------------------------------|
| <b>Metadata</b>                 | Información acerca de los datos.                                                                    |
| <b>Miembros<br/>o Atributos</b> | Valores en las columnas de la dimensión que definen los niveles.                                    |
| <b>MOLAP</b>                    | <i>Multidimensional Online Analytical Processing</i> . Proceso Analítico en Línea Multidimensional. |
| <b>MSF</b>                      | <i>Microsoft Solution Framework</i> . Marco de solución de trabajo <i>Microsoft</i> .               |
| <b>Multidimensional</b>         | Permite el análisis de información en base a varios criterios.                                      |
| <b>Nivel</b>                    | Elemento de una jerarquía de dimensión.                                                             |
| <b>ODBC</b>                     | <i>Open Database Connectivity</i> . Conexión de orígenes de Datos.                                  |
| <b>OLAP</b>                     | <i>On-Line Analytical Processing</i> . Proceso Analítico en línea.                                  |
| <b>OLTP</b>                     | <i>On-Line Transaction Processing</i> . Proceso transaccional en línea                              |
| <b>PC</b>                       | <i>Personal Computer</i> . Computador Personal.                                                     |
| <b>Planificación</b>            | Es un área de la empresa que analizan los indicadores.                                              |
| <b>Poblar</b>                   | Cargar la información al <i>Data Warehouse</i> .                                                    |
| <b>RDBMS</b>                    | <i>Relational Database Management System</i> . Sistema Gestor de Base de datos Relacional.          |
| <b>RDSI</b>                     | Red Digital de Servicios Integrados.                                                                |

|                           |                                                                                           |
|---------------------------|-------------------------------------------------------------------------------------------|
| <b>Rendimiento</b>        | Duración de utilización de un recurso dividido entre la duración total de observación.    |
| <b>Retrospección</b>      | Almacenamiento histórico, la historia se altera cuando algún valor de los objetos cambia. |
| <b>ROLAP</b>              | <i>Relational Online Analytical Processing</i> . Proceso Analítico en Línea Relacional.   |
| <b>Roll up</b>            | Consolidar o Comprimir información.                                                       |
| <b>RUP</b>                | <i>Rational Unified Process</i> . Proceso Unificado Racional.                             |
| <b>Segmento 6</b>         | Redondeo de los minutos en el consumo de las llamadas internacionales.                    |
| <b>Semántica</b>          | Determina el formato y el significado de un atributo.                                     |
| <b>Snowflake</b>          | Copo de nieve.                                                                            |
| <b>SQL</b>                | <i>Structure Query Language</i> . Lenguaje Estructural de Consultas.                      |
| <b>Sumarización</b>       | Nivel de resumen.                                                                         |
| <b>Tabla de Hecho</b>     | Tabla en el centro del esquema dimensional. Contiene los valores numéricos.               |
| <b>Tomas</b>              | Utilización de un recurso de la red telefónica.                                           |
| <b>Tráfico telefónico</b> | Duración de una llamada que puede ser expresada en una unidad de tiempo o en Erlangs.     |

|                        |                                                                                    |
|------------------------|------------------------------------------------------------------------------------|
| <b><i>Triggers</i></b> | Son secuencias de comandos que se ejecutan al producirse un evento.                |
| <b><i>UML</i></b>      | <i>Unified Modeling Language</i> . Lenguaje de Modelado Unificado.                 |
| <b><i>URA</i></b>      | Unidad de conexión del abonado es parte de una central telefónica.                 |
| <b><i>VIM</i></b>      | <i>Validation Integration Mapping</i> . Validación, integración y Mapeo.           |
| <b><i>Xlcubed</i></b>  | Es un complemento de Excel para realizar los reportes utilizando tablas dinámicas. |

## BIBLIOGRAFÍA

ALCATEL. “Gestión de las Observaciones de Carga y Tráfico”. 2002.

BITAM. Acerca de *Business Intelligence* (Inteligencia de Negocios).  
<http://www.bitam.com/spanish/TecAnalisis.html> 2004, [consulta 10 de abril de 2006].

DEPARTAMENTO DE SISTEMAS INFORMÁTICOS Y COMPUTACIÓN  
UNIVERSIDAD POLITÉCNICA DE VALENCIA. Almacenes de Datos.  
Hernández Orayo José.  
[http://www.dsic.upv.es/~jorayo/cursoDWDM/dwdm\\_11.pdf](http://www.dsic.upv.es/~jorayo/cursoDWDM/dwdm_11.pdf), [consulta 24 de abril de 2006].

IBM. Sistemas *OLAP*. Ibarra María de los Ángeles.  
[http://exa.unne.edu.ar/depar/areas/informatica/sistemasoperativos\\_olap.pdf](http://exa.unne.edu.ar/depar/areas/informatica/sistemasoperativos_olap.pdf) . 2005,  
[consulta 15 de abril de 2006].

IBM. Variante en el Tiempo. Ibarra María de los Ángeles.  
[http://exa.unne.edu.ar/depar/areas/informatica/sistemasoperativos\\_olap.pdf](http://exa.unne.edu.ar/depar/areas/informatica/sistemasoperativos_olap.pdf) . 2005,  
[consulta 1 de abril de 2006].

JACOBSON, Ivar; BOOCH, Grady; RUMBAUGH, James. “El Proceso Unificado de desarrollo de Software”. *Pearson Education*. S.A. España, 2000.

MICROSOFT CORPORATION, “*Database creation, warehousing, and optimization*” (Creación de una base de datos, procedimiento y optimización), *Microsoft Press, United States of America*, 2001.

MICROSOFT CORPORATION. *Dimensions Hierarchies* (Jerarquías de dimensión).  
[http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions\\_42ib.htm](http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions_42ib.htm),  
1988, 2000, [consulta 20 de abril de 2006].

MICROSOFT CORPORATION. *Dimensions* (Dimensiones).  
[http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions\\_12pf.htm](http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions_12pf.htm),  
1988, 2000, [consulta 20 de abril de 2006]

MICROSOFT CORPORATION. *Optimizing Cube Performance Using Microsoft SQL Server 2000 Analysis Services* (Optimizar la ejecución de un cubo usando *Microsoft Sql Server 2000*). <http://www.msdn.microsoft.com/library>. *Analysis Services*. 2001, [consulta 15 de abril de 2006].

MICROSOFT CORPORATION. *Introduction To Dimensions* (Introducción a dimensiones).  
[http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions\\_30mr.htm](http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agdimensions_30mr.htm),  
1988, 2000, [consulta 21 de abril de 2006].

*MICROSOFT CORPORATION. Levels and members* (Niveles y miembros).  
[http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/aglevels\\_7w4z.htm](http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/aglevels_7w4z.htm), 1988,  
2000, [consulta 21 de abril de 2006].

*MICROSOFT CORPORATION. Measures* (Medidas).  
[http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agmeasures\\_9fsj.htm](http://www.doc.ddart.net/mssql/sql2000/html/olapdmat/agmeasures_9fsj.htm), 1988,  
2000, [consulta 20 de abril de 2006].

*MICROSOFT. Microsoft Solutions Framework* (Marco de Trabajo de Soluciones de Microsoft). <http://www.microsoft.com/MSF> Junio 2003, [consulta 15 de octubre de 2005].

*ORACLE CORPORATION. Data Warehousing Guide* (Guía para el Proceso de una bodega de datos).  
<http://www.lc.leidenuniv.nl/awcouse/oracle/server.920/a96520/doc.htm> Marzo 2002,  
[consulta noviembre de 2005].

*REALITECH. Data Warehousing* (Proceso Bodega de Datos).  
<http://www.sqlmax.com/dataw1.asp> Junio 2001, [consulta 15 de abril de 2006].

*TODDMAN, Chris. "Designing a Data Warehouse"* (Diseño de una Bodega de datos). Hall PTR, *New Jersey*. Octubre 2001.