

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS

FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

E.A.P DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

**Solución Integral para explotar eficientemente la
Información de los contactos con los clientes utilizando
Datamart en Telefónica del Perú**

TESINA

para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Alejandro Javier Gamarra Ramírez

ASESOR

David Armando Espinoza Robles

Lima – Perú

2010

FICHA CATALOGRÁFICA

GAMARRA RAMÍREZ, ALEJANDRO JAVIER

SOLUCIÓN INTEGRAL PARA EXPLOTAR EFICIENTEMENTE LA INFORMACIÓN DE LOS CONTACTOS CON LOS CLIENTES UTILIZANDO DATAMART EN TELEFÓNICA DEL PERÚ

Sistemas, Informática y Sociedad
(Lima, Perú 2010)

Tesina, Facultad de Ingeniería de Sistemas, Pregrado, Universidad Nacional Mayor De San Marcos
Formato 21 x 29.7 cm

Páginas 139

DEDICATORIA:

La presente tesis está dedicada al esfuerzo de mis padres por permitirme una educación, a mi abuela Blanca por ser una fuente de inspiración y fortaleza todo este tiempo y para ti Evelyn Peña que me brindaste tu apoyo y confianza durante tantos años. Todos mis logros están dedicados a ustedes, no hubiese podido lograrlo sin su ayuda.

Gracias a todos.

AGRADECIMIENTOS

Agradezco a toda mi familia, mi papá, mi mamá, mis hermanos, mis tíos que en su momento me han dado consejos y sugerencias para ser una mejor persona y espero nunca defraudarlos.

A mis amigos y amigas Yaneet, Rosa, Ruth, Verónica, Consuelo, Lorena, Steve, Jaime, Diana, Yvett, con los que he pasado unos inolvidables momentos y que me han apoyado cuando más los necesitaba.

A los ingenieros del proyecto que me han inculcado los valores de la responsabilidad, el respeto, entre otros, además de sus conocimientos en cada una de las materias.

A mi asesor, Ingeniero David Espinoza Robles que con sus consejos y sugerencias me han apoyado para culminar con éxito este proyecto.

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
FACULTAD DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

**SOLUCIÓN INTEGRAL PARA EXPLOTAR EFICIENTEMENTE LA INFORMACIÓN
DE LOS CONTACTOS CON LOS CLIENTES UTILIZANDO DATAMART EN
TELEFÓNICA DEL PERÚ**

GAMARRA RAMÍREZ, ALEJANDRO JAVIER

ESPINOZA ROBLES, DAVID ARMANDO

Título: Tesina para optar el Título Profesional de Ingeniero de Sistemas

Fecha: Noviembre del 2010

RESUMEN

El presente trabajo muestra el proceso seguido para el desarrollo de un Datamart cuyo objetivo será facilitar la información de los contactos con el cliente de Telefónica del Perú a todos los usuarios internos de la empresa y proporcionar las herramientas necesarias que ayuden a la correcta explotación y uso eficiente de la misma. La presencia de este Datamart contribuirá a comprender, aún mejor, cuáles son las necesidades reales del cliente y contribuir a la satisfacción de las mismas.

La tesis realizada pone de manifiesto la gran utilidad que tiene la metodología propuesta por Gill Harjinder y Rao Prakash, la cual es muy flexible y ayuda a realizar eficazmente este tipo de proyectos. Para la implementación del Datamart se seleccionó SQL Server como motor de base de datos, Analysis Services para la construcción de cubos multidimensionales, Reporting Services para la elaboración de reportes y Macromedia Dreamweaver para la creación del Sitio Web de consultas online de los contactos de los clientes.

Palabras Claves: Datamart, contactos, metodología, SQL, Sitio Web, consultas online

MAJOR NATIONAL UNIVERSITY OF SAN MARCOS
FACULTY OF SYSTEMS ENGINEERING
ESCUELA ACADÉMICO PROFESIONAL DE INGENIERÍA DE SISTEMAS E INFORMÁTICA

**INTEGRAL SOLUTION TO EXPLOIT EFFICIENTLY THE CUSTOMERS
CONTACTS INFORMATION BY USING DATAMART IN TELEFÓNICA DEL PERÚ**

GAMARRA RAMÍREZ, ALEJANDRO JAVIER
ESPINOZA ROBLES, DAVID ARMANDO

Título: Tesina para optar el Título Profesional de Ingeniero de Sistemas

Fecha: Noviembre del 2010

ABSTRACT

The present work shows the process followed for the development of a Datamart whose objective will be to provide information of the customer's contacts with Telefónica del Perú to all company users and providing the necessary tools to assist the correct operation and efficient use of the same. The presence of this Datamart will contribute to understand, even better, which are the real needs of the client and to contribute to the satisfaction of the same.

The thesis made shows the great utility of the methodology proposed by Gill Harjinder and Rao Prakash, which is very flexible and helps to realize this kind of projects effectively. For Datamart's implementation, SQL Server was selected as the database engine, Analysis Services as multidimensional cubes builder, Reporting Services for reporting, and Macromedia Dreamweaver to create the Web Site for online consultations of customer contacts.

Key words: Datamart, contacts, methodology, SQL, Web Site, online consultations

ÍNDICE DE CONTENIDO

DEDICATORIA.....	ii
AGRADECIMIENTOS.....	iii
RESUMEN.....	iv
ABSTRACT.....	v
ÍNDICE DE CONTENIDO.....	vi
ÍNDICE DE FIGURAS.....	xi
ÍNDICE DE TABLAS.....	xiii
INTRODUCCIÓN.....	xv

Capítulo 1. PLANTEAMIENTO METODOLÓGICO

1.1 ANTECEDENTES.....	2
1.2 DEFINICIÓN DEL PROBLEMA.....	7
1.3 OBJETIVOS.....	8
1.4 JUSTIFICACIÓN.....	9
1.5 HIPÓTESIS.....	10
1.6 UNIVERSO Y MUESTRA.....	13
1.7 PROPUESTA.....	14
1.8 ORGANIZACIÓN DE LA TESINA.....	15

Capítulo 2. MARCO TEÓRICO

2.1 DATAMART.....	17
2.1.1 Conceptos Asociados.....	17
2.1.1.1 Datawarehouse.....	17
2.1.1.2 CRM (Customer RelationShip Management)	20
2.1.1.3 Inteligencia de negocios.....	21
2.1.2 Definición de Datamart.....	22
2.1.3 Características del Datamart.....	22
2.1.4 Clasificación del Datamart.....	23
2.1.4.1 Datamart dependiente.....	23
2.1.4.2 Datamart independiente.....	24

2.1.4.3	Datamart híbrido.....	24
2.1.5	Ventajas del Datamart.....	24
2.1.6	Desventajas del Datamart.....	25
2.1.7	Diferencias entre un OLTP y un Datamart.....	26
2.1.8	Diferencias entre un Datawarehouse y Datamart.....	27
2.2	MODELO MULTIDIMENSIONAL.....	28
2.2.1	Definiciones.....	28
2.2.1.1	Dimensión.....	28
2.2.1.2	Tabla de hechos.....	29
2.2.2	Modelo Copo de Nieve (Snowflake)	29
2.3	PROCESOS ETL.....	30
2.3.1	Definición.....	30
2.3.2	Etapas de los procesos ETL.....	31
2.3.2.1	Extracción.....	32
2.3.2.2	Transformación.....	33
2.3.2.3	Carga.....	35
2.4	DATAMINING.....	35
2.4.1	Definición.....	35
2.4.2	Ventajas del Datamining.....	36
2.4.3	Técnicas del Datamining.....	37
2.4.3.1	Análisis estadístico.....	37
2.4.3.2	Árboles de decisión.....	38
2.4.3.3	Redes neuronales.....	39
2.4.3.4	Inducción de reglas.....	39
2.4.3.5	Modelos de asociación.....	39
2.4.3.6	Lógica difusa.....	40
2.4.3.7	Series temporales.....	40
2.4.3.8	Algoritmos genéticos.....	40
2.4.4	Diferencias entre consultas, Datamining y sistemas OLAP.....	40
2.5	SISTEMAS DE SOPORTE A LAS DECISIONES.....	41
2.5.1	Definición.....	41
2.5.2	Tipos de Sistemas de Soporte a las Decisiones.....	42
2.5.3	Características de los Sistemas de Soporte a las Decisiones.....	42
2.6	REDES.....	43

2.6.1	Redes de computadoras.....	43
2.6.2	Componentes de las redes de computadoras.....	43
2.6.2.1	Capa de aplicación.....	44
2.6.2.2	Capa de transporte.....	44
2.6.2.3	Capa de Internet.....	44
2.6.2.4	Capa de interfaz de red.....	44
2.7	SOLUCIONES INTEGRALES.....	44

Capítulo 3. ESTADO DEL ARTE METODOLÓGICO

3.1	METODOLOGÍAS PARA EL DESARROLLO DE UN DATAMART.....	46
3.1.1	Metodología propuesta por Gill Harjinder y Rao Prakash.....	46
3.1.1.1	Planeación.....	46
3.1.1.2	Requerimientos.....	52
3.1.1.3	Análisis.....	56
3.1.1.4	Diseño.....	57
3.1.1.5	Construcción.....	58
3.1.1.6	Despliegue.....	58
3.1.1.7	Expansión.....	59
3.1.2	Metodología propuesta por Ralph Kimball y Margy Ross.....	59
3.1.2.1	Planificación.....	61
3.1.2.2	La administración de la planificación.....	62
3.1.2.3	Definición de requerimientos del negocio.....	62
3.1.2.4	Diseño de la Arquitectura Técnica.....	63
3.1.2.5	Selección e instalación del producto.....	64
3.1.2.6	Diseño Dimensional.....	64
3.1.2.7	Diseño físico.....	65
3.1.2.8	Diseño y Desarrollo del Almacén de Datos.....	66
3.1.2.9	Despliegue.....	66
3.1.2.10	Mantenimiento Incremental.....	66
3.2	SELECCIÓN DE LA METODOLOGÍA A APLICARSE.....	67
3.3	HERRAMIENTAS PARA EL DESARROLLO DE UN DATAMART.....	68
3.3.1	Herramientas de consulta y reportes de usuario final.....	69
3.3.2	Herramientas de Datawarehouse.....	74

3.4 SELECCIÓN DE LAS HERRAMIENTAS PARA EL DESARROLLO DE UN DATAMART.....	78
3.5 HERRAMIENTAS PARA EL DESARROLLO DE UN SITIO WEB.	80
3.6 SELECCIÓN DE LAS HERRAMIENTAS PARA EL DESARROLLO DE UN SITIO WEB.....	85

Capítulo 4. IMPLEMENTACIÓN

4.1 PLANEACIÓN.....	88
4.1.1 Selección de la estrategia de Implementación.....	88
4.1.2 Selección de la metodología de desarrollo.....	88
4.1.3 Desarrollo de los objetivos empresariales.....	88
4.1.4 Selección del ámbito de implementación.....	89
4.1.5 Selección de enfoque arquitectónico.....	89
4.1.6 Desarrollo de un programa y presupuesto para planes del proyecto... ..	90
4.1.7 Desarrollo de escenarios de uso empresarial.....	91
4.1.8 Recopilación de los metadatos.....	92
4.2 REQUERIMIENTOS.....	92
4.2.1 Requerimientos del propietario.....	92
4.2.2 Requerimientos del arquitecto.....	93
4.2.3 Requerimientos del desarrollador.....	95
4.2.4 Requerimientos de los usuarios finales.....	96
4.3 ANÁLISIS.....	97
4.3.1 Visión de arriba hacia abajo.....	97
4.3.2 Visión de la fuente de datos.....	97
4.3.3 Visión del Datamart.....	98
4.3.4 Visión de consulta empresarial.....	99
4.4 DISEÑO.....	99
4.4.1 Diseño de la arquitectura de datos.....	99
4.4.2 Diseño de la arquitectura de la aplicación.....	101
4.5 CONSTRUCCIÓN.....	103
4.5.1 Modelo de Base de Datos.....	105
4.5.2 Paquetes de Transformación.....	108
4.5.3 Construcción del Cubo del Datamart.....	123

4.5.4 Construcción de Web de Contactos del Simulador	128
4.6 DESPLIEGUE	129
CONCLUSIONES.....	135
RECOMENDACIONES.....	137
BIBLIOGRAFÍA.....	138

ÍNDICE DE FIGURAS

Capítulo 2. MARCO TEÓRICO

FIGURA II.1 DATAMART DEPENDIENTE.....	23
FIGURA II.2 DATAMART INDEPENDIENTE.....	23
FIGURA II.3 DATAMART HÍBRIDO.....	23
FIGURA II.4 EJEMPLO DE CUBO DE DIMENSIONES.....	29
FIGURA II.5 EJEMPLO DE MODELO SNOWFLAKE O COPO DE NIEVE.....	30
FIGURA II.6 EJEMPLO DE UN FLUJO DE TRANSFORMACIONES.....	33
FIGURA II.7 CLASES DE USUARIO DENTRO DE LA ORGANIZACIÓN.....	37
FIGURA II.8 CLASES DE HERRAMIENTAS USADAS POR LOS USUARIOS DE LA ORGANIZACIÓN.....	37

Capítulo 3. ESTADO DEL ARTE METODOLÓGICO

FIGURA III.1 PASOS PARA IMPLEMENTAR LA FASE DE PLANEACIÓN DEL SISTEMA DE DATAWAREHOUSE.....	47
FIGURA III.2 MODELO DE DESARROLLO ESPIRAL.....	49
FIGURA III.3 PASOS PARA IMPLEMENTAR LA FASE DE REQUERIMIENTOS DEL SISTEMA DATAWAREHOUSE.....	53
FIGURA III.4 DIAGRAMA DEL CICLO DE VIDA DEL DATAMART.....	60
FIGURA III.5 MATRIZ DE PRIORIZACIÓN DE PROCESOS.....	63
FIGURA III.6 EJEMPLO DE MATRIZ DE PROCESOS Y DIMENSIONES.....	65

Capítulo 4. IMPLEMENTACIÓN

FIGURA IV.1 DIAGRAMA DE LA ARQUITECTURA DE APLICACIÓN (FASE 1).....	102
FIGURA IV.2 DIAGRAMA DE LA ARQUITECTURA DE APLICACIÓN (FASE 2).....	103
FIGURA IV.3 MODELO DE BASE DE DATOS DEL SIMULADOR DEL PROYECTO (PARTE I).....	106
FIGURA IV.4 MODELO DE BASE DE DATOS DEL SIMULADOR DEL PROYECTO (PARTE II).....	107
FIGURA IV.5 DTS DE CARGA DE LA FUENTE DEL 104.....	109
FIGURA IV.6 DTS DE CARGA DE LA FUENTE DE AVERÍAS.....	110
FIGURA IV.7 DTS DE CARGA DE LA FUENTE DE CABLE MÁGICO.....	111
FIGURA IV.8 DTS DE CARGA DE LA FUENTE DEL CANAL ONLINE.....	112
FIGURA IV.9 DTS DE CARGA DE LA FUENTE DEL CROSS SELLING.....	113
FIGURA IV.10 DTS DE CARGA DE LA FUENTE DEL MULTICENTRO.....	114
FIGURA IV.11 DTS DE CARGA DE LA FUENTE DEL CANAL OUT.....	115

FIGURA IV.12 DTS DE CARGA DE LA FUENTE DE RECLAMOS DE CABLE MÁGICO.....	116
FIGURA IV.13 DTS DE CARGA DE LA FUENTE DE RECLAMOS TDP.....	117
FIGURA IV.14 DTS DE CARGA DE LA FUENTE DEL CANAL SPEEDY.....	118
FIGURA IV.15 DTS DE CARGA DE LA FUENTE DE TELEFÓNICA ONLINE.....	119
FIGURA IV.16 DTS DE CARGA DE LA FUENTE DEL LABORATORIO.....	120
FIGURA IV.17 DTS DE INTEGRACIÓN DE LOS CONTACTOS CON EL CLIENTE DE TELEFÓNICA.....	122
FIGURA IV.18 ASISTENTE PARA GENERAR CUBOS DEL ANALYSIS MANAGER.....	123
FIGURA IV.19 SELECCIÓN DE UN NUEVO ORIGEN DE DATOS.....	124
FIGURA IV.20 SELECCIÓN DEL PROVEEDOR DEL VÍNCULO DE DATOS.....	124
FIGURA IV.21 SELECCIÓN DE LA BASE DE DATOS DEL DATAMART.....	125
FIGURA IV.22 DEFINICIÓN DE LA TABLA DE HECHOS Y SUS MEDIDAS.....	125
FIGURA IV.23 DEFINICIÓN DE LAS DIMENSIONES Y JERARQUÍAS DEL CUBO.....	125
FIGURA IV.24 CREACIÓN DE LA DIMENSIÓN TIEMPO.....	126
FIGURA IV.25 VISTA DEL CUBO DEL SIMULADOR DEL DATAMART FINALIZADO.....	127
FIGURA IV.26 WEB DE CONTACTOS CON CONEXIÓN AL DATAMART DE CONTACTOS (PAGINA DE BÚSQUEDA).....	128
FIGURA IV.27 WEB DE CONTACTOS CON CONEXIÓN AL DATAMART DE CONTACTOS (PAGINA DE RESULTADOS).....	129
FIGURA IV.28 DASHBOARD – REPORTE POR MEDIO DE CONTACTOS EN EL ÚLTIMO MES.....	130
FIGURA IV.29 DASHBOARD – REPORTE DE LA EVOLUCIÓN MENSUAL DE LOS RECLAMOS EN EL 2010.....	131
FIGURA IV.30 DASHBOARD – REPORTE DE SATISFACCIÓN DEL CLIENTE POR MEDIO DE CONTACTO.....	131

FIGURA IV.31 DASHBOARD – EVOLUCIÓN MENSUAL DE LOS RECLAMOS EN EL 2010 POR EL CANAL OUT.....	132
---	-----

ÍNDICE DE TABLAS

Capítulo 2. MARCO TEÓRICO

TABLA II.1 DIFERENCIA ENTRE OLTP Y DATAMART.....	27
TABLA II.2 DIFERENCIA ENTRE DATAWAREHOUSE Y DATAMART.....	28
TABLA II.3 DIFERENCIA ENTRE CONSULTAS, OLAP Y DATAMINING.....	41

Capítulo 3. ESTADO DEL ARTE METODOLÓGICO

TABLA III.1 PROS Y CONTRAS DEL ENFOQUE DE ABAJO HACIA ARRIBA.....	48
TABLA III.2 CUADRO COMPARATIVO DE METODOLOGÍAS DE DESARROLLO DE DATAWAREHOUSE.....	67
TABLA III.3 JUSTIFICACIÓN DE PUNTUACIONES ASIGNADAS A LOS PARÁMETROS DE SELECCIÓN.....	68
TABLA III.4 SELECCIÓN DE LA METODOLOGÍA.....	68
TABLA III.5 CUADRO COMPARATIVO DE HERRAMIENTAS PARA REALIZACIÓN DE REPORTE DISPONIBLES.....	73
TABLA III.6 JUSTIFICACIÓN DE PUNTUACIONES ASIGNADAS A LOS PARÁMETROS DE SELECCIÓN.....	73
TABLA III.7 SELECCIÓN DE HERRAMIENTA PARA ELABORACIÓN DE REPORTE.....	73
TABLA III.8 CUADRO COMPARATIVO DE HERRAMIENTAS DATAWAREHOUSE DISPONIBLES EN EL MERCADO.....	79

TABLA III.9 JUSTIFICACIÓN DE PUNTUACIONES ASIGNADAS A LOS PARÁMETROS DE SELECCIÓN.....	79
TABLA III.10 SELECCIÓN DE HERRAMIENTA DE DATAWAREHOUSE.....	79
TABLA III.11 CUADRO COMPARATIVO DE HERRAMIENTAS PARA EL DESARROLLO DE UN SITIO WEB.....	85
TABLA III.12 JUSTIFICACIÓN DE PUNTUACIONES ASIGNADAS A LOS PARÁMETROS DE SELECCIÓN.....	86
TABLA III.13 SELECCIÓN DE HERRAMIENTA PARA EL DESARROLLO DE UN SITIO WEB.....	86

Capítulo 4. IMPLEMENTACIÓN

TABLA IV.1 TIEMPO ESTIMADO DEL PROYECTO.....	90
TABLA IV.2 PRESUPUESTO DEL PROYECTO.....	91
TABLA IV.3 DISEÑO DE LA TABLA DE DATOS DEL CLIENTE.....	100
TABLA IV.4 DISEÑO DE LA TABLA DE CAMPAÑAS ASIGNADAS.....	100
TABLA IV.5 DISEÑO DE LA TABLA DE CONTACTOS CON EL CLIENTE.....	101

INTRODUCCIÓN

Actualmente la distribución de la información en el negocio, es vital para la toma de decisiones acertadas y oportunas bajo el manejo de datos confiables y actualizados. Esta distribución de información, se hace aún más fiable y fácil de manejar, al contar con componentes y procesos que agilicen esta operación. Uno de los conceptos que permite centrarse en un solo segmento del negocio y determinar lo antes expuesto, es el Datamart.

El objetivo fundamental del presente trabajo especial de grado se centra en la construcción, para el Segmento Residencial de Telefónica del Perú, de un Datamart con datos de los clientes residenciales de las filiales de la empresa.

El estudio y desarrollo de dicho Datamart se estructuró en 5 capítulos. *El Capítulo I. Planteamiento Metodológico*, presenta un esbozo del problema y sus alcances, de tal manera que permita al lector conocer la situación y las necesidades del negocio así como los objetivos a alcanzar con el desarrollo del mismo. En el *Capítulo II, referido al Marco Teórico* se describen las bases teóricas que sustentan el

desarrollo del Datamart. *El Capítulo III. Estado del Arte Metodológico*, define las acciones a tomar dentro del despliegue del Datamart. *El Capítulo IV, referido a la Implementación*, en este capítulo se va a estructurar las acciones en diseños que permitan el desarrollo del mismo. *El Capítulo V. Conclusiones y Recomendaciones*, comprenderán las conclusiones del desarrollo de la tesina hasta esta etapa.

Con esta estructura se espera cubrir los aspectos importantes del despliegue del proyecto, así como también el fácil entendimiento de la secuencia del mismo.

CAPÍTULO I

PLANTEAMIENTO METODOLÓGICO

1.1. ANTECEDENTES

A través del procesamiento de información, una compañía crea valor, en especial si se trata de una empresa que ofrece servicios. Por lo tanto, en este caso, la información tiene un valor aún mayor porque ayuda a alcanzar los objetivos de la compañía.

Los Sistemas de Información (SI) y las Tecnologías de Información (TI) han cambiado la forma en que operan las organizaciones actuales. A través de su uso se logran importantes mejoras, pues automatizan los procesos operativos, suministran una plataforma de información necesaria para la toma de decisiones y, lo más importante, su implantación logra ventajas competitivas o reducir la ventaja de los rivales¹.

Con frecuencia, el uso de la tecnología de información para la globalización da como resultado el desarrollo de sistemas de información que ayudan a una empresa a darle ventaja competitiva en el mercado, utilizándolos para desarrollar productos, servicios, procesos y capacidades que dan a una empresa una ventaja estratégica sobre las fuerzas competitivas que enfrenta una empresa. Estrategias de diferenciación, como desarrollar maneras de utilizar la tecnología de información para diferenciar productos o servicios de una empresa, de los de sus competidores, de manera que los clientes perciban los productos o servicios como poseedores de atributos o beneficios únicos.

¹ Impacto de Tecnologías Informáticas I – (En Línea)
<http://www.deltaasesores.com>

Por ejemplo, suministrar servicios rápidos y completos de soporte al cliente por medio de un sitio Web en Internet, o utilizar sistemas de marketing como objetivo para ofrecer a clientes individuales los productos y servicios que le atraen.

La evolución de los Sistemas, y la creciente competencia en el mercado, hacía que cada vez más fuera necesario eliminar trabas a los usuarios en sus consultas de la información procedente del negocio, capturada por los Sistemas Operacionales normales de la Compañía, pero que antes sólo se podía ver de forma agregada, por ejemplo: Ventas totales en el mes. En una palabra, había que estructurar el mensaje comercial para convencer a las empresas de que les sería muy útil gastarse una fortuna en un nuevo gran sistema, probablemente incompatible con todos los que tenía ya instalados.

El resultado fue el concepto de Datawarehouse, allá por los años 93-94 del siglo pasado. Se habían hecho ya ciertos pinitos en los años 80, sobre todo los de una pequeña compañía californiana llamada Teradata, pionera absoluta en el concepto de máquina paralela dedicada exclusivamente a la resolución de consultas.

Mediados de los noventa. Tenemos Sistemas UNIX con capacidad multiprocesador y diversas arquitecturas para proceso paralelo. Existen Bases de Datos Relacionales con capacidad de procesar Series en paralelo. Las redes y los PC's ya tienen capacidad suficiente como para mover una cierta cantidad de información por las redes. Sabemos dónde están los datos en nuestros Sistema Operacionales, sabemos diseñar y programar el software, tenemos experiencia en el uso de las Bases de Datos Relacionales.

Y hubo algunas empresas, no demasiadas todavía, que comenzaron sus proyectos de Datawarehouse alegremente. Hicieron costosos benchmarking entre Sistemas, pruebas exhaustivas, estudios comparativos rigurosísimos, compraron un gran servidor de varios procesadores y una base de datos (los que mejor salieron en las interminables pruebas... o los que decidieron a dedo los directores, que de todo hubo), y comenzaron el proyecto para crear, por fin, el Datawarehouse.

Los sistemas de Data Warehousing son el centro de la arquitectura de los Sistemas de Información de los 90's. Han surgido como respuesta a la problemática de extraer información sintética a partir de datos atómicos almacenados en bases de datos de producción. Uno de los objetivos principales de este tipo de sistemas es servir como base de información para la toma de decisiones. Los beneficios obtenidos por la utilización de este tipo de sistemas se basan en el acceso interactivo e inmediato a información estratégica de un área de negocios. Este acercamiento de la información al usuario final permite una toma de decisiones rápida y basada en datos objetivos obtenidos a partir de las bases de datos (eventualmente heterogéneas) de la empresa. Estos beneficios aumentan cuanto más importantes son las decisiones a tomar y cuanto más crítico es el factor tiempo.

Hoy en día las empresas cuentan en su mayoría con la automatización de sus procesos, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea. En esta información descansa el "Know How" de la empresa, constituyendo un recurso corporativo primario y parte importante de su patrimonio.

El nivel competitivo alcanzado en las empresas les ha exigido desarrollar nuevas estrategias de gestión. En el pasado, las organizaciones fueron típicamente estructuradas en forma piramidal con información generada en su base fluyendo hacia lo alto; y era en el estrato de la pirámide más alto donde se tomaban decisiones a partir de la información proporcionada por la base, con un bajo aprovechamiento del potencial de esta información. Estas empresas, han reestructurado y eliminado estratos de estas pirámides y han autorizado a los usuarios de todos los niveles a tomar mayores decisiones y responsabilidades. Sin embargo, sin información sólida para influenciar y apoyar las decisiones, la autorización no tiene sentido.

Esta necesidad de obtener información para una amplia variedad de individuos es la principal razón de negocios que conduce al concepto de Datawarehouse. El énfasis no está sólo en llevar la información hacia lo alto sino que a través de la organización, para que todos los empleados que la necesiten la tengan a su disposición.

Data Warehousing y Data Mining son dos procesos, entre otros, que posibilitarán la Gestión del Conocimiento en una compañía. La aparición, relativamente reciente de estas tecnologías, que han surgido con una inusitada fuerza en el mercado de IT, han sorprendido a muchos profesionales del sector, y han generado una tremenda confusión en el mercado.

El concepto Datamart es una extensión natural del Datawarehouse, y está enfocado a un departamento o área específica, como por ejemplo los departamentos de

Finanzas o Marketing. Permitiendo así un mejor control de la información que se está abarcando.

En Telefónica del Perú, empresa dedicada a prestar servicios de telecomunicaciones, la información es un factor estratégico, no sólo para la interrelación entre áreas sino también en el contacto con los clientes. Al igual que muchas otras empresas en el mundo tiene sus propios sistemas de información que tienen por objetivo darle al cliente, tanto interno como externo, la mejor respuesta a sus necesidades. Sin embargo, en los últimos meses se ha detectado cierta ineficiencia en los sistemas que manejan los diferentes puntos de contacto con los clientes. Este esencialmente se debe a la falta de un sistema que integre la información recabada por las diferentes agencias.

Las empresas de telecomunicaciones en todo el mundo están enfrentando una reducción de su participación en el mercado, la desregulación y los veloces avances tecnológicos. La competitividad y supervivencia final de las compañías del sector depende de su habilidad para tomar decisiones en forma rápida e informada.

La industria de las telecomunicaciones está ingresando a una nueva y desafiante era. El ritmo de los cambios en el sector de las telecomunicaciones es tan dramático que muchas veces se lo llama la “nueva frontera”. Desregulación, globalización, nueva tecnología y mayor competencia son las tendencias que están creando una atmósfera de incertidumbre. Pese a lo cual el potencial de incrementar la participación en el mercado y las ganancias nunca ha sido mayor.

Pero, la inevitable competencia que resulta de estas influencias externas está forzando a las compañías de telecomunicaciones a examinar de cerca sus iniciativas de negocios. En otras palabras, las empresas están re-examinando su enfoque del mercado, sus gastos y sus estructuras.

Para establecer un liderazgo como empresa de telecomunicaciones, se debe tener un detallado conocimiento (no sólo datos) de los clientes y competidores. Las decisiones de negocios basadas en este conocimiento permitirán crear estrategias efectivas para proveer una gama de productos y servicios que satisfagan las necesidades de todos los clientes. Teniendo a que todas las necesidades de telecomunicaciones sean provistas por un solo proveedor, la convergencia exitosa de todos estos servicios puede ser la clave para asegurar una saludable participación en el mercado.

1.2. DEFINICIÓN DEL PROBLEMA

La carencia de integración de la información en Telefónica representa un problema para los jefes, analistas, supervisores, operadores e incluso para los propios clientes (estos últimos en forma indirecta). No existe un ente integrador que reúna los contactos con el cliente y los publique en un único lugar al cual puedan acceder los usuarios internos y utilizar esta información para análisis, seguimientos, búsquedas, cruce de información, etc.

El problema se da cuando un usuario interno tiene que realizar algún cruce de información con estas fuentes, ya que debe buscar a las personas responsables de dicha información, solicitarla, esperar la confirmación, acceder a ésta y exportarla a

algún motor de base de datos, analizarla, entenderla, depurarla, para finalmente poder realizar el cruce que se necesitaba inicialmente. Ahora imagínese tener que hacer el cruce con 2 o más fuentes; es realizar el mismo procedimiento multiplicado por el número de fuentes que necesites.

Un simple cruce de información con fuentes de múltiples contactos con el cliente toma para un usuario promedio alrededor de 1 semana, sólo para conseguir integrar esta información. Ahora imagínese que esta simple tarea se deba realizar periódicamente, se convertiría prácticamente en tu función diaria; función que abarcaría todo el tiempo del analista en su día de trabajo.

El enunciado del problema es: la complejidad actual en Telefónica del Perú para integrar eficientemente la información de los contactos con los clientes en un repositorio de datos ideal para su explotación.

1.3. OBJETIVOS

Objetivo General

- Desarrollar una solución integral para explotar eficientemente la información de los contactos con el cliente.

Objetivos Específicos

- Investigar y estudiar las diferentes técnicas de integración del conocimiento enfatizando en las empresas de comunicaciones.

- Consultar los adelantos e innovaciones en el tema que se han hecho hasta el momento al interior de entidades u organizaciones relacionadas con empresas de servicio al cliente.
- Establecer las principales características y diferencias entre algunas de las diversas herramientas que han sido diseñadas en los últimos años para solucionar el problema de la integración de la información.
- Identificar la información necesaria que el usuario de Telefónica necesita conocer y tener disponible para la óptima realización de su trabajo.
- Definir las herramientas que se desarrollarán o adoptarán en la empresa para solucionar el problema de la falta de integración de la información.

1.4. JUSTIFICACIÓN

Dentro de las empresas de servicio al cliente, la información se ha posicionado como uno de los principales recursos que poseen las empresas actualmente. La información no es sólo un subproducto de la conducción empresarial, sino que a la vez alimenta a los negocios y puede ser uno de los tantos factores críticos para la determinación del éxito o fracaso de éstas.

Si deseamos maximizar la utilidad que posee nuestra información, ésta se debe manejar de forma correcta y eficiente, tal y cómo se manejan los demás recursos existentes. Los administradores deben comprender de manera general que hay costos asociados con la producción, distribución, seguridad, almacenamiento y recuperación de toda la información que es manejada en la organización. Aunque la

información se encuentra a nuestro alrededor, debemos saber que ésta no es gratis, y su uso es estrictamente estratégico para posicionar de forma ventajosa la empresa dentro de un negocio.

Por lo dicho anteriormente, tener la información adecuada disponible, normalizada y actualizada sería uno de los pilares para las estrategias de marketing y comunicación con las que debería de contar Telefónica al momento de pensar en cómo llegar al cliente.

Después de todo, si el objetivo de la empresa es servir al cliente, lo primero que debería tener en cuenta es escuchar al cliente, y para esto, es que es necesario un sistema de información que pueda ayudar a la empresa a escucharlos, entenderlos para finalmente ayudarlos a elegir el producto que necesitan.

1.5. HIPÓTESIS

A continuación se plantearán las hipótesis que se pueden deducir de la tesis expuesta, revisaremos los problemas y las variables e indicadores que les podrían dar solución.

Hipótesis General

La implementación de un Datamart de Contactos, resolverá la necesidad de los usuarios de Telefónica de contar con la información de los contactos con los clientes completa, validada y disponible cuando ellos lo requieran, asimismo, los análisis que se realicen de ésta incrementará el conocimiento de las necesidades del cliente.

- **Hipótesis Específica 1**

La integración y disposición de los contactos con el cliente permitirá a los analistas realizar estudios más acertados para generar campañas que el cliente realmente desea recibir.

- **Hipótesis Específica 2**

La implementación de la web de consultas permitirá a los asesores comerciales tener las herramientas necesarias para resolver las dudas del cliente en el momento y generar así la confianza de éste último hacia la empresa.

Variables e Indicadores

A continuación se describirán los problemas, variables, indicadores e índices que nos ayudarán en el desarrollo de la tesina.

- **Variable Independiente:**

Implementación de un Datamart de Contactos

Indicadores: Total de fuentes de contacto con el cliente agregados al Datamart.

Índices: Catorce fuentes de contacto con el cliente en la actualidad.

- **Variable Dependiente:**

Conocimiento de las necesidades del Cliente.

Indicadores:

- Cantidad de rechazos en las campañas comerciales.
- Motivos de rechazo por tipo de contacto.
- Productos vendidos en el último mes por tipo de contacto.
- Evolución de las respuestas del cliente en el último trimestre.

Índices: Durante el mes de Agosto 2010 se registraron 207,398 rechazos aproximadamente.

Hipótesis Específica 1

La integración y disposición de los contactos con el cliente permitirá a los analistas realizar estudios más acertados para generar campañas que el cliente realmente desea recibir.

Variables e Indicadores

- **Variable Independiente:**

Integración y disposición de los contactos con el cliente.

Indicadores: Cantidad de usuarios que acceden a la información de los contactos.

Índices: Cero usuarios.

- **Variable Dependiente:**

Generar campañas que el cliente realmente desea recibir.

Indicadores: % de efectividad sobre contactos en Campañas Comerciales.

Índices: Durante el mes de Agosto 2010 se registró una efectividad sobre contacto del 14.52% aproximadamente en Campañas Comerciales.

Hipótesis Específica 2

La implementación de la web de consultas permitirá a los asesores comerciales tener las herramientas necesarias para resolver las dudas del cliente en el momento y generar así la confianza de éste último hacia la empresa.

Variables e Indicadores

- **Variable Independiente:**

Disponer las herramientas necesarias a los asesores comerciales.

Indicadores: Cantidad de herramientas a disposición de los asesores comerciales.

Índices: Hasta el mes de Agosto 2010 se registran 3 diferentes sistemas a disposición de los Multicentros.

- **Variable Dependiente:**

Mejorar la confianza del cliente hacia la empresa.

Indicadores: % de satisfacción del cliente en los Multicentros.

Índices: Durante el mes de Agosto 2010 se registró un índice de satisfacción del 8.5% en los clientes que visitaban los Multicentros.

1.6. UNIVERSO Y MUESTRA

Universo

El universo de clientes en Telefónica del Perú, que están aptos para generar Campañas Comerciales y sobre los cuales se extenderá el proyecto de tesina es de 1'623,519 clientes.

Muestra

Está conformada por la cuarta parte de clientes distribuidos proporcionalmente entre clientes con líneas libres, control y prepago, de lima y provincias, la cual asciende a una cantidad de 405,880 clientes.

1.7. PROPUESTA

Los analistas de la empresa, es decir, usuarios internos de la misma, utilizamos la información que les proporciona los sistemas actuales para realizar análisis y desarrollar estrategias de mercado.

Por los años de experiencia en el trabajo, se hace necesario un sistema de información que nos pueda facilitar realmente la información que necesitamos.

La propuesta que se propone en este documento es la realización de un sistema integral que beneficie tanto a los usuarios internos como a los mismos clientes de Telefónica, tal sistema integral lo vamos a dividir de la siguiente manera:

- **Datamart de Información con los contactos de los clientes de Telefónica.**

El cual se encontrará dentro del Datawarehouse de la empresa como un Universo más al cual puedan acceder todos los usuarios internos y les permita descargar una información actualizada, normalizada y confiable que les permita desarrollar sus análisis, consultas, seguimientos, en conclusión, la totalidad de sus tareas más dinámicamente.

- **Solución web de consultas a la información del cliente**

El cual será una solución web que estará a disposición de todos los asesores de Telefónica que tengan contactos con los clientes, esto es, que esta herramienta se encuentre instalada en los Multicentros, 104, Multimedia, Centros de Cobro, etc. El cual permitirá buscar simplemente ingresando algún identificador del cliente toda la información integrada de sus más recientes contactos con la empresa, así como más

información adicional que les permita desarrollar su gestión más eficientemente, y genere hacia el cliente una confianza para con la empresa.

1.8. ORGANIZACIÓN DE LA TESINA

El desarrollo de la tesina se estructuró en 5 capítulos. El Capítulo I. Planteamiento Metodológico, presenta un esbozo del problema y sus alcances, de tal manera que permita al lector conocer la situación y las necesidades del negocio así como los objetivos a alcanzar con el desarrollo del mismo. En el *Capítulo II, referido al Marco Teórico* se describen las bases teóricas que sustentan el desarrollo del Datamart. *El Capítulo III. Estado del Arte Metodológico*, define las acciones a tomar dentro del despliegue del Datamart. *El Capítulo IV, referido a la Implementación*, en este capítulo se va a estructurar las acciones en diseños que permitan el desarrollo del mismo. *El Capítulo V. Conclusiones y Recomendaciones*, comprenderán las conclusiones del desarrollo de la tesina hasta esta etapa. A estos 5 capítulos, se anexan los apéndices, que complementan la información manejada en el desarrollo del proyecto.

Con esta estructura se espera cubrir los aspectos importantes del despliegue del proyecto, así como también el fácil entendimiento de la secuencia del mismo.

CAPÍTULO II
MARCO TEÓRICO

2.1. DATAMART

2.1.1. Conceptos Asociados

2.1.1.1. Datawarehouse

Definición

De acuerdo a William H. Inmon, considerado el padre del concepto de Datawarehouse (almacén de datos) en los años ochenta, Datawarehouse es "un conjunto de datos integrados orientados a una materia, que varían con el tiempo y que no son transitorios, los cuales soportan el proceso de toma de decisiones de una administración" (Harjinder & Prakash, 1996, p.4).

El Datawarehouse es un conjunto de componentes y procesos que integran una arquitectura diseñada para el almacenamiento de la información y orientada a satisfacer las necesidades de análisis y consultas que realizan los actores responsables de la empresa para dar soporte a los procesos de acción y decisión.

El propósito del Datawarehouse es asistir a los responsables ejecutivos de la empresa para la comprensión de las realidades del pasado y adoptar elementos para la planificación y toma de decisiones de futuro a corto, medio y largo plazo.

El Datawarehouse ofrece algunas de las siguientes funcionalidades:

- Integración de bases de datos heterogéneas de orígenes diferentes.
- Ejecución de consultas complejas, no predefinidas, con diferentes niveles de agrupamiento, totalización de datos y ofreciendo la visualización de la información de forma gráfica e interpretable.
- Agrupación de los datos de forma interactiva.

- Análisis de los datos de forma interactiva.
- Análisis de los problemas en términos de diferentes dimensiones (por ejemplo, dimensión tiempo).
- Control de calidad de los datos para asegurar consistencia, homogeneidad y relevancia.

Características

Diché (2000) plantea las siguientes características de un Datawarehouse:

- **Integrado:** los datos almacenados integrarse en una estructura consistente, por lo que las inconsistencias existentes entre los diversos sistemas operacionales deben ser eliminadas.
- **Temático:** contiene sólo los datos necesarios para el proceso de generación del conocimiento del negocio.
- **Histórico:** el Datawarehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- **No volátil:** la información de un Datawarehouse existe para ser leída, y no modificada, por tanto permanente, significando la actualización del Datawarehouse la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que ya existía.

Ventajas

- Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio, por lo que las decisiones empresariales se hacen más rápidas debido a que el personal ejecutivo está más y mejor informado.

- Los procesos empresariales pueden ser optimizados al poder simplificar dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.
- Las conexiones y dependencias entre procesos empresariales se vuelven más claros.
- La visibilidad, accesibilidad y conocimiento de los datos producen mayor confianza en los sistemas operacionales.
- Facilita la aplicación de técnicas estadísticas de análisis y modelado para encontrar relaciones ocultas entre los datos del almacén; obteniendo un valor añadido para el negocio de dicha información.
- Proporciona la capacidad de aprender de los datos del pasado y de predecir situaciones futuras en diversos escenarios.
- Supone una optimización tecnológica y económica en entornos de centro de información, estadística o de generación de informes.

Desventajas

- Los Datawarehouse crecen y se vuelven cada vez más complejos, lo que ocasiona un bajo rendimiento en las consultas al mismo y el modelo se hace cada vez más ineficiente.
- El Datawarehouse al tener una administración central puede ocasionar que los requerimientos de información de unidades se vean afectados entre sí.
- La elaboración de un Datawarehouse corporativo a partir de los sistemas OLTP plantea un desafío técnico superior, necesidades de personal y un tiempo de diseño, desarrollo e implantación sumamente altos, que puede no ser costado a largo plazo.

- El desarrollo del Datawarehouse es a largo plazo que puede no satisfacer necesidades y requerimientos de información a corto plazo necesarios en la empresa. (Strange, 2002).

2.1.1.2. CRM (Customer Relationship Management)

Es una estrategia de negocios para seleccionar y manejar clientes con la finalidad de optimizar su valor a largo plazo. CRM requiere una filosofía y cultura de negocios centrada en el cliente para soportar efectivamente los esfuerzos de mercadeo, ventas y proceso de servicio.

Para Conde (2002), CRM es un nuevo enfoque de la comercialización basado en la gestión de las relaciones individualizadas con el cliente para conseguir un incremento de los beneficios a través de una oferta personalizada del producto y/o servicio. Se basa en la utilización de la información recogida en la relación con los clientes, con objetivo de establecer en la gestión con el cliente, el ciclo CRM: identificar, diferenciar, personalizar, interactuar.

El objetivo primario del CRM es obtener mayores ingresos, y no recortar costos.

Para ello Conde (2002) plantea que las soluciones de CRM deben:

- Mejorar los esfuerzos de ventas y marketing, y le permite a las organizaciones proporcionar un mejor servicio a los clientes.
- Asistir a las organizaciones a mejorar sus procesos de ventas remotas y el manejo de cuentas.
- Permitir la formación de relaciones individualizadas con los clientes, con el objetivo de mejorar el nivel de satisfacción de los mismos y maximizando los

beneficios, identificando los clientes más rentables y dándoles a éstos el mejor nivel de servicio posible.

- Proveer a los empleados de la información y procesos necesarios para conocer mejor a sus clientes, entender sus necesidades, y efectivamente construir relaciones productivas y mutuamente beneficiosas entre la compañía, sus clientes y sus asociados.

2.1.1.3. Inteligencia de negocios

Inteligencia de negocios o Business Intelligence (BI) es un término acuñado por la consultora Gartner Group a finales de la década de los 80 y describe, básicamente, la capacidad de los integrantes de una empresa para acceder a la información residente en una base de datos y explorarla, de manera que el usuario pueda analizar esa información y desarrollar con ella teorías y conocimientos que serán básicos para la toma de determinadas decisiones críticas para el negocio.

Es imperativo para las empresas, dotar a sus directivos y los que toman las decisiones del negocio, de rápidas respuestas a sus preguntas accediendo inmediatamente a la información que necesitan.

El problema de las empresas no es la falta de datos como tal, sino el gran volumen de estos, que impide ver de forma global el estado del negocio.

Con herramientas de Business Intelligence, las compañías pueden dotar a sus usuarios autorizados de acceso a un único punto de información que acceda a todos los sistemas de la empresa: intranet, extranet, o E-Business, Datawarehouse, sistemas ERP's, Datamart o aplicaciones a la medida, de una manera fácil sin necesidad de conocimiento técnico y lo más gráfico de entender posible. La clave

para convertir la información en accesible, es dar a los usuarios, herramientas para encontrar respuestas a sus preguntas de una manera sencilla.

Los beneficios de un sistema de Business Intelligence abarcan la reducción de costos, aumento de beneficios, mejora en la satisfacción del cliente y mejoras en la comunicación empresarial.

2.1.2. Definición de Datamart

Un Datamart según Lane (1999) es una forma más sencilla de un Datawarehouse que está enfocado a una sola área funcional tales como ventas, finanzas o mercadeo. Debido a que se centra únicamente en una sola área, los Datamart se constituyen de menor cantidad de fuentes de datos que los Datawarehouse, las cuales pueden ser sistemas operacionales internos o un Datawarehouse interno o externo.

Para Bill Inmon (1999), lo más importante en la definición de un Datamart, constituye que el departamento de la organización propietario del mismo posea el hardware, el software y datos que lo constituyen. Al poseer los derechos de propiedad de Datamart el departamento tiene el control y disciplina de los datos encontrados en el mismo.

Los Datamart presentan las mismas características de integración, no volatilidad, orientación temática e integración que el Datawarehouse.

2.1.3. Características del Datamart

Bill Inmon (1999) plantea que un Datamart debe presentar las siguientes características:

- Se centra únicamente en un área de la empresa.
- El volumen de datos que maneja es inferior al de un Datawarehouse.
- No dispone del nivel de detalle ni los históricos que puedan disponer otros sistemas.
- Son empleados en su mayoría como soporte para la toma de decisiones.

2.1.4. Clasificación del Datamart

Según Inmon, un Datamart se clasifica de la siguiente manera:

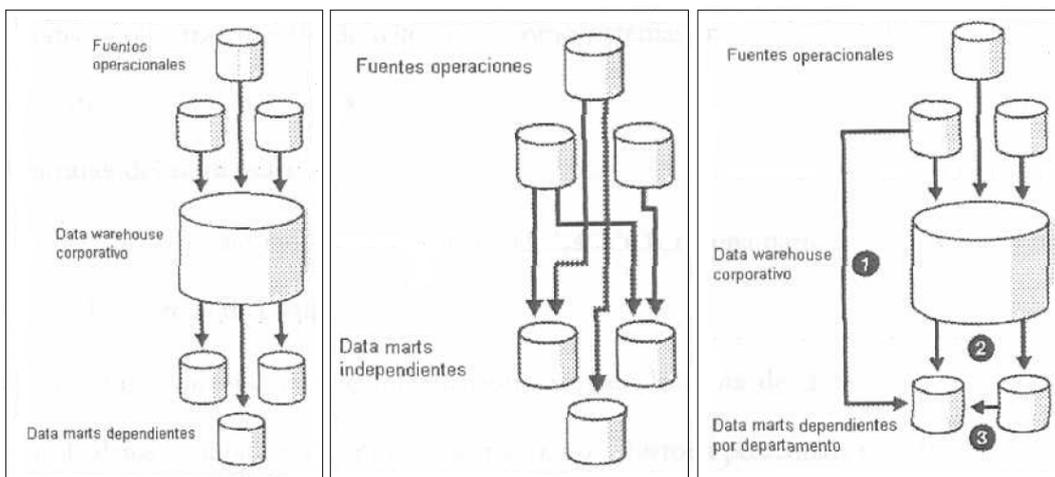


Figura II.1 Datamart dependiente.
Tomado de "Oracle8i Data Warehousing Guide". Traducción: Alvaro Trak y Emma Zambrano

Figura II.2 Datamart independiente.
Tomado de "Oracle8i Data Warehousing Guide". Traducción: Alvaro Trak y Emma Zambrano

Figura II.3 Datamart híbrido.
Tomado de "Oracle8i Data Warehousing Guide". Traducción: Alvaro Trak y Emma Zambrano

2.1.4.1 Datamart dependiente

Los Datamart dependientes son aquellos que reciben los datos desde una Datawarehouse. En este tipo de Datamart la fuente de los datos es única, tal como se muestra en la figura II.1.

2.1.4.2. Datamart independiente

Los Datamart denominados independientes son aquellos que toman sus datos directamente desde los sistemas transaccionales y no dependen de otros Datawarehouse. Este tipo de Datamart se alimenta generalmente de los denominados “*legacy systems*” de las organizaciones. Un esquema de este tipo de Datamart se muestra en la figura II.2

La ventaja de los Datamart dependientes según Inmon (1999) es que son de estructura y arquitectura estable a medida que crece la organización porque se alimentan con el Datawarehouse, el cual crece junto a la organización. Sin embargo, los Datamart independientes manifiestan sus deficiencias posterior a su elaboración cuando existen ya diversos Datamart inconexos.

2.1.4.3. Datamart híbrido

Los Datamart híbridos permiten combinar las fuentes de datos de un Datawarehouse corporativo con otras fuentes de datos tales como sistemas transaccionales y/o operacionales, tal como se muestra en la figura II.3

2.1.5. Ventajas del Datamart

- Los Datamart se ajustan mejor a las necesidades que tiene una parte específica de un negocio, más que a la de toda una empresa.
- Optimizan la distribución de información útil para la toma de decisiones y se enfocan al manejo de datos resumidos o de muestras, más que a la historia presentada con detalle.

- No necesitan ser administrados centralmente por el departamento de sistemas de una organización, sino que pueden estar a cargo de un grupo específico dentro del área de la empresa que los utilice.
- Disminuyen significativamente el coste de creación y de operación, lo cual los pone al alcance de muchas compañías.
- Con los Datamart se puede llegar a prototipos más rápidamente y obtener sistemas completamente desarrollados e implementados en tiempos menores a los del Datawarehouse.
- Las pequeñas empresas y los departamentos autónomos de una organización prefieren utilizar los Datamart para construir su propio mecanismo para toma de decisiones.
- Los departamentos de sistemas construyen un Datamart que les permite ganar experiencia y el apoyo dentro de cuadros gerenciales de la organización.

Al comenzar con un plan modesto e ir creciendo conforme se aprende más sobre la fuente de los datos y sobre las necesidades finales del usuario, permite a las organizaciones justificar el uso de los Datamart conforme estos avanzan.

- Los proyectos que comienzan como Datawarehouse pueden evolucionar a Datamart. Cuando las organizaciones acumulan grandes cantidades de datos históricos para el apoyo de decisiones, que rara vez o nunca usan, puede reducir la información guardada y convertir su Datawarehouse en uno o varios Datamart mejor enfocados.

2.1.6. Desventajas del Datamart

While (1999) revela que puede existir la tendencia en la organización de elaborar numerosos Datamart que posteriormente se conviertan en especie de islas de

información inconexas y con datos replicados de los sistemas operacionales y transaccionales.

2.1.7. Diferencias entre un OLTP y un Datamart

Las diferencias de un Datamart con un sistema OLTP se pueden resumir en las siguientes:

	OLTP	Datamart
Orientación	Orientado hacia las transacciones	Orientado hacia conceptos de negocio
	Orientado hacia las operaciones	Orientado al análisis
Tipo de operaciones	Predomina la actualización e inserción de registros	Predomina la consulta
Frecuencia de actualización	Se actualiza en línea con las operaciones diarias.	Representa valores por un período de tiempo mensual, trimestral, etc.
Actividad más frecuente	La actividad más importante es de tipo operativo (día a día)	La actividad más importante es el análisis y la decisión estratégica
	Predomina el proceso puntual, es decir, trabaja con base en transacciones	Predomina el proceso masivo. Se trabaja con base en resúmenes, agrupaciones y/o totalizaciones de grandes volúmenes de datos
Características a nivel de sistema	Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Características de los datos	Datos en general desagregados, detallados	Datos en distintos niveles de detalle y agregación, totalizados.

	Importancia del dato actual	Importancia del dato histórico
Tiempo de respuesta y desempeño	Sensitivo al tiempo de respuesta y al desempeño	Poco sensitivo al desempeño y a los tiempos de respuesta
Tipo de estructura	Estructura relacional	Visión multidimensional
	Estructura estática	Estructura flexible
Tipo de usuarios que lo emplean	Manejado generalmente por usuarios operativos	Los datos son manejados por personal de asuntos estratégicos, mercadeo, personal gerencial.
Información contenida	Contiene información procesada	Contiene información bruta (datos)
Tipo de actualización	Actualización en línea	Actualización en batch
Diseño de base de datos	Generalmente contiene muchas tablas de base de datos con pocas columnas	Generalmente contiene pocas tablas de base de datos con muchas columnas
Utilización de la información	Utilización de la información relacionada con la operación de cada aplicación	Utilización de toda la información interna y externa relacionada con el negocio

Tabla II.1 Diferencia entre OLTP y Datamart
Elaboración: Alvaro Trak y Emma Zambrano

2.1.8. Diferencias entre un Datawarehouse y Datamart

Las diferencias entre un Datawarehouse y un Datamart se pueden resumir en las siguientes:

	Datawarehouse	Datamart
Alcance	Construido para satisfacer las necesidades de información de toda la organización	Construido para satisfacer las necesidades de un área de negocios específica.
Objetivo	Diseñado para optimizar la integración y la administración de los datos fuente	Diseñado para optimizar la entrega de información de soporte a decisiones
Características de los datos	Administra grandes cantidades de datos históricos a nivel atómico	Se concentra en administrar resúmenes y/o datos totalizados
Pertenencia	Pertenece a toda la organización	Pertenece al área de negocio al cual está orientado
Administración	Es administrado por la unidad de sistema de la organización	Es administrado por el personal de sistema de la unidad propietaria del Datamart

Tabla II.2 Diferencia entre Datawarehouse y Datamart
Elaboración: Alvaro Trak y Emma Zambrano

2.2. MODELO MULTIDIMENSIONAL

2.2.1. Definiciones

2.2.1.1. Dimensión

Una dimensión es una estructura que clasifica los datos en el orden que los usuarios establecen mediante las consultas empresariales. Algunas de las dimensiones más utilizadas son Cliente, Producto, Región Geográfica y Tiempo. Para identificar las dimensiones es necesario revisar las tablas, especialmente las llamadas tablas de

hechos. El conjunto de dimensiones define lo que se denomina Cubo de Dimensiones el cual se puede apreciar en la figura II.4.

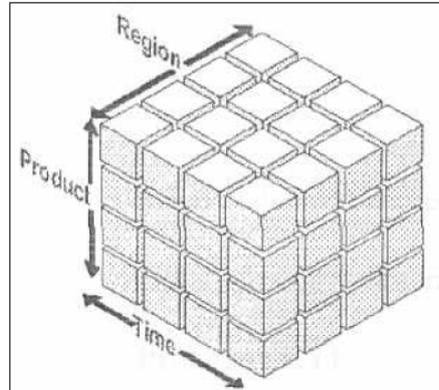


Figura II.4 Ejemplo de cubo de dimensiones
Fuente: Oracle Data Warehousing Guide

2.2.1.2. Tabla de hechos

La tabla de hechos, también llamada de detalles, es el núcleo central en un esquema de modelado de estrella y de copo de nieve y se caracteriza por poseer una clave compuesta, la cual se compone de las tablas de dimensiones. Un ejemplo de la tabla de hechos son los detalles de las ventas y los detalles de los consumos de un producto.

2.2.2. Modelo Copo de Nieve (Snowflake)

En este modelo la estructura de las tablas de dimensiones es descompuesta o normalizada en una estructura de árbol que puede disponer de muchos niveles o jerarquías.

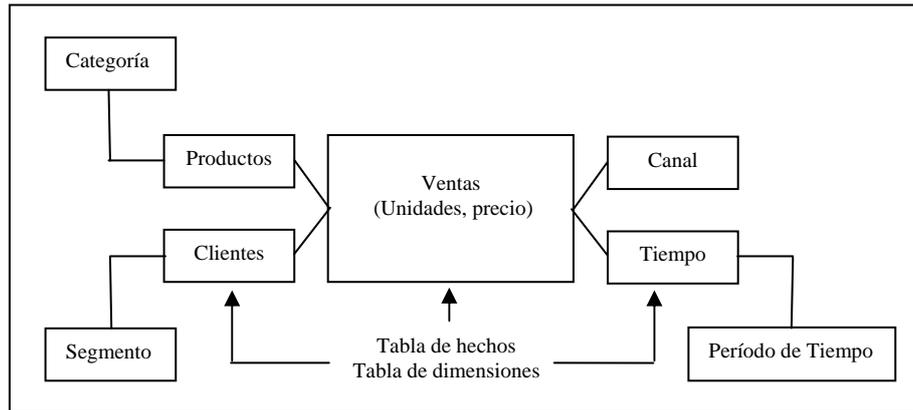


Figura II.5 Ejemplo de modelo Snowflake o copo de nieve
Fuente: Elaboración propia

Algunas de las ventajas de este modelo son las siguientes:

- Requiere menor espacio de almacenamiento que el modelo tipo estrella al minimizar la redundancia de datos.
- El modelo puede mejorar el desempeño, flexibilidad y mantenimiento con respecto a otros.

En cuanto a las desventajas, se pueden mencionar las siguientes:

- El modelo puede comprometer el entendimiento de los usuarios finales porque es de mayor complejidad.
- El modelo puede impactar sobre el desempeño de las búsquedas en caso de requerir relaciones entre tablas de hechos y dimensiones no relacionadas directamente.

2.3. PROCESOS ETL

2.3.1. Definición

Los procesos ETL1 (Extracción, Transformación y Carga) son aquellos procesos encargados de alimentar el Datawarehouse o Datamart de datos. (Palomar & Trujillo, 2001). Sin embargo, Lane (1999) prefiere denominarlos ETT2

(Extracción, Transformación y Transporte, por sus siglas en inglés) los cuales tienen como misión extraer los datos desde distintas fuentes y colocarlos en el Datawarehouse, luego de una serie de transformaciones.

Los procesos ETL surgen a raíz de la aparición de los primeros Datawarehouse, cuando empezaron a aparecer los primeros problemas para transformar los datos de los legacy systems, a menudo incomprensible y formatos transaccionales a información comprensible que un hombre de negocios pudiera consultar.

Los procesos ETL no se limitan a tomar datos de un lugar y colocarlos en otro; son capaces de hacer otras tareas como:

- Reunir datos de sistemas operacionales separados.
- Convertir los datos de un formato a otro.
- Modificar los datos para que sean más comprensibles o más completos.
- Cargar los datos modificados en el Datawarehouse o Datamart destino.

Al momento de realizar un análisis del negocio, se requiere que los datos de uno o más sistemas operacionales sean extraídos y copiados en un mismo repositorio. Este proceso de la lectura y de elaborar los datos puede ser difícil, por lo que necesita ser realizado sobre una base regular de reglas de negocio.

2.3.2. Etapas de los procesos ETL

Los procesos ETL1 (Extracción, Transformación y Carga) son aquellos procesos encargados de alimentar.

2.3.2.1. Extracción

La extracción es la operación que tiene como meta obtener los datos desde las fuentes de información. Esta operación se puede llevar a cabo mediante la extracción de los datos desde un archivo que los contenga o directamente mediante una conexión de red desde el sistema que los contiene. Los datos obtenidos de esta etapa serán posteriormente transformados y cargados en el almacén de datos, ya sea un Datawarehouse o un Datamart.

Según Palomar & Trujillo (2001), las fuentes de datos que normalmente suministran los insumos para esta operación son:

- Base de datos operacionales de la organización o bases de datos externas.
- Archivos de texto plano.
- Datos provenientes de documentos de texto, hojas de cálculos, etc.
- Datawarehouse empresarial, en caso que el destino del proceso ETL sea un Datamart.
- Datamart de la organización.

Técnicas de Extracción de datos

Las técnicas para la extracción se pueden clasificar en dos categorías: (Lane, 1999)

- Mediante el uso de archivos de datos (Data files)

Mediante la conexión distribuida de las fuentes de datos de origen con el repositorio de destino.

- Mediante la captura de los cambios ocurridos desde la última actualización.

2.3.2.2. Transformación

Lane (1999) plantea que la transformación de los datos para el Datamart se puede definir como una serie de pasos, que consiste en llevar a cabo las transformaciones de cada atributo a considerar en el Datamart de forma separada mediante operaciones de SQL y/o procedimientos en lenguajes estructurados; por lo que se crean tablas temporales donde se almacenan los resultados parciales de estas operaciones. Para llevar a cabo esta estrategia se debe proveer de checkpoints o puntos seguros dentro del flujo de transformaciones con miras a llevar el monitoreo del proceso y poder reiniciar el mismo en caso de alguna falla.

Esta estrategia provee al proceso un mejor desempeño por cuanto se lleva a cabo en etapas, pero dificulta la modificación, inserción y eliminación de las transformaciones y por ello se utilizan los checkpoints.

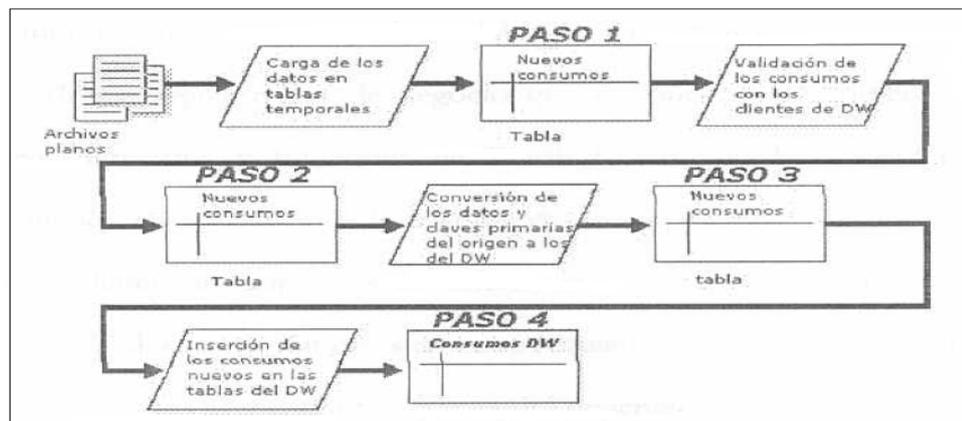


Figura II.6 Ejemplo de un flujo de transformaciones
Fuente: Lane, 1999. Traducción: Alvaro Trak y Emma Zambrano

Algunas de las técnicas empleadas en la transformación son las siguientes:

Transformaciones durante el proceso de carga

Existen herramientas de carga que proveen la funcionalidad de llevar a cabo transformaciones básicas en los datos tales como conversiones en los tipos de datos, manejo de la nulidad en los campos, sustituciones de valores y comprobación de integridad. Según Palomar y Trujillo (2001), estas herramientas son de migración de datos, de limpieza de datos (data scrubbing) y herramientas de auditoría de datos.

Transformaciones mediante SQL y /o procedimientos de PL/SQL

Otra técnica empleada es usar consultas de SQL y operaciones mediante PL/SQL una vez que se ha extraído la carga de datos desde su origen y han sido depositados en los repositorios temporales. El uso de SQL se manifiesta con la creación de tablas mediante consultas SQL, lo cual es una herramienta poderosa para la manipulación de grandes volúmenes de datos. Mediante SQL se puede llevar a cabo sustitución o actualización de datos, es decir, se pueden realizar modificaciones o actualizaciones sobre algunos atributos de las filas. También se puede realizar la transformación de la clave primaria de las tuplas extraídas para hacer su equivalencia con las claves que se manejan dentro del Datamart.

PL/SQL se emplea generalmente cuando se deben hacer complejas transformaciones en los datos y cuando se deben seguir reglas de negocio que no pueden ser cubiertas con simples consultas. Los procedimientos y funciones de PL/SQL ejecutan las consultas de SQL y almacenan los resultados en las tablas de la base de datos. Un ejemplo del uso de PL/SQL se manifiesta cuando se deben unificar datos de diferentes sistemas operacionales con diferentes formatos para llevarlos al Datamart. En estos casos se requiere un programa que analice las tuplas

entrantes para decidir cuál será el tratamiento que tendrá dependiendo de sus características.

2.3.2.3. Carga

Consiste en depositar los datos en el sistema destino y puede llevarse de la siguiente forma:

- **Cargas secuenciales:** se realiza una serie de cargas organizadas una tras otra y generalmente de grandes volúmenes de datos.
- **Procesos por lotes (batches):** se realiza una serie de cargas generalmente largas que son supervisadas por el administrador de la base de datos y son generalmente realizadas por programas de forma periódica con la finalidad de mantener actualizado el Datamart.
- **Procesamiento paralelo y técnicas incrementales:** esta técnica se caracteriza por cargar sólo los datos nuevos y modificar aquellos que han sido alterados desde su origen. Esta técnica permite la consulta del Datamart aún cuando se esté llevando a cabo la actualización y realiza comprobaciones periódicas (auditorias) con la finalidad de revisar la consistencia en los estados del proceso.

2.4. DATAMINING

2.4.1. Definición

Data Mining es el proceso que permite el descubrimiento y cuantificación de las relaciones predictivas en los datos, permitiendo transformar la información disponible en conocimiento útil de negocio. Estas relaciones predictivas se refieren a patrones, perfiles y tendencias que se descubren en los datos a través del uso de tecnologías de reconocimiento de patrones, técnicas estadísticas y matemáticas.

Posee un gran potencial para ayudar a las organizaciones a concentrarse en la información más importante de su Datawarehouse o de sus Datamart.

Un factor clave para el éxito de los proyectos de Datamining es presentar los resultados de una forma clara, haciendo hincapié en la información realmente útil.

Según Bernal y Moneo (2002), para construir un modelo útil para Datamining se deben descubrir dos aspectos dentro del Datamart o Datawarehouse:

- Los atributos significativos o de relevancia en el estudio.
- Los intervalos que identifican un patrón o segmento de clientes de la organización.

2.4.2. Ventajas del Datamining

- **Descubrimiento automatizado de modelos previamente desconocidos:** las herramientas de data Mining realizan una búsqueda minuciosa en las bases de datos identificando modelos desconocidos. Una aplicación de esta característica es la identificación de operaciones fraudulentas de tarjetas de crédito.
- **Transformar la información en decisiones útiles y eficaces:** el Datamining solo descubre relaciones entre datos. Son los usuarios que emplean las herramientas de minería de datos los que interpretan la información que arrojan estas herramientas para la toma de decisiones en la organización. Por ello que el factor más importante en el Datamining es el conocimiento y la experiencia de estos usuarios que utilizaran información concerniente a las relaciones de datos que arroje la minería de datos.

Las conclusiones en la minería de datos no son valiosas por sí mismas, sino en la medida en que se apliquen para obtener resultados. Por esta razón que la

herramienta de Datamining debe ser empleada por aquellos usuarios, denominados por Diché (2001, p.243) “trabajadores del conocimiento”. En la figura II.7 y II.8 se puede observar las clases de usuarios dentro de la organización y el tipo de herramientas que emplean.

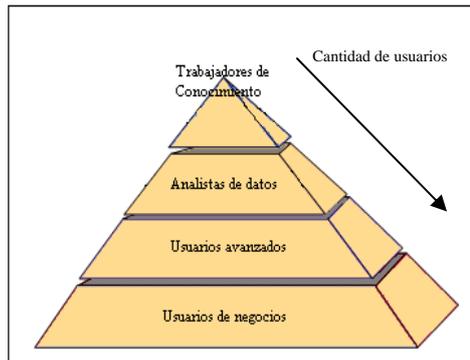


Figura II.7 Clases de usuario dentro de la organización
Fuente: Diché, 2001, p.243

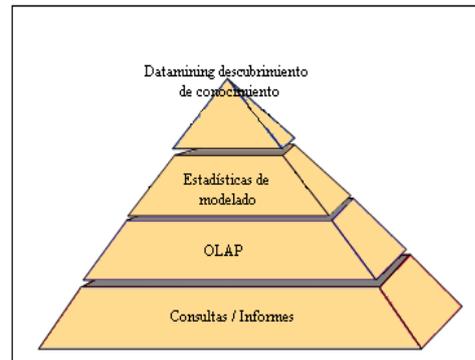


Figura II.8 Clases de herramientas usadas por los usuarios de la organización.
Fuente: Diché, 2001, p.243

- Permite la construcción de modelos automáticos e inteligentes: el Datamining automatiza el proceso de búsqueda para hallar información valiosa dentro de Datawarehouse o Datamart. Algunas aplicaciones de este proceso son los estudios de mercado y análisis de riesgo financiero en préstamos.

2.4.3. Técnicas del Datamining

2.4.3.1. Análisis estadístico: utilizando las siguientes herramientas de estadística:

- **ANOVA:** o análisis de la varianza, el cual contrasta si existen diferencias significativas entre las medidas de una o más variables continuas en grupo de población distintos.
 - **REGRESIÓN:** Define la relación entre una o más variables y un conjunto de variables predictoras de las primeras.
 - **CHI CUADRADO:** Contrasta la hipótesis de independencia entre variables.

- **CLUSTER:** Es una técnica de reducción de datos que agrupa elementos basados en características similares. Los elementos de cada conglomerado son lo más parecidos posible entre sí y lo más diferenciados posible respecto a los elementos de los otros conglomerados. Esta técnica es útil para encontrar segmentos de clientes basándose en características demográficas, financieras, etc., dando como resultado segmentos homogéneos de datos.

2.4.3.2. Árboles de decisión:

Los árboles de decisión son estructuras en forma de árbol que representan un conjunto de decisiones. Estas decisiones generan reglas para la clasificación de un conjunto de datos. Entre los métodos específicos de árboles de decisión se incluyen el método de Árboles de Clasificación y Regresión (CART: Classification And Regression Tree) y el método de Detección de Interacción Automática de Chi Cuadrado (CHAID: Chi Square Automatic Interaction Detection).

Los árboles de decisión agrupan los datos en conjuntos, cuyas reglas tienen un efecto distinto en una variable objetivo. Por ejemplo, se pueden buscar las características de una persona que respondería con mayor probabilidad a una acción de marketing directo, maximizando la efectividad de las campañas de la organización. Estas características podrían ser trasladadas a un conjunto de reglas. Mediante estas reglas se pueden identificar a los clientes que respondan a la campaña basándose en las características de los que respondieron a las campañas previas.

2.4.3.3. Redes neuronales:

Son modelos predecibles no-lineales que aprenden a través del entrenamiento, tal y como funciona la estructura de una red neuronal biológica de los seres vivos. Las

redes neuronales aprenden a partir de un conjunto de entradas, y ajustan los parámetros del modelo según esta nueva información para hallar patrones en los datos. Es una herramienta adecuada para la predicción, modelos de respuesta, análisis de riesgo y asignación de crédito.

2.4.3.4. Inducción de reglas:

Es una técnica que sirve para descubrir en los datos un conjunto de reglas del tipo “si – entonces” (if-then) asignando una determinada probabilidad al caso. Es una técnica potente que busca todos los patrones posibles y es de fácil comprensión para el usuario.

2.4.3.5. Modelos de asociación:

Son modelos que examinan en qué medida los valores de un campo dependen de, o se predicen por, valores de otro campo. Estos modelos descubren reglas acerca de artículos que aparecen juntos en transacciones de compra. Se utilizan en estudios de empresas de distribución, para analizar patrones de compra y decidir la ubicación de los productos en el establecimiento.

2.4.3.6. Lógica difusa:

A diferencia de la lógica booleana, la lógica difusa rompe con el conjunto de valores dicotómicos (sólo dos) para tener un conjunto de valores que se acerquen más al razonamiento humano.

2.4.3.7. Series temporales:

Es el conocimiento de una variable a través del tiempo para, a partir de este conocimiento, y bajo el supuesto de que no van a producirse cambios estructurales, poder realizar predicciones. Se pueden aplicar de forma conjunta con los métodos anteriores.

2.4.3.8. Algoritmos genéticos:

Son técnicas de optimización que usan procesos tales como combinaciones genéticas, mutaciones y selección natural en un diseño, basado en los conceptos de evolución biológica.

2.4.4. Diferencias entre consultas, Datamining y sistemas OLAP

Entre las diferencias entre minería de datos y OLAP se pueden mencionar:

- Las herramientas OLAP y consulta ofrecen información limitada para la toma de decisiones, ya que aportan información sobre lo que ha ocurrido pero no indica cuáles son las causas que han originado esa situación, siendo el analista quién debe deducirlo. En cambio, las herramientas de Datamining pueden determinar los factores que originan la situación mediante el descubrimiento de patrones o hechos del negocio relevantes en el estudio.
- El análisis que realizan las herramientas OLAP es dirigido por el usuario, es deductivo, parte de una pregunta o una hipótesis, analizándose los datos para resolver esa consulta concreta. Por el contrario, la minería de datos permite razonar de forma inductiva para llegar a una hipótesis general que modele el problema.
- Las aplicaciones OLAP trabajan generalmente con datos agregados para obtener una visión global del negocio. Mientras que el Datamining trabaja con datos

individuales, concretos, descubriendo los patrones que existen entre sí y generalizando a partir de ellos.

	Consultas	OLAP	Datamining
Alcance	Extracción de detalles de información resumida.	Tendencias, pronósticos e información resumida.	Descubrimiento de conocimiento y patrones ocultos.
Resultado	Información	Análisis	Conocimientos y pronósticos
Características de los resultados	Información específica de los datos.	Información resumida y analizada según requerimiento.	Pronósticos y causas que lo ocasionan basado en los datos.

Tabla II.3 Diferencia entre consultas, OLAP y Datamining
Basado en: Oracle9i Datamining. Traducción: Alvaro Trak y Emma Zambrano

2.5. SISTEMAS DE SOPORTE A LAS DECISIONES

2.5.1. Definición

Los Sistemas de Soporte a la Decisión (Decision Support Systems, DSS) son sistemas basados en computación que ayudan a quienes deciden a tomar mejores decisiones.

Los SSD pueden dar soporte individual o de grupo, por una vez o recurrente, tanto en decisiones de problemas estructurados como semi-estructurados que requieren del juicio humano.

Los SSD son sistemas interactivos que incorporan datos y modelos de datos. El personal que toma las decisiones y los profesionales en todos los niveles de la

organización emplean uno o más tipos de SSD incluyendo los denominados “ad-hoc”, organizacionales, de grupo y sistemas de información ejecutiva.

Los SSD permiten al personal que toma las decisiones combinar su juicio personal con las salidas de computadora utilizando modelos cuantitativos y elementos de bases de datos para la solución de problemas.

2.5.2. Tipos de Sistemas de Soporte a las Decisiones

Los sistemas de soporte a las decisiones existentes son: Sistema de Soporte a la toma de decisiones (DSS), Sistemas de Información para Ejecutivos (EIS), Sistemas para la toma de decisiones en grupo (GDSS) y Sistemas Expertos de Soporte a la toma de decisiones (EDSS).

2.5.3. Características de los Sistemas de Soporte a las Decisiones

- Interactividad.
- Apoya el proceso de toma de decisiones estructuradas y no estructuradas.
- Tiene una utilización frecuente por parte de la administración media y alta para el desempeño de su función.
- Puede emplearse por usuarios de diferentes áreas funcionales como ventas, mercadeo, producción, administración, finanzas y recursos humanos.
- Permite que el usuario desarrolle de manera directa modelos de decisión sin la participación operativa de profesionales en informática.
- Permite la posibilidad de interactuar con información externa como parte de los modelos de decisión.
- Facilita la comunicación de información relevante de los niveles altos a los niveles operativos y viceversa, a través de gráficas.

2.6. REDES

2.6.1. Redes de computadoras

Es el conjunto de técnicas, conexiones físicas y programas informáticos empleados para conectar dos o más computadoras. Los usuarios de una red pueden compartir ficheros, impresoras y otros recursos, enviar mensajes electrónicos y ejecutar programas en otras computadoras. Para posibilitar la comunicación entre éstas es necesario un conjunto de reglas formales para su interacción, a las cuales se les denomina protocolos.

TCP/IP, acrónimo de Transmission Control Protocol/Internet Protocol (protocolo de control de transmisiones / protocolo de Internet), es un conjunto de protocolos usados para el control de la transmisión Internet. Permite que diferentes tipos de computadoras se comuniquen a través de redes heterogéneas.

2.6.2. Componentes de las redes de computadoras

Las redes de computadoras están constituídas de varios niveles de componentes, siendo el más utilizado el modelo del TCP/IP que emplea cuatro niveles:

2.6.2.1. Capa de aplicación

Es el nivel más alto y permite acceder a los servicios disponibles a través de la red de redes TCP/IP. Una aplicación interactúa con uno de los protocolos de nivel de transporte para enviar o recibir datos.

2.6.2.2. Capa de transporte

Proporciona la comunicación entre un programa de aplicación y otro. La capa de transporte regula el flujo de información y puede proporcionar un transporte confiable, asegurando que los datos lleguen sin errores.

2.6.2.3. Capa de Internet

Es la encargada de manejar la comunicación de una computadora a otra, aceptando una solicitud para enviar un paquete desde la capa de transporte, junto con un id del equipo, hacia la que se debe enviar el paquete.

2.6.2.4. Capa de interfaz de red

Se refiere a la capa que interactúa directamente con el hardware y dispositivos físicos necesarios para la comunicación. Algunas de las tecnologías más empleadas en este nivel son Ethernet, ATM, Token Ring, entre otros.

2.7. SOLUCIONES INTEGRALES

El término “Solución Integral” al que hacemos referencia en la presente tesina, se referirá al conjunto de herramientas tecnológicas diseñadas para resolver tanto el problema principal como los problemas secundarios que se han identificado en el capítulo I, logrando de esta manera un valor agregado en la solución planteada.

CAPÍTULO III

ESTADO DEL ARTE METODOLÓGICO

3.1. METODOLOGÍAS PARA EL DESARROLLO DE DATAMART

Para realizar el desarrollo del presente trabajo se han tomado en cuenta las propuestas por Harjinder – Prakash y Kimball – Ross, las cuales son aplicables tanto para el desarrollo de un Datawarehouse como para un Datamart, ya que este último se define como un subconjunto del primero.

3.1.1. Metodología propuesta por Gill Harjinder y Rao Prakash

La metodología para el desarrollo de una Datawarehouse convencional, propuesta por Harjinder & Prakash, en su libro llamado “Data Warehousing, la integración de información para la mejor toma de decisiones” (1996), en la cual se define que el Datawarehouse, al igual que el Datamart, sigue el mismo ciclo de perfeccionamiento que todos los desarrollos de software. Aunque las fases son las mismas, existen variantes únicas que se relacionan específicamente con el Datawarehouse agregando o quitando tareas dentro de estas fases.

3.1.1.1. Planeación

Durante la fase de planeación se toman las decisiones que tienen mayor impacto en el ámbito de implementación y en la magnitud del esfuerzo. Las actividades se pueden realizar en paralelo y dentro de las más destacadas se pueden mencionar: la selección de una metodología de desarrollo, la selección de la arquitectura adecuada de implementación, la estimación de planes de programa y las justificaciones de presupuesto (véase figura III.1).

Los pasos se exponen a continuación:

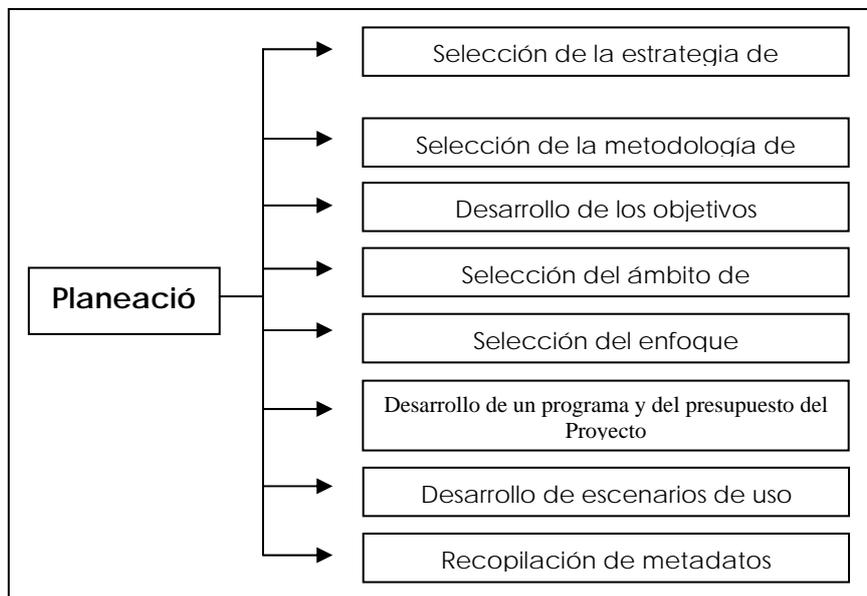


Figura III.1 Pasos para implementar la fase de planeación del sistema de Datawarehouse
Fuente: Harjinder & Prakash, 1996

Selección de la estrategia de implementación

Durante este paso, se decide la estrategia general de implementación, tomando como base la cultura de la organización y en cómo se llevan a cabo las tareas dentro de la misma. La estrategia de implementación que más se adecúa a las estrategias de la organización es:

- **El enfoque de arriba hacia abajo (Top Down)**

En esta estrategia de implementación se abordan todos los requerimientos empresariales que cubrirá el Datawarehouse propuesto; es útil cuando la organización ya tiene experiencia en la implementación de la tecnología y por lo tanto, están bien entendidos y claros los problemas que la solución resolverá.

- **El enfoque de abajo hacia arriba (Bottom Up)**

En esta estrategia de implementación se selecciona un subconjunto específico, bien entendido, de la problemática empresarial y se formula una solución para este subconjunto. Este enfoque se toma por lo regular para implementar un Datamart, un pequeño sistema de información ejecutivo, o un Datawarehouse departamental que está claramente orientado a responder unas cuantas consultas bien escogidas en un dominio determinado, como en el análisis de mercados y la administración de productos. Se recomienda en casos donde la organización todavía no está comprometida con la tecnología del Datamart, pero busca una evaluación tecnológica para determinar cómo, dónde y cuándo desplegar esta tecnología, también en casos donde la organización trata de obtener una idea de los costos y gastos adicionales de implementar y desplegar la tecnología del Datamart. Los pros y contras de este enfoque se muestran en la Tabla III.1.

Pros	Contras
En ocasiones, las necesidades de implementación y de comenzar de una vez, superan por mucho el análisis de arriba hacia abajo y las consideraciones a largo plazo.	Después de la implementación inicial, es bueno retroceder y observar cómo puede ampliarse la solución para dar servicio a toda la empresa.
En las primeras etapas de madurez de la tecnología, este enfoque permite a una organización evaluar los beneficios de la tecnología sin grandes compromisos.	La falla de un solo proyecto de abajo hacia arriba puede demorar la implementación de una tecnología benéfica en potencia.
La participación de poca gente trabajando en un ámbito reducido puede acelerar la implementación y la toma de decisiones.	El equipo inicial debe retroceder e integrarse a un mayor equipo para ampliar el ámbito de la solución inicial.

Tabla III.1 Pros y contras del enfoque de abajo hacia arriba (Bottom-Up)

Fuente: Harjinder & Prakash, 1996, p.81

- **El enfoque combinado**

Es una combinación de los dos anteriores; y es útil cuando la organización cuenta con un equipo comprometido que tiene claro donde se debe aplicar la tecnología de Datawarehouse, generalmente dicho equipo está compuesto por usuarios finales y personal de Sistemas.

Selección de la metodología de desarrollo

En teoría, un Datamart puede desarrollarse por medio de cualquiera de las metodologías de desarrollo existentes, sin embargo, los requerimientos de la implementación de un Datamart descartan el uso de una metodología que requiera una fase prolongada de acopio de requerimientos y análisis, una fase de desarrollo monolítico que tome muchos meses y una fase de despliegue que ocupe también varios meses. A continuación se describe una de las metodologías más utilizada para el desarrollo de una Datamart. (Harjinder & Prakash, 1996).

- **Método de Desarrollo Espiral**

El enfoque se basa en la observación de que es más fácil redirigir un sistema desplegado con base en nuevos requerimientos que construir una solución completa basada en requerimientos inadecuados o no disponibles. Se presta principalmente al desarrollo de aplicaciones de base de datos, al desarrollo de una Datamart y al desarrollo de sistemas Orientados a Objetos (sistemas OO).

La figura III.2 muestra las fases del modelo en espiral.

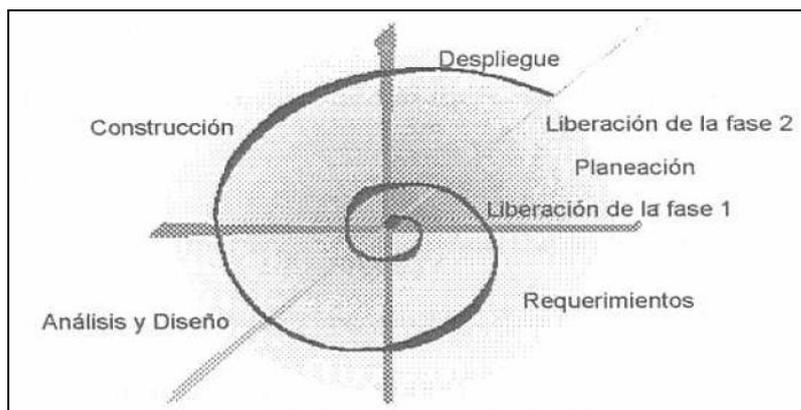


Figura III.2 Modelo de Desarrollo Espiral
Fuente: Harjinder & Prakash, 1996, p.84

Desarrollo de los Objetivos Empresariales

Durante este paso se desarrolla una lista de objetivos empresariales que el sistema debe satisfacer, es decir, en este paso se busca planear los objetivos del Datamart determinando una serie de aspectos como el mercado de destino (audiencia potencial), las plataformas de uso actual o planeadas, las diversas fuentes de datos que pueden y/o deben integrarse al Datamart y las características y funciones del mismo.

Selección de un ámbito inicial de implementación

En este paso se definen los alcances del proyecto del Datamart, se define un rumbo general y un conjunto general de objetivos, es decir, se define cuál es el departamento que necesita utilizar inicialmente el Datamart, el rango de consultas empresariales a las que debe responder inicialmente, cuáles y cuántas son las fuentes de entrada de datos, qué tan utilizables son los datos de las fuentes, qué tan bien documentadas están las fuentes de datos, cuál es la disponibilidad de modelos lógicos y herramientas CASE, si se pueden utilizar las habilidades y recursos humanos existentes, entre otros.

Enfoque de selección de una arquitectura

En este paso se selecciona la arquitectura para implementar el Datamart, ya que esta arquitectura proporciona un esquema de comunicaciones comprensible y definido, un vocabulario y un lenguaje común, que permite a todas las personas involucradas en el proyecto hablar entre ellos y entender lo que cada uno dice. Proporciona un diagrama conceptual para analizar las diversas opciones de implementación. Se cuenta con diversas opciones arquitectónicas, dentro de la que destaca la siguiente:

- **Sólo Datamart**

Esta arquitectura reconoce que cada departamento funcional en una organización tiene sus propias necesidades específicas y que un Datawarehouse corporativo no puede satisfacer todas estas necesidades. El concepto de Datamart es el de una tienda local, que da servicio a un vecindario, en vez de una gran tienda departamental que atiende a un suburbio o a una ciudad completa.

Desarrollo de un programa y un presupuesto para planes del proyecto

Uno de los aspectos más importantes de la planeación consiste en poder realizar lo siguiente:

- **Articular tanto un plan de programa como un conjunto de planes de proyecto**

Un plan de programa es una visión general de la actividad del Datamart y su función en la vida diaria y semanal de la organización. El plan de programa identifica los diversos departamentos y unidades empresariales que usarán el Datamart. Asimismo, establece prioridades sobre la implementación, para atender

primero las necesidades empresariales críticas. Sin embargo, los planes de proyecto son los planes para implementaciones específicas del Datamart.

- **Reservar un presupuesto adecuado para el programa**

La planeación de este presupuesto se basa en el siguiente enfoque:

- **Estimación del costo, con base en la arquitectura de referencia.**

Cada componente de la arquitectura de referencia puede dividirse en subcomponentes y obtenerse su costo por separado.

Desarrollo de escenarios de uso empresarial

Los escenarios empresariales ayudan a definir las expectativas del usuario final con respecto al Datamart. Algunas de estas expectativas se satisfacen, en última instancia, mediante el uso de fuentes de datos apropiadas. Otras, requieren de datos externos de fuentes adquiridas. La construcción de escenarios de uso empresarial ayuda a articular estas expectativas.

Recopilación de los metadatos

Durante esta fase, se recopilan los metadatos necesarios para la construcción del Datamart, éstos pueden venir de modelos de datos abstractos de la organización, de diccionarios de datos y de otras fuentes externas de datos.

3.1.1.2. Requerimientos

La fase de requerimientos de la implementación del Datamart es una especificación precisa de las funciones que se obtendrán del mismo. Además de las características y funciones necesarias, los requerimientos describirán con claridad el ambiente

operativo en el que se entregará el Datamart. Una forma directa de ver los requerimientos es a través de las diversas perspectivas del marco de referencia de Zachman, ya que los requerimientos de un Datamart son tan variados y diversos como las clases de usuarios que lo utilizan para obtener beneficios empresariales.

En la figura III.3 se muestran las distintas perspectivas a tomar en cuenta bajo el marco de referencia de Zachman.

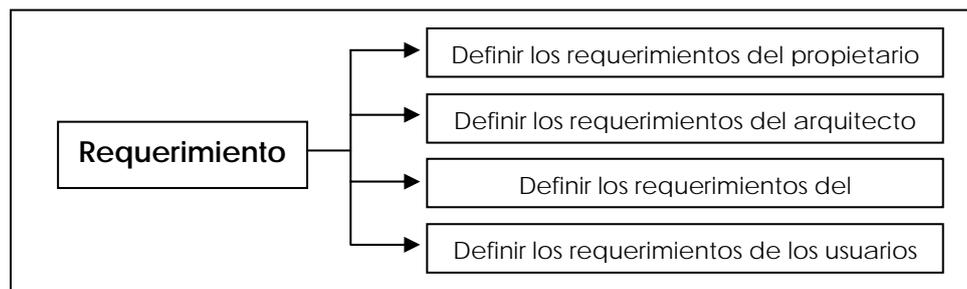


Figura III.3 Pasos para implementar la fase de requerimientos del sistema de Datawarehouse
Fuente: Harjinder & Prakash, 1996

Requerimientos del Propietario

El inversionista / dueño de una Datamart participa en la generación del concepto del Datamart por su capacidad de aprobación y revisión. Autoriza la adquisición inicial y, en ocasiones, los costos para mantener la operación y las actualizaciones necesarias para mantener el sistema útil. Durante el inicio, el apoyo principal que aporta la arquitectura a este miembro del equipo es un diagrama para realizar actividades de planeación y coordinación, predecir costos e ingresos y el manejo de riesgos. Por lo general, el punto de vista del dueño se mide en dinero, el interés del dueño en la tecnología se limita al análisis de inversiones y pagos, y la evaluación

de riesgos y programas. El inversionista o patrocinador ejecutivo considera que el Datamart forma parte del SSD (Sistema de Soporte de Decisiones) dentro de la perspectiva empresarial total.

Requerimientos del Arquitecto

El arquitecto es la persona responsable de diseñar los diversos componentes del Datamart para sustentar las necesidades actuales y futuras. Éste debe entender tanto las necesidades empresariales del Datamart, como la tecnología de implementación que se necesita para solventarlas. Para el diseño de la arquitectura, se maneja un enfoque que se conoce como Planeación de Arquitectura Empresarial (EAP por las siglas de Enterprise Architecture Planning). En esta metodología se han desarrollado los tres tipos siguientes de arquitectura:

- **Arquitectura de datos**

Describe los elementos de datos y sus relaciones, las arquitecturas de datos se describen en general como Modelos Entidad – Relación.

- **Arquitectura de Aplicación**

La arquitectura de aplicación es un catálogo de aplicaciones junto con las funciones que ofrecen y las interfaces entre ellas. También se confronta con la arquitectura de datos utilizando una matriz “CRUD” (Create, Read, Update o Delete), se establecen referencias cruzadas desde cada aplicación con uno o más elementos de datos que la aplicación crea, lee, actualiza o elimina.

- **Arquitectura de Tecnología**

Se construye dividiendo un sistema en elementos de componentes de tecnología como una computadora servidor, una estación de trabajo de usuario, la interfaz

gráfica de usuario, el sistema de base de datos y el diccionario / depósito de datos, y luego seleccionando candidatos con base en criterios de evaluación, tal como el apego a los estándares ANSI y de la industria, el costo y el apego a los estándares internos.

Requerimientos del desarrollador

Los requerimientos del desarrollador son un refinamiento de los del arquitecto, con decisiones tomadas respecto a la selección de plataformas y la separación de la arquitectura de datos y la arquitectura de aplicaciones sobre las plataformas seleccionadas. También los requerimientos del desarrollador se relacionan con descripciones detalladas de la arquitectura de tecnología para la especificación de elementos tales como el lenguaje de programación, el acceso a la base de datos y los protocolos de comunicación. Además los desarrolladores necesitan también de los siguientes requerimientos:

- **Requerimientos de Tecnología**

Los requerimientos de tecnología se desglosan según la arquitectura de referencia.

- **Requerimientos de Despliegue**

Los requerimientos de despliegue se relacionan con la capacidad del Datamart para proporcionar acceso y para distribuir información de manera oportuna y conveniente, considerando aspectos tales como los métodos de acceso, los métodos de recuperación, las herramientas de acceso, los requerimientos de conectividad y los requerimientos de plataforma cliente.

- **Requerimientos de Disposición para la producción del Datamart**

Los requerimientos de disposición para la producción se relacionan principalmente con el manejo de la solidez y la disponibilidad, la conservación de la consistencia y la precisión de la información, el manejo del desempeño al crecer el almacenamiento de datos, la definición de políticas y procedimientos para la actualización y el mantenimiento tanto del modelo del Datamart como de los datos, proporcionando control de acceso y procedimientos de seguridad.

- **Requerimientos para el desarrollo y despliegue del personal y sus habilidades**

Consiste en identificar cuáles habilidades se requieren para qué fase del ciclo de vida del desarrollo del Datamart.

Requerimientos de los usuarios finales

Los requerimientos del usuario final se pueden ubicar en categorías como el flujo de trabajo, requerimientos de consultas de datos, requerimientos de reportes, tipos de análisis de datos que deseen realizar sobre los datos del Datamart.

3.1.1.3 Análisis

La fase de análisis del ciclo de desarrollo del Datamart significa convertir los requerimientos acopiados en la fase anterior de requerimientos, en un conjunto de especificaciones que puedan apoyar el diseño, derivando modelos físicos y lógicos de datos para el Datamart y definir los procesos necesarios para conectar las fuentes de datos, el Datawarehouse, los Datamart y las herramientas de acceso del usuario final. Para entender y analizar las necesidades empresariales se utilizan las siguientes visiones: (Harjinder & Prakash, 1996, p.104)

Visión de arriba hacia abajo.

Permite la selección de la información correcta para el Datamart, de tal manera que éste cumpla con la finalidad principal de un negocio.

Visión de la fuente de datos.

La visión de la fuente de datos se modela mediante técnicas estándar de modelado de datos como la del modelo Entidad – Relación y es importante reducir los modelos de fuentes de datos en un modelo único que represente la información fuente.

Visión del Datamart.

En esta visión se utilizan técnicas de modelado como el modelo copo de nieve (Snowflake Model) y otros. Cada uno de éstos se desvía de las técnicas clásicas de aplicación de normas y refleja más de cerca la visión que tiene el usuario del Datamart.

Visión de consulta empresarial.

La visión de consulta empresarial es la perspectiva de los datos del Datamart desde el punto de vista del usuario.

3.1.1.4. Diseño

En la fase de diseño, los modelos lógicos desarrollados en la fase de análisis se convierten en modelos físicos. Los procesos identificados en la fase de análisis para conectar las fuentes de datos con el Datamart, y el Datamart con las herramientas

de estación de trabajo del usuario final, se convierten en diseños para programas que realizarán las tareas requeridas para los procesos. También se identifican y detallan los procesos que requiere, de manera interna, cada bloque de la arquitectura de referencia del Datamart. (Harjinder & Prakash, 1996, p.105). En la fase de diseño se encuentran las actividades principales como el diseño detallado de la arquitectura de datos y el diseño detallado de la arquitectura de aplicación.

Dentro de las actividades del diseño de la arquitectura de datos está el desarrollo de modelos físicos de datos para las bases de datos de almacenamiento del Datamart y la correspondencia de los modelos físicos de datos de las fuentes de datos con los modelos físicos del Datamart.

El diseño de la arquitectura de aplicación está comprendido en procesos internos a las fuentes de datos que se relacionan con depuraciones o extracciones parciales de información, procesos que conectan las fuentes de datos con el Datamart, procesos que son internos al Datamart y se usan para fines de manejo interno, procesos que conectan al Datamart con los Datawarehouse, procesos internos a los Datamart y se emplean para fines de manejo interno.

3.1.1.5. Construcción

La fase de construcción es responsable de implementar físicamente los diseños desarrollados durante la fase de diseño. Las aplicaciones que se necesitan construir son programas que crean y modifican las bases de datos para el Datamart, programas que extraigan, programas que realicen transformaciones de datos, programas que realicen actualizaciones de bases de datos relacionales y programas que efectúen búsquedas en bases de datos muy grandes.

3.1.1.6. Despliegue

La fase de despliegue tiene que ver principalmente con los retos de la instalación, puesta en servicio y uso de la solución de Datamart. Las actividades requeridas en la fase de despliegue son:

- Proporcionar la instalación inicial, incluyendo facilidades para las conexiones básicas de datos con las fuentes y para la actualización y sincronización de datos.
- Planeación y entrega de una implementación por etapas.
- Proporcionar capacitación y orientación a todo tipo de usuarios.
- Planeación e implementación de la actualización de plataformas y el mantenimiento necesario por el Datamart cuando se requiere.
- Proporcionar la administración de usuarios y sistemas.
- Proporcionar la capacidad de generar archivos permanentes y respaldos.
- Proporcionar la capacidad de recuperación.
- Asegurar la integración dentro de la infraestructura existente.
- Proporcionar controles de accesos y seguridad.
- Promover activamente entre los usuarios finales la información que proporciona el Datamart.

3.1.1.7. Expansión

Tiene que ver con el mejoramiento del Datamart en las siguientes situaciones:

- Consultas empresariales que no pueden formularse o satisfacerse.
- Consultas empresariales que requieren de datos externos que no se tomaron en cuenta en la implementación inicial.

- Algunos departamentos no quisieron implementar sus Datamart pero ahora ven la necesidad de hacerlo.

3.1.2. Metodología propuesta por Ralph Kimball y Margy Ross

Proponen que un Datawarehouse no es más que: "la unión de todos los Datamart de una entidad". Defiende por tanto una metodología ascendente (Bottom-Up) a la hora de diseñar un almacén de datos. Antes de iniciar la construcción del Datawarehouse hay que tener en cuenta lo siguiente:

- El proyecto Datawarehouse debería enfocarse a las necesidades del negocio.
- Los datos de los usuarios deberían de estar contenidos en las dimensiones propuestas por el Datawarehouse.
- Todo proyecto Datawarehouse debe tener inicio y fin.

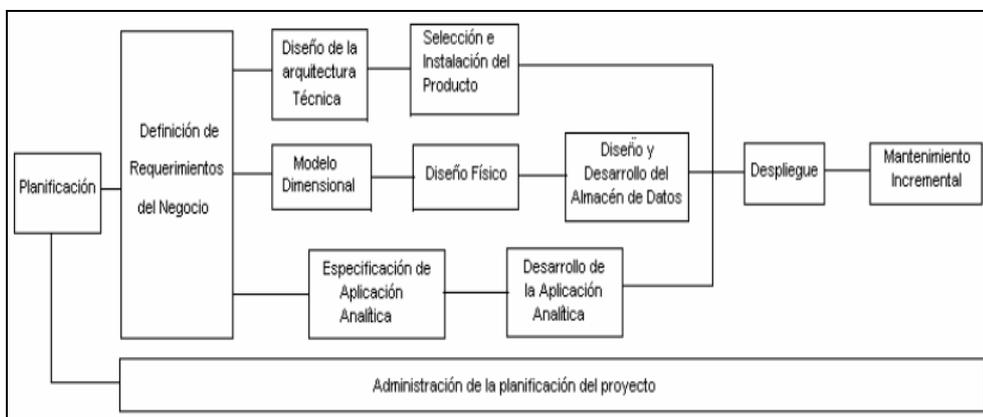


Figura III.4 Diagrama del Ciclo de vida del Datamart
Fuente: Ralph Kimball y Margy Ross

- Planificación, en esta etapa se establece el ámbito preliminar que va a tener el Datawarehouse y la justificación, se obtiene recursos y se pone en marcha el proyecto Datawarehouse.

- La administración de la planificación, este proceso se lo realiza en todas las etapas del Datawarehouse para controlar que los tiempos y alcances de cada una se cumplan.
- Definición de los requerimientos del negocio, en donde se consideran los requerimientos del negocio para la construcción del Datawarehouse, se debe tener en cuenta que esta etapa está muy relacionada con la etapa de la planificación ya que dependiendo de la cantidad de requerimientos que tenga el usuario determinará los tiempos y el alcance del Datawarehouse para su correspondiente planificación.
- Diseño Técnico de la Arquitectura, en esta etapa se establece la estructura del Datawarehouse para establecer la integración de múltiples tecnologías.
- Selección e Instalación del Producto, aquí se evalúan las opciones presentadas en el Diseño Técnico de la Arquitectura y se selecciona los productos a utilizarse.
- Modelo dimensional, en esta etapa se trasladan los Requerimientos del negocio que están enfocados hacia el modelo de datos Dimensional.
- Modelo Físico, en el cual se transforma el modelo Dimensional al modelo físico, esta etapa también involucra otras actividades como agregaciones, particionamiento, indexación, etc.
- Diseño y Desarrollo del almacén de Datos, en la que se desarrollan los procedimientos de carga y transformación de datos.
- Especificación de la aplicación Analítica, en donde se especifica las aplicaciones que van a acceder al Datawarehouse.
- Desarrollo de la aplicación Analítica.
- Despliegue
- Mantenimiento incremental

3.1.2.1. Planificación

- **Selección del ámbito**

Para seleccionar el ámbito se debe tener en cuenta hacia qué sector del negocio se va a enfocar el Datawarehouse para responder preguntas comunes de dicho sector por ejemplo: Ventas, ¿Cuánto se vendió en el mes anterior?, Clientes ¿Qué cliente compro más? Entre otras.

- **Justificación**

Se trata presentar el beneficio que se va a obtener luego de la implementación del proyecto Datawarehouse y el impacto que va a tener en otros sectores de la organización.

- **Desarrollo del plan del Proyecto**

Involucra la identificación de todas las tareas necesarias para implementar el Datawarehouse y el esfuerzo y tiempo que van a utilizarse en cada una de estas.

3.1.2.2. La administración de la planificación

Este proceso se lo realiza en todas las etapas del Data Warehouse para controlar que los tiempos y alcance de cada una se cumplan.

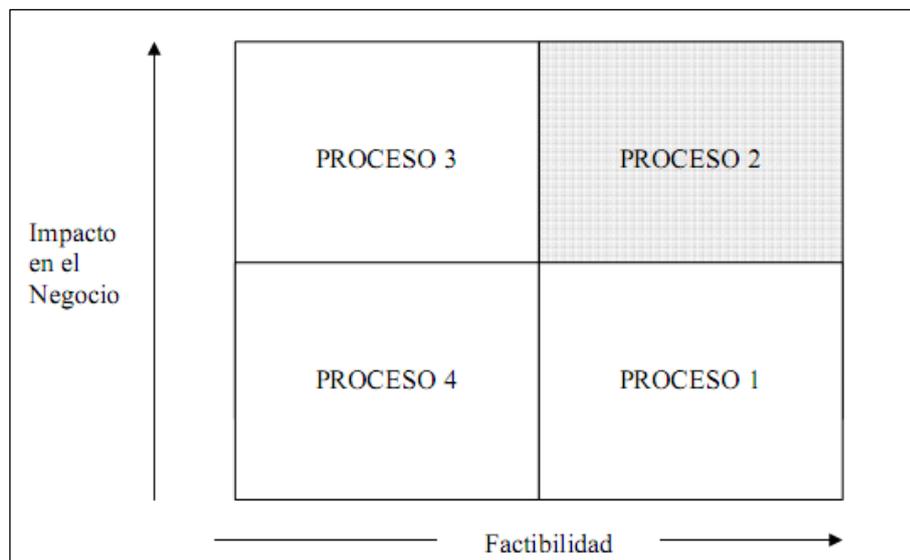
3.1.2.3. Definición de requerimientos del negocio

Antes de recopilar las necesidades del negocio es necesario tomar en cuenta los requerimientos del equipo de trabajo, seleccionar el calendario de actividades y establecer los representantes del negocio.

- **Recolectando requerimientos del negocio**

Para la recolección de los requerimientos del negocio es necesario sentarse personalmente con los encargados del sector del negocio al cual va a estar dirigido el Datawarehouse, presentarles primeramente los objetivos tanto de las entrevistas como del Datawarehouse en sí. Para lo cual se debe seguir los siguientes, pasos:

- Realización de entrevistas, cuyo objetivo es recopilar datos de los usuarios del negocio averiguando lo que hacen y porque lo hacen, y se debería empezar por preguntar acerca de las responsabilidades de sus respectivas labores y como impactan las mismas en la organización.
- Documentación de Entrevistas, lo primero que se debe realizar es la documentación de cada una de las entrevistas, y luego, un documento que contenga un consolidado de todas las entrevistas realizadas, identificando los procesos principales del negocio y luego identificando los requerimientos del mismo.
- Luego de realizar la documentación, se realizan reuniones con los representantes de la organización para discutir acerca de los procesos identificados en las entrevistas, para luego priorizar a cada una de estas como muestra la figura III.5



3.1.2.4. Diseño de la Arquitectura Técnica

La arquitectura técnica es como el plano de una casa el cual contiene el modelo y funcionalidad del almacén de datos así como de sus componentes, además provee soporte para la integración de múltiples tecnologías, el éxito de la arquitectura técnica depende mucho de la comunicación que se tenga entre cada uno de los involucrados del proceso y los administradores del proyecto Datawarehouse.

Dado que todos los proyectos de Datawarehouse deberían tener una arquitectura, esta metodología propone 8 pasos para el diseño de la Arquitectura Técnica:

- Establecer tareas.
- Recolectar requerimientos de Arquitectura.

- Documentar los requerimientos de la Arquitectura.
- Diseñar el modelo de la arquitectura de alto nivel.
- Especificar y diseñar los subsistemas para acceso al Datawarehouse.
- Determinar fases de implementación.
- Documentar la arquitectura técnica.
- Revisión y finalización del diseño de la arquitectura técnica.

3.1.2.5. Selección e instalación del producto

Para la selección e instalación del producto es necesario conocer las capacidades de adquisición de la organización así como también el enfoque al cumplimiento de los requerimientos del negocio para lo cual se desarrolla una matriz de evaluación de productos determinado los criterios entre lo que se podría incluir: funcionalidad, arquitectura técnica, características de diseño de software, entre otras. Además realizar un prototipo con las herramientas de ser necesario.

3.1.2.6. Diseño Dimensional

Una vez definidos los requerimientos del negocio, se elabora una matriz la cual contenga los procesos del negocio identificados, y las posibles dimensiones que va a contener el Datawarehouse. La figura III.6 muestra un ejemplo del diseño de la misma.

PROCESOS DEL NEGOCIO	DIMENSIONES COMUNES				
	Fecha	Clientes	Productos	Proveedores	Promociones
Ventas	X	X	X	X	X
Compras	X	X	X	X	
Inventario	X		X		X

Figura III.6 Ejemplo de Matriz de Procesos y Dimensiones
Fuente: Ralph Kimball y Margy Ross

Luego de construir esta matriz se debe priorizar los procesos del negocio que se van a cubrir primeramente y además identificar, las tablas, columnas, y definiciones de las dimensiones identificadas.

3.1.2.7. Diseño Físico

El diseño físico se lo realiza a partir del diseño dimensional generado en el paso anterior, y se definen los tipos de datos a almacenarse, así como también la estructura física de las tablas contenidas en el modelo dimensional.

- **Estrategias de Agregación**

Todos los Datawarehouse deberían contener tablas de agregación pre-almacenadas y pre-calculadas teniendo en cuenta que cada agregación de la tabla de hechos debería ocupar su propia tabla de hechos física.

Para definir las estrategias de agregación se debe considerar lo siguiente:

- Los usuarios que van a acceder al Datawarehouse.

- Determinar la distribución de los datos.
- **Estrategia Inicial de Indexación**

Las tablas de dimensiones deberían tener solo un índice por la clave primaria, y es recomendable tener índices de árboles B.

La clave primaria de la tabla de hechos es un subconjunto de claves foráneas provenientes de las tablas de dimensiones.

3.1.2.8. Diseño y Desarrollo del Almacén de Datos

Esta actividad tiene que ver con los procesos de extracción, transformación y carga de datos desde las fuentes de datos hacia el Datawarehouse.

3.1.2.9. Despliegue

Se refiere al diseño y desarrollo de las aplicaciones que van a tener acceso al Datawarehouse, y en la manera en que los usuarios van a acceder a la información presentada por el mismo.

3.1.2.10. Mantenimiento Incremental

Después de la implantación del Datawarehouse en la organización se deben tener actividades de mantenimiento, soporte y capacitación a los usuarios del mismo con el fin de recopilar de ser el caso, nuevos requerimientos para ser incorporados en el Datawarehouse implementado.

3.2. SELECCIÓN DE LA METODOLOGÍA A APLICARSE

La tabla 1.1 muestra cualitativamente las características de cada una de las metodologías descritas anteriormente tomando como valoraciones las siguientes:

Cuadro de valores

M (Muy Bueno)

B (Bueno)

R (Regular)

N (No disponible)

Nº	Parámetro de Evaluación	Gill Harjinder y Rao Prakash	Puntaje	Ralph Kimball y Margy Ross	Puntaje
1	Facilidad de seguimiento	M	3	B	2
2	Semejanza con fases de desarrollo de todo proyecto informático	M	3	R	1
3	Detalle de pasos a seguir en cada una de las fases	M	3	R	1
4	Rapidez de Implementación	B	3	R	2
5	Cantidad de fases	B	3	R	2
6	Antigüedad	R	1	M	3
7	Difusión	B	2	M	3
8	Presentación de casos de estudio	N	0	M	1
Puntaje Total			18		15

Tabla III.2 Cuadro Comparativo de Metodologías de Desarrollo de Datawarehouse

Para la selección de la metodología se ha tomado en consideración las siguientes puntuaciones para los criterios de evaluación presentados en la Tabla III.3:

Parámetros	Justificación
1,2,3,6 y 7	Se dan las puntuaciones de M=3, B=2, R=1 y N=0 debido a que estos parámetros se relacionan con el ciclo propuesto por la metodología y su facilidad de seguimiento, que si bien es un factor muy importante en la selección, no es determinante.
4 y 5	Se dan puntuaciones de M=4, B=3, R=2 y N=1 que son más altas que en el caso anterior debido a que reflejan la rapidez que ofrece la metodología para la obtener el producto final que en este caso es un factor determinante en la selección.
8	Se dan puntuaciones de N=0 que en este caso significa que carece de esta característica y M=1 en el caso contrario.

Tabla III.3 Justificación de puntuaciones asignadas a los parámetros de selección

Con las puntuaciones descritas anteriormente se obtienen los siguientes resultados:

Metodología propuesta por Gill Harjinder y Rao Prakash	18
Metodología propuesta por Ralph Kimball y Margy Ross	15

Tabla III.4 Selección de la Metodología

Por lo cual se selecciona la Metodología propuesta por Gill Harjinder y Rao Prakash para el desarrollo del Datamart propuesto.

3.3. HERRAMIENTAS PARA EL DESARROLLO DE UN DATAMART

La correcta selección de las herramientas para el desarrollo de un Datawarehouse es muy importante, ya que no solo se limita a la presentación de reportes de la información contenida en las diferentes fuentes de datos, sino que también debe tener la capacidad de acoplarse a los diferentes requerimientos de información por parte de los usuarios.

Las herramientas para la realización de un Datawarehouse se clasifican en dos categorías básicas:

- Herramientas para Consulta y Reportes de Usuario Final.
- Herramientas de Datawarehouse, dentro de las cuales se consideran: las herramientas de Almacenamiento de datos, las herramientas de Extracción, Transformación - Carga y las herramientas de Análisis Inteligente.

3.3.1. Herramientas de consulta y reportes de usuario final

Permiten la construcción de las interfaces vistas por el usuario. Al usuario se le debe proveer de un mecanismo para que obtenga los datos a un alto nivel y luego encuentre con ello la solución a preguntas específicas tales como: ¿"Qué sucedió"? (Ejemplo: ¿"Cómo comparar las ventas de los productos X, Y y Z del mes pasado con las ventas del presente mes y las ventas del mismo mes del año pasado?").

En esta categoría se encuentran diversos productos, entre los más populares están:

- Crystal Reports de Business Objects
- Oracle Reports
- SAS Web Reports
- Servicio de Reportes (Reporting Services) de Microsoft SQL, entre otras.
- Web Intelligence de Business Objects

Crystal Reports XI

Es un producto elaborado por Bussiness Objects, entre las características principales están:

- Publica fácilmente reportes y otro contenido en la Web.
- Puede guiar paso a paso para la configuración del contenido mostrado en el reporte tanto de Bases de Datos como de archivos.
- Provee de calendarización de reportes para correr a horas específicas o basadas en eventos específicos.
- Posee interactividad con los reportes en la Web, puede Imprimir, exportar y profundizar en cuadros y otros objetos de reportes
- Puede realizar búsquedas condicionales dentro de los reportes.
- Buen nivel de seguridad.

Oracle Report

- Puede guiar paso a paso para la configuración del contenido mostrado en el reporte tanto de Bases de Datos como de archivos.
- Soporta una gran variedad de fuentes de datos.
- Puede presentar los reportes en formato PDF,XML, Excel
- Se puede crear Sitio Web de Reportes Dinámicos usando código JSP
- Proporciona acceso seguro a Reportes, controlar el procesamiento y la distribución de informes, y normalizar el uso de los informes en la empresa.
- Permite la calendarización de los reportes.

SAS Web Reports

- Guía paso a paso la generación del reporte con preguntas intuitivas para el usuario en términos del negocio.
- Ofrece una rápida organización y búsqueda de reportes.

- Permite exportar los datos del reporte a formatos como PDF o Excel.
- Puede integrarse fácilmente con aplicaciones Web con arquitectura J2EE.

SQL Reporting Services

- Proporciona integración con Microsoft Internet Information Services (IIS) y permite generar un entorno de informes sobre una infraestructura de servidores Web existente.
- Los programadores pueden crear informes para publicarlos en el Servidor de informes mediante herramientas de diseño de Microsoft o de otros fabricantes que utilicen el lenguaje RDL (Report Definition Language), un estándar de la industria basado en XML que se usa para definir informes.
- Se pueden utilizar informes con parámetros para filtrar datos basados en valores que se proporcionan en tiempo de ejecución.
- Los informes se pueden representar con formatos de escritorio o basados en Web.
- Proporciona un modo de centralizar el almacenamiento y la administración de informes, proporciona acceso seguro a informes y carpetas, controlar el procesamiento y la distribución de informes, y normalizar el uso de los informes en la empresa.
- Admite la entrega de informes a petición (extracción) y basada en eventos (inserción). Los usuarios pueden ver los informes con formato Web o en el correo electrónico.

Web Intelligence

- Se puede visualizar reportes de los metadatos
- Es posible ver, editar, remover reportes, secciones o filtros.

- Proporciona la facilidad de formatear y cambia de tamaño las celdas, tablas y gráficos en línea a través de la Web.

- Se pueden insertar cálculos personalizados, añadir filas y columnas a tablas, crear y duplicar tablas y gráficos, crear fórmulas y variables en línea.

En el siguiente cuadro comparativo se muestra cualitativamente las características de cada uno de los productos antes descritos. Tomando como valoraciones las siguientes:

Cuadro de valores

M (Muy Bueno)

B (Bueno)

R (Regular)

N (No disponible)

Nº	Parámetro de Evaluación	Crystal Reports XI	SAS Web Reports	Oracle Report	Reporting Services	Web Intelligence
1	Facilidad de uso para los usuarios finales	M (3)	M (3)	B (2)	M (3)	M (3)
2	Facilidad de creación de informes	M (3)	B (2)	B (2)	M (3)	M (3)
3	Facilidad de integración a sistemas informáticos	M (3)	R (1)	B (2)	M (3)	M (3)
4	Calendarización de tareas	B (2)	N (0)	M (3)	M (3)	N (0)
5	Interactividad de reportes en web	N (0)	N (0)	R (1)	B (2)	M (3)
6	Nivel de seguridad para visualización de reportes	R (1)	R (1)	M (3)	M (3)	B (2)
7	Manera de acceder a los reportes	B (2)	B (2)	M (3)	M (3)	M (3)

Capítulo III. Estado del Arte Metodológico

8	Enviar notificaciones de error cuando los reportes fallan	N (0)	N (0)	B (2)	N (0)	N (0)
9	Jerarquía de carpetas en reportes	N (0)	N (0)	B (2)	M (3)	M (3)
10	Costo de Licencias	M (1)	B (2)	M (1)	B (2)	B (2)
Puntaje Total		15	11	21	25	22

Tabla III.5 Cuadro Comparativo de Herramientas para realización de reportes disponibles

Para la selección de la herramienta se ha tomado en consideración las siguientes puntuaciones para los parámetros de comparación que se muestran en la Tabla III.4:

Parámetros	Justificación
1-9	Se dan las puntuaciones de M=3, B=2, R=1 y N=0 donde, M tiene el valor más alto debido a que representa la existencia y mayor utilidad para el usuario final.
10	Se dan puntuaciones de M=1, B=2 y R=3 donde, R tiene el valor más alto porque representa el costo más bajo que es la característica que busca la mayor parte de las organizaciones.

Tabla III.6 Justificación de puntuaciones asignadas a los parámetros de selección

Con las puntuaciones descritas anteriormente se obtienen los siguientes resultados:

Crystal Reports XI	15
SAS Web Reports	11
Oracle Reports	21
Reporting Services	25
Web Intelligence	22

Tabla III.7 Selección de Herramienta para elaboración de Reportes

De los resultados mostrados en la Tabla III.6, se concluye que la mejor opción para las necesidades de la organización es Reporting Services.

3.3.2. Herramientas de Datawarehouse

Las herramientas de Datawarehouse las dividimos en los grupos que describimos a continuación para facilitar su entendimiento.

Herramientas de Almacenamiento

Los generadores de reporte tienen sus limitaciones cuando los usuarios finales necesitan más que una vista estática de los datos. Para estos usuarios, las herramientas del procesamiento analítico en línea (OLAP - On Line Analytical Processing), proveen capacidades "Slide and Dice" que contestan a la pregunta "¿qué sucedió?" al analizar por qué los resultados están como están.

Herramientas de Extracción, Transformación y Carga

La Extracción es el primer paso para obtener la información hacia el ambiente del Datawarehouse.

Una vez que la información es extraída hacia el área de tráfico de datos, hay posibles paso de transformación como; limpieza de la información, seleccionar únicamente los campos necesarios para el Datawarehouse, combinar fuentes de datos, haciéndolas coincidir por los valores de las llaves, creando nuevas llaves para cada registro de una dimensión.

Al final del proceso de transformación, los datos están en forma para ser cargados.

Herramientas de Análisis Inteligente:

Estas herramientas han sido construidas utilizando inteligencia artificial que buscan alrededor del Datawarehouse modelos y relaciones en los datos. Estas herramientas utilizan una técnica conocida como Data Mining o Minería de datos.

En todas estas categorías existen productos que integran todos estos procesos, entre los más comunes están:

- **Business Objects, Inc.**

- **OLAP Intelligence**

Posee fácil y rápido acceso a los datos multidimensionales a la medida del negocio. Posee una interfaz intuitiva y altamente funcional. Proporciona fácil y abierto acceso a Servidores OLAP de otros fabricantes, como: Microsoft, Hyperion, entre otros.

Se puede crear flujos de trabajo personalizado para cada área del negocio.

Proporciona publicación fácil y personalizada de la información OLAP a través de la Web.

- **Data Integrator**

Se encarga del proceso de extracción de las diferentes fuentes de datos, la transformación y carga de los mismos en el Datawarehouse.

- **Cognos**

- **Cognos Data Integration**

Esta herramienta permite asistir al usuario en las etapas de extracción transformación y carga de los datos dentro del Datawarehouse. Provee interfaz intuitiva para realizar estos procesos con mayor facilidad.

- **Cognos Bussiness Event Management**

Puede controlar y manejar los eventos a través de cualquier combinación entre fuentes OLAP y fuentes relacionales de datos.

- **Oracle**

- **Oracle Warehouse Builder**

Permite la conexión a diferentes fuentes de Bases de datos. Se puede crear módulos de fuentes que contienen la información de conexiones e información de las tablas.

Utiliza funciones de “Drag and Drop”. Provee un ambiente gráfico del mapeo de los objetos. Puede generar código de extracción, transformación y carga de los datos.

- **Oracle Data Mining**

Provee una interfaz fácil de usar. Permite a los analistas de datos construir modelos predictivos, y generar código JAVA o PL/SQL dependiendo de las necesidades. Puede integrarse fácilmente con aplicaciones para automatizar el proceso de Minado de Datos.

- **Oracle OLAP**

Los resultados son desplegados en tablas en forma gráfica tomando en cuenta la estructura OLAP. Es fácil integración en las consultas con los términos de negocios.

- **Sql Server**

- **Sql Server Integration Services**

Permite la conexión a diferentes fuentes de datos. Utiliza funciones de “Drag and Drop”. Provee un ambiente gráfico para la ejecución de tareas. Provee utilidades para la extracción Transformación y cargas fáciles de utilizar. Los paquetes creados son de fácil instalación, mantenimiento y ejecución.

Se puede agregar tareas calendarizadas para la ejecución de los paquetes DTS. Permite la ejecución de los paquetes desde aplicaciones de consola.

En los procesos de transformación de datos permite utilizar código Visual Basic para incrementar la funcionalidad en la transformación.

- **Sql Server Analysis Services**

Permite la conexión a diferentes fuentes de datos. Provee una interfaz fácil de usar. Los resultados son desplegados en tablas en forma gráfica tomando en cuenta la estructura OLAP.

Es de fácil integración en las consultas con los términos de negocios.

Permite crear formulas personalizadas para determinar los valores de las medidas. Dispone de Asistentes para la creación de cubos multidimensionales.

3.4. SELECCIÓN DE LAS HERRAMIENTAS PARA EL DESARROLLO DE UN DATAMART

La Tabla III.7 muestra cualitativamente las características de cada una de las herramientas de Datawarehouse antes descritas tomando como valoraciones las siguientes:

Cuadro de valores

M (Muy Bueno)

B (Bueno)

R (Regular)

N (No disponible)

N°	Parámetro de Evaluación	Business Objects	Cognos	Oracle	Sql Server
1	Facilidad de uso para los usuarios finales	M (3)	M (3)	B (2)	M (3)
2	Acceso Web	M (3)	B (2)	B (2)	M (3)
3	Facilidad de integración a sistemas informáticos	M (3)	R (1)	B (2)	M (3)
4	Calendarización de tareas	B (2)	N (0)	M (3)	M (3)
5	Lenguaje MDX de consulta claro	N (0)	N (0)	R (1)	B (2)
6	Nivel de seguridad	R (1)	R (1)	M (3)	M (3)
7	Soporte a diferentes fuentes de datos	B (2)	B (2)	M (3)	M (3)
8	Categorización de usuarios	N (0)	N (0)	B (2)	N (0)
9	Acceso remoto seguro	N (0)	N (0)	B (2)	M (3)
10	Ayuda en línea	Si (1)	Si (1)	Si (1)	Si (1)

11	Costo de Licencias	M (1)	M (1)	M (1)	B (2)
Puntaje Total		16	11	22	26

Tabla III.8 Cuadro Comparativo de Herramientas Datawarehouse disponibles en el Mercado

Para la selección de la herramienta se ha tomado en consideración las siguientes puntuaciones para los parámetros de comparación:

Parámetros	Justificación
1-9	Se dan las puntuaciones de M=3, B=2, R=1 y N=0 donde, M tiene el valor más alto debido a que representa la existencia y mayor utilidad para el usuario final.
10	Se dan puntuaciones de Si=1 y No=0, donde Si representa la existencia de esta característica y No la carencia de la misma.
11	Se dan puntuaciones de M=1, B=2 y R=3 donde, R tiene el valor más alto porque representa el costo más bajo que es la característica que busca la mayor parte de las organizaciones.

Tabla III.9 Justificación de puntuaciones asignadas a los parámetros de selección

Con los valores presentados en la tabla anterior se obtienen los siguientes resultados:

Business Objects	16
Cognos	11
Oracle	22
Sql Server	26

Tabla III.10 Selección de Herramienta de Datawarehouse

De los resultados mostrados en la Tabla III.9, se concluye que la mejor opción para las necesidades de la organización es Sql Server.

3.5. HERRAMIENTAS PARA EL DESARROLLO DE UN SITIO WEB

Algunos editores muy conocidos son Macromedia Dreamweaver, NVU, Microsoft Frontpage, Microsoft Publisher y WYSIWYG Web Builder. Son software especializado en la creación de sitios web, contruidos para crear y modificar el código HTML, CSS, PHP, etc. de las páginas web.

Tienen funciones muy útiles para un diseñador web, como el editor gráfico, que te permite crear webs con pocos conocimientos de HTML ya que se asemeja a un procesador de textos. Otras funciones muy interesantes de estos programas son el permitir remplazar código de múltiples páginas a la vez, la comprobación de errores de programación en el código, pre-visualizar las páginas en diferentes navegadores, etc.

Las herramientas para el desarrollo de un Sitio Web que analizaremos son las siguientes:

- **NVU**

Las páginas web son visuales por naturaleza, pero su creación ha implicado siempre la edición de código. Los programas gratuitos que permiten editar páginas web como si fuesen documentos son muy pocos. NVU es uno de ellos.

NVU es un editor visual de páginas web, esto es, permite crear páginas web sin tener que escribir código. La maquetación y el contenido son visibles desde el primer momento como si estuvieras viendo la página en tu navegador. En este sentido, es similar a otros programas como Dreamweaver o Composer.

En NVU, insertar imágenes, crear tablas o modificar estilos CSS son maniobras bastante fáciles. Las páginas se cargan en pestañas en la barra superior, mientras que la inferior contiene accesos rápidos a la vista previa y al código fuente.

Robusto y con un buen abanico de herramientas a disposición, NVU es ideal para modificar rápidamente una página o crear una desde cero. Sin embargo, no esperes la misma potencia de otros programas: NVU es un editor sencillo, no apto para proyectos complejos.

- **Microsoft Front Page**

Al igual que otros productos de Microsoft, FrontPage incluye una serie de plantillas prediseñadas listas para usar si se trata de una pequeña página web personal o sitio grande para negocios.

Sólo se debe de ingresar un poco de texto e imágenes y eso es todo, una de las mejores cosas acerca de FrontPage es que le permite ver las páginas tal y como aparecen en Internet Explorer y otros navegadores. También tiende a funcionar en perfecta armonía con otras tecnologías de Microsoft tales como MS SQL y de bases de datos de MS Access y el lenguaje de programación ASP, FrontPage interactúa fácilmente con otras aplicaciones de Microsoft Office también.

El principal inconveniente de FrontPage es que usa tablas con un ancho fijo por defecto. Esto podría ser frustrante para alguien que planea crear páginas de distintos tamaños.

- **Microsoft Publisher**

El programa de autoedición de publicaciones Microsoft Publisher ofrece las herramientas necesarias para crear materiales de comunicación. Los familiares menús y herramientas Microsoft Office, las plantillas de diseño profesional y sus flexibles opciones de personalización convierten la creación de publicaciones de

calidad profesional para impresión o para la web en una tarea más sencilla que nunca.

Facilita más que nunca la creación, diseño y publicación de materiales de marketing y comunicación con apariencia profesional. Se pueden crear materiales para imprimir, enviar por correo electrónico y usar en la web con la misma interfaz de usuario conocida de otros programas del Microsoft Office System.

Eleva a un nivel superior la creación y publicación de materiales de marketing.

Publisher ayuda a crear fácilmente boletines, folletos, sitios web y catálogos, permitiendo crear, personalizar y publicar publicaciones empresariales únicas con menos trabajo y mejores resultados que con un procesador de textos.

Ayuda también a proyectar una imagen corporativa sofisticada y homogénea en todas sus comunicaciones. Las guías de diseño automatizadas y los más de 40 conjuntos de diseño coordinados ayudan a dar una apariencia homogénea a todos nuestros materiales.

Resulta fácil aprender a usar Publisher, puesto que opera como Microsoft Word incluso a nivel de menús, barras de herramientas y accesos directos.

- **Macromedia Dreamweaver**

Dreamweaver es la herramienta de diseño de páginas web más avanzada, tal como se ha afirmado en muchos medios. Aunque sea un experto programador de HTML el usuario que lo maneje, siempre se encontrarán en este programa razones para utilizarlo, sobre todo en lo que a productividad se refiere.

Cumple perfectamente el objetivo de diseñar páginas con aspecto profesional, y soporta gran cantidad de tecnologías, además muy fáciles de usar:

- Hojas de estilo y capas
- Javascript para crear efectos e interactividades
- Inserción de archivos multimedia.

Además es un programa que se puede actualizar con componentes, que fabrica tanto Macromedia como otras compañías, para realizar otras acciones más avanzadas.

En resumen, el programa es realmente satisfactorio, incluso el código generado es de buena calidad. La única contrariedad consiste en que al ser tan avanzado, puede resultar un poco difícil su manejo para personas menos experimentadas en el diseño de webs.

Dreamweaver ha evolucionado mucho en su versión 4, que incluye soporte para la creación de páginas dinámicas de servidor en ASP, con acceso a bases de datos (versión Ultradev) y una mayor integración con otras herramientas de Macromedia como Fireworks.

• **WYSIWYG Web Builder**

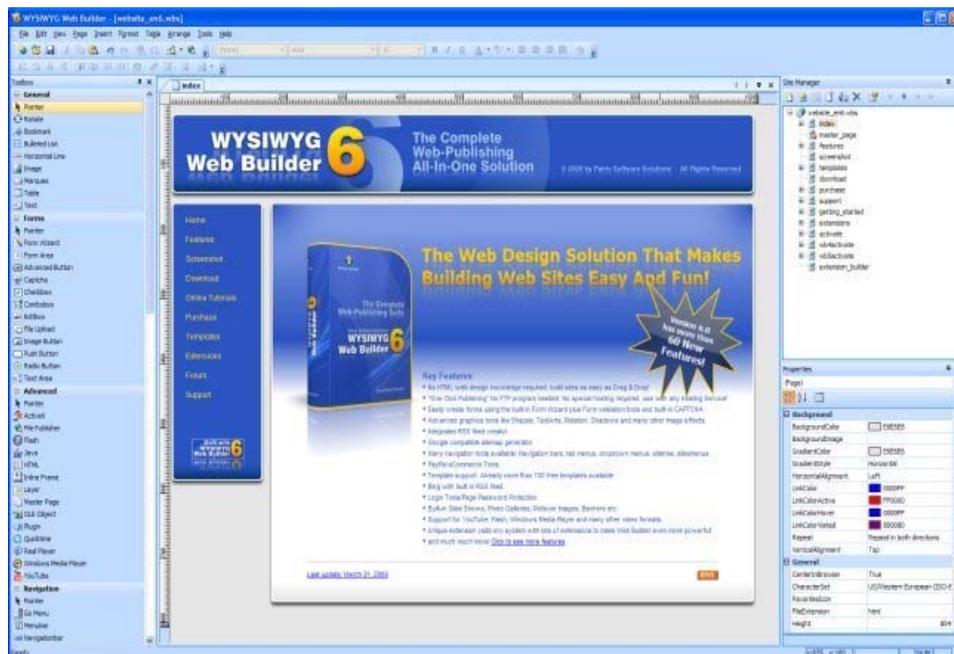
Con esta herramienta se puede crear fácilmente una página web estándar sin necesidad de tener conocimiento alguno de HTML.

Se trata de una utilidad enfocada a los usuarios con menos experiencia en este tipo de tareas, que recuerda mucho al editor de HTML que venía integrado con

Windows (FrontPage Express) y que incluye todas aquellas herramientas que se puedan necesitar en el diseño y publicación de una página web personal.

El programa te da libertad absoluta a la hora de “colocar” en la web los diversos elementos que la integran; sólo tienes que crearlos (área de texto, imágenes, tablas, enlaces, formularios, flash, etc.) y ponerlos donde quieras, distribuyéndolos por la superficie de la web con un simple movimiento de ratón. Los usuarios más avanzados también pueden editar e insertar código HTML cuando lo necesiten.

Se puede introducir además elementos extras como fragmentos de JavaScript listos para usar, imágenes “roll-over” o incluso colorear las barras de desplazamiento del navegador.



3.6. SELECCIÓN DE LAS HERRAMIENTAS PARA EL DESARROLLO DE UN SITIO WEB

La Tabla III.10 muestra cualitativamente las características de cada una de las herramientas web antes descritas tomando como valoraciones las siguientes:

Cuadro de valores

M (Muy Bueno)

B (Bueno)

R (Regular)

N (No disponible)

Nº	Parámetro de Evaluación	Macromedia Dreamweaver	NVU	Microsoft Frontpage	Microsoft Publisher	Web Builder
1	Sistema Web profesional	M(3)	B(2)	R(1)	R(1)	R(1)
2	Acceso en Multiplataforma	B(2)	B(2)	R(1)	R(1)	R(1)
3	Multilinguaje	M(3)	B(2)	R(1)	R(1)	R(1)
4	Vistas de entorno	M(3)	M(3)	R(1)	M(3)	M(3)
5	Funciones de Diseño Web	M(3)	M(3)	R(1)	M(3)	M(3)
6	Nivel de seguridad	B(2)	R(1)	R(1)	R(1)	R(1)
7	Espacio en disco	B(2)	M(2)	R(1)	B(2)	M(3)
8	Consumo de Memoria RAM	B(2)	M(3)	B(2)	R(1)	B(2)
9	Ayuda en línea	Si(1)	No(0)	Si(1)	Si(1)	No(0)
10	Conocimiento previo en HTML	Si(0)	No(1)	No(1)	No(1)	No(1)
11	Costo de Licencias	B(2)	R(3)	B(2)	B(2)	B(2)
Puntaje Total		23	22	13	17	18

Tabla III.11 Cuadro Comparativo de Herramientas para el desarrollo de un Sitio Web

Para la selección de la herramienta se ha tomado en consideración las siguientes puntuaciones para los parámetros de comparación:

Parámetros	Justificación
1-8	Se dan las puntuaciones de M=3, B=2, R=1 y N=0 donde, M tiene el valor más alto debido a que representa la existencia y mayor utilidad para el usuario final.
9	Se dan puntuaciones de Si=1 y No=0, donde Si representa la existencia de esta característica y No la carencia de la misma.
10	Se dan puntuaciones de Si=0 y No=1, donde Si representa la existencia de esta característica y No la carencia de la misma.
11	Se dan puntuaciones de M=1, B=2 y R=3 donde, R tiene el valor más alto porque representa el costo más bajo que es la característica que busca la mayor parte de las organizaciones.

Tabla III.12 Justificación de puntuaciones asignadas a los parámetros de selección

Con los valores presentados en la tabla anterior se obtienen los siguientes resultados:

Macromedia Dreamweaver	23
NVU	22
Microsoft Frontpage	13
Microsoft Publisher	17
SWYSIWYG Web Builder	18

Tabla III.13 Selección de herramienta para el desarrollo de un Sitio Web

De los resultados mostrados en la Tabla III.12, se concluye que la mejor opción para desarrollar un Sitio Web es Macromedia Dreamweaver.

CAPÍTULO IV
IMPLEMENTACIÓN

4.1. PLANEACIÓN

4.1.1. Selección de la estrategia de Implementación

Se seleccionará el enfoque combinado, ya que se cuenta con la colaboración del Área de Sistemas y además se conoce donde se aplicará la tecnología de Datamart, que en este caso será en el Segmento Residencial de Telefónica del Perú.

4.1.2. Selección de la metodología de desarrollo

Se aplicará el método de desarrollo espiral, ya que tiene principal énfasis en la velocidad de culminación del proyecto.

4.1.3. Desarrollo de los objetivos empresariales

A continuación se definen los objetivos de la empresa.

Datamart de Contactos

- Proporcionar a diversas áreas de negocio de Telefónica del Perú un repositorio único de datos que permita manejar de manera centralizada las diversas fuentes de Contactos del Cliente.
- Permitir la generación de Reportes de forma automatizada e intuitiva a través de una herramienta de explotación de datos, eliminando la dependencia con el área de Sistemas para la obtención de los mismos.
- Automatización de los procesos de carga, lo cual permitirá la consolidación periódica de la información que se genera en cada contacto.
- Procesamiento rápido, preciso y confiable de los datos.
- Acoplamiento con las plataformas tanto de software como de hardware disponibles en la organización.

- Almacenamiento de una gran cantidad de información (Gigabytes)

Web de Contactos

- Construcción de una Web de consultas a nivel de teléfono con el que los asesores de Telefónica podrán revisar los contactos con el cliente.
- Acceso eficiente y ágil de la información a los usuarios de Telefónica.
- Disponibilidad de la información las 24 horas del día.

4.1.4. Selección del ámbito de implementación

Para limpiar el alcance de implementación del Datamart se tomaron en cuenta las consideraciones tecnológicas del caso y las dimensiones requeridas por el usuario final del Datamart.

Las consideraciones tecnológicas fueron las siguientes:

- Tamaño o longitud de los datos del Datamart.
- Cantidad de fuentes de datos del Datamart.
- Organización y calidad de los datos de las fuentes de datos del Datamart.
- Documentación de las fuentes de datos del Datamart.
- Manejador de base de datos de las fuentes de datos.
- Disponibilidad de modelos lógicos de las fuentes de datos.
- Recursos disponibles para la implementación del Datamart.

4.1.5. Selección del enfoque arquitectónico

El enfoque arquitectónico utilizado es de Datamart debido a que esta arquitectura está dirigida a una unidad o departamento de la organización (Segmento

Residencial). Con la selección de este enfoque arquitectónico se obtienen las ventajas del Datamart en cuanto a tiempo de implementación, el cual es menor al de un Datawarehouse.

El Datamart construido será de tipo independiente, por cuanto se alimenta de sistemas de información inconexos.

4.1.6. Desarrollo de un programa y presupuesto para planes del proyecto

Estimación de Tiempos del Proyecto

En la Tabla IV.1 se detallan los tiempos estimados para el desarrollo de las distintas fases:

Fase	Tiempo estimado (Días)
Planeación	31 días
Requerimientos	35 días
Análisis	32 días
Diseño	45 días
Construcción y Pruebas	60 días
Despliegue	60 días
Expansión	30 días

Tabla IV.1 Tiempo estimado del Proyecto

Presupuesto del Proyecto

Para la elaboración del presupuesto del proyecto se han tomado en consideración los rubros mostrados en la Tabla IV.2

N°	Elementos	Datamart	Web de Contactos
1	Duración	237 días	56 días
2	Jornadas	160 Jornadas / Hombre	115 Jornadas / Hombre
3	Desviación	+/- 30%	+/- 20%
Costo aproximado		S/. 76,088	S/. 60,000

Tabla IV.2 Presupuesto del Proyecto

4.1.7. Desarrollo de escenarios de uso empresarial

El escenario donde se desenvuelve el Datamart es el siguiente:

- Las unidades de mercado de la corporación son las áreas de negocio y los puntos de contacto con el cliente.
- El personal que labora en estas unidades es quien dará uso a la solución para realizar el estudio de las unidades de clientes residenciales de la corporación y de los productos y consumos que poseen.
- El Datamart ofrece la información de productos y consumos de un cliente Residencial dentro de la corporación. Esta información ha sido recopilada desde los datos suministrados por las filiales² dando la prioridad a la filial que tenga datos de mayor confiabilidad, lo cual no puede ser cubierto por alguno de los sistemas operacionales existentes.

² Filiales: Puntos de contacto con el cliente

4.1.8. Recopilación de los metadatos

La recopilación de los metadatos fue producto de las reuniones de trabajo con los líderes tecnológicos de las filiales, los cuales aportaron la información necesaria acerca de sus sistemas.

4.2. REQUERIMIENTOS

En esta etapa definimos una especificación de las funciones que contendrá el Datamart como son:

4.2.1. Requerimientos del propietario

El propietario, en el proyecto del trabajo, es el que redacta esta tesina, Responsable de Proyectos en Campañas Dirigidas, Bach. Alejandro Gamarra Ramírez, quien estuvo encargado de dirigir al equipo de trabajo del proyecto, mantener en contacto a la gerencia del estado del mismo y proporcionar las facilidades para su desarrollo.

Los requerimientos del propietario se centraron en el concepto del proyecto, los cuales fueron expuestos a través de varias sesiones de trabajo con los arquitectos, desarrolladores y usuarios finales. En estas sesiones de trabajo, el propietario del Datamart hizo énfasis en las funcionalidades que debía brindar la solución y las reglas de negocio que debían cumplirse en su diseño y desarrollo.

Las indicaciones hechas por el propietario del proyecto a los arquitectos y desarrolladores en las sesiones de trabajo estaban orientadas de acuerdo con los siguientes aspectos:

- Contar con información centralizada y depurada: Registrar todas las interacciones entre el cliente y la empresa o viceversa, en un único repositorio (Datamart).
- Facilitar la consulta de información por parte de los diferentes usuarios del sistema.
- Eliminar los tiempos dedicados a la producción de información básica.
- Optimizar los costos de administración de la información.
- Tener un Datamart cuyas fuentes sean 100% ATIS Contactos.³

4.2.2. Requerimientos del arquitecto

El equipo de arquitectos del proyecto está conformado por personal de las áreas de Negocio y Sistemas de Telefónica, los cuales tienen los siguientes roles:

- Roxanna Cochachi, líder de tecnología del proyecto.
- Responsables del Negocio, encargados de la supervisión del diseño y desarrollo realizado por equipo de desarrollo.
- Equipo de Sistemas, desarrolladores del proyecto.

Los arquitectos del proyecto debieron tomar en cuenta los requerimientos del propietario y del usuario final y buscar la forma de solventar las necesidades de información de ambas partes mediante la tecnología de Datamart.

La arquitectura planteada para solventar las necesidades del propietario y de los usuarios finales se compone de varios elementos, los cuales se mencionan a continuación:

³ ATIS Contactos: Universo dentro del Datawarehouse de la empresa destinada a guardar la información de los contactos con el cliente.

Arquitectura de datos

Se refiere al modelo entidad-relación, elementos de datos y relaciones de la base de datos única. Mediante este modelo se da solución a la carga inicial de los datos de las filiales, la generación de Reportes automatizados y la generación de la Web de Contactos.

Arquitectura de aplicación

La arquitectura de aplicación definida por los arquitectos cubre las siguientes áreas:

- Carga de datos de las filiales en un único Repositorio (Datamart).
- Generación de reportes a través de un Datamart de forma automatizada que serán depositados en un servidor con la suficiente capacidad para almacenar la información histórica y la información que se genere a partir de este proceso automatizado.
- Se necesita que ATIS Contactos incorpore dentro de sus módulos los tipos de contacto que hoy no registra (Reclamos, Averías, Ventas por Call Out, Crossselling, E-Mailing, etc).
- Web de Contactos para uso de los asesores comerciales en cada punto de contacto con el cliente.

Arquitectura de tecnología

Constituida por los elementos de hardware y software que se mencionan a continuación:

- Manejador de base de datos relacional.
- Base de datos denominada Base de Datos Única, la cual es la base del Datamart.

- Servidor donde se ejecuta la instancia del manejador de base de datos.
- Servidor web donde se alojará la Web de Contactos.
- Archivos de carga de datos (Scripts de SQL Loader).
- Librería de funciones y procedimientos
- Tecnología para independizar los datos que manejará el Datamart de los utilizados por los sistemas transaccionales, gracias a lo cual se dispondrá de la información en todo momento.

4.2.3. Requerimientos del desarrollador

La tecnología a utilizarse se dividirá de la siguiente manera:

Para la fuente de datos

- SQL Server 2000 como Sistema de Administración de Bases de Datos Relacionales.
- Power Designer, como herramienta para documentación de metadatos.

Para el Datamart

- Data Transformation Services de SQL Server 2000: como herramienta de extracción, transformación y carga de datos.
- Analysis Services de SQL Server 2000: como herramienta de generación de cubos multidimensionales.
- Power Designer, como herramienta para documentación de metadatos.

Para el Usuario Final

- SQL Reporting Services como herramienta OLAP.

- Macromedia Dreamweaver como herramienta de generación del Sitio Web de consultas para los asesores comerciales en las filiales.

Además, la solución se implementará sobre las siguientes plataformas:

- Windows NT 4.0, XP y 2000 para los clientes y Windows Server 2003 y NT 4.0 para los Servidores.

4.2.4. Requerimientos de los usuarios finales

Para satisfacer por completo a los usuarios finales se debe solicitar información referente a requerimientos de consulta y de reportes como se detalla a continuación:

Requerimientos de consultas

- Total de llamadas entrantes o visitas por cada punto de contacto.
- Motivos de llamadas o visitas por cada punto de contacto.
- Ranking de los 10 principales motivos por cada punto de contacto.
- Máximos y mínimos intervalos de tiempo que se registran llamadas entrantes o visitas por cada punto de contacto.
- Recurrencia de motivos y alerta temprana por cada punto de contacto.
- Número de veces que un cliente se pone en contacto con Telefónica al día/mes/año.
- Número de veces que un cliente se comunica por el mismo motivo en cualquiera de los puntos de contacto con Telefónica.
- Historial de los contactos con el cliente disponibles para cada asesor comercial.

Requerimientos de reportes

Los usuarios finales requieren que los reportes posean las siguientes características:

- Acceso rápido
- Visualización gráfica de los datos
- Fácil elaboración de reportes por parte de los usuarios.

4.3. ANÁLISIS

El análisis realizado tuvo cuatro visiones:

4.3.1. Visión de arriba hacia abajo

La visión de arriba hacia abajo buscó la definición de “cliente” en el proyecto, debido a que cada filial tenía una visión diferente del cliente. Mediante este análisis se incorporó los elementos más importantes de cada filial para la determinación del cliente Telefónica. También mediante esta visión se buscó la definición del "producto" dentro de la base de datos única.

4.3.2. Visión de la fuente de datos

La visión de fuente de datos determinó la manera en que se alimentaría el Datamart con los datos de las filiales. Existieron dos métodos considerados para tal fin, el primero consiste en la fusión de las fuentes de datos para así conformar la base de datos única. El segundo consiste en la creación de una base de datos semilla, la cual dispone inicialmente de datos "confiables".

Las consideraciones más relevantes analizadas para realizar una posible fusión de datos fueron:

- Carga inicial más complicada porque tiene rutinas de validación y calidad de datos específicas por cada filial.
- Inicialmente más completa que la semilla porque dispone de la información de todas las filiales.
- Precisión baja y/o poco aceptable de la data inicial.
- Mayor esfuerzo para mantener la calidad de los datos.
- Retardo en el tiempo de entrada a producción (disponibilidad).

Las consideraciones más relevantes para la construcción de la semilla fueron:

- Carga inicial con complicación manejable debido a que tiene las mismas rutinas de validación y calidad de datos para las filiales.
- Inicialmente incompleta pero representativa del universo de clientes Residenciales.
- Precisión suficiente en el inicio.
- Uso de la información inmediata por los usuarios.
- Cruces posteriores para actualizar los datos de la semilla con las filiales.

4.3.3. Visión del Datamart

La visión de Datamart utilizada se centró en disponer de la información que requería el usuario final, por lo que el modelado Snowflake utilizado para reflejar el Datamart fue el que se ajustó con mejor precisión para recoger los requerimientos.

Debido a que la base de datos única de Clientes Residenciales de Telefónica se centra principalmente en dichos clientes, el modelo seleccionado tiene como tabla de hechos los contactos realizados por los clientes en las diferentes filiales. Las dimensiones utilizadas en el modelo son las diferentes perspectivas que desea el

usuario final de los clientes de Telefónica, y las mismas se alimentan de los datos suministrados por las filiales.

4.3.4. Visión de consulta empresarial

La consulta empresarial fue uno de los principales factores que influyeron en el diseño de Datamart, por cuanto refleja las perspectivas del usuario final. En las primeras sesiones del usuario final con el propietario y los arquitectos, con base a las consultas requeridas por el departamento de mercadeo se plantearon cuáles serían las informaciones que estarían contenidas en el Datamart.

4.4. DISEÑO

El Diseño se dividió en las siguientes partes detalladas a continuación:

4.4.1. Diseño de la arquitectura de datos

La arquitectura de datos comprende el diseño de la base de datos única, la cual contiene el Datamart de clientes, para contemplar los campos y tipos de datos que contiene utilizaremos las Tablas que se describe a continuación:

- Datos del Cliente
- Información de Campañas comerciales asignadas
- Información de los contactos con el cliente

- **Datos del Cliente**

Información	Tipo de Dato	Ejemplo
Nombre del Cliente	Varchar(255)	Alejandro Gamarra Ramírez
Segmento del Cliente	Varchar(100)	Platino
Provincia	Varchar(255)	Lima
Departamento	Varchar(255)	Lima
Distrito	Varchar(255)	Santiago de Surco
Tipo de línea	Varchar(150)	Tarifa Plana Local
Tipo de Speedy	Varchar(150)	Speedy Convencional 500
Tipo de Cable	Varchar(150)	Estándar
Paquetización	Varchar(150)	Paq. Dúo Plano 500
Top Residencial	Varchar(2)	No
Robinson	Varchar(2)	No

Tabla IV.3 Diseño de la Tabla de Datos del Cliente

- **Información de Campañas comerciales asignadas**

Información	Tipo de Dato	Ejemplo
Tipo de envío	Varchar(150)	E-Mailing
Campaña	Varchar(150)	Newsletter Telefónica
Acción	Varchar(150)	Informar a los clientes de las novedades en Telefónica
Destino	Varchar(100)	Terra
Oferta	Varchar(100)	Inscribirse en el servicio
Fecha de envío	Datetime	2008-02-19

Tabla IV.4 Diseño de la Tabla de Campañas asignadas

- **Información de los contactos con el cliente**

Información	Tipo de Dato	Ejemplo
Fuente	Varchar(150)	Averías
Medio	Varchar(150)	Cable
Sub Medio	Varchar(150)	
Fecha	Datetime	2010-01-26
Hora	Datetime	08:26:25
Tipo Contacto	Varchar(150)	Llamada telefónica
Motivo1	Varchar(255)	Avería
Motivo2	Varchar(255)	
Observaciones	Varchar(255)	No cuenta con señal

Tabla IV.5 Diseño de la Tabla de contactos con el cliente

4.4.2. Diseño de la arquitectura de la aplicación

La arquitectura de la aplicación estará dividida en 2 etapas:

- En la primera fase se comenzará a trabajar con FTP para la transferencia de información de las fuentes (104, MC, Out, Cross Selling, Reclamos y Averías, Cobros, On Line, E-Mailing y encartes, IVR, Gestión de Cobranza, Back Office, Eventos especiales) al Servidor del Datamart.

La extracción, carga y transformación se hará con el Data Transformation Services (DTS), ya que nos permitirán extraer, transformar y consolidar datos de fuentes disímiles a la base de datos SQL Server, tal como se puede apreciar en la Figura IV.1 que se muestra a continuación:

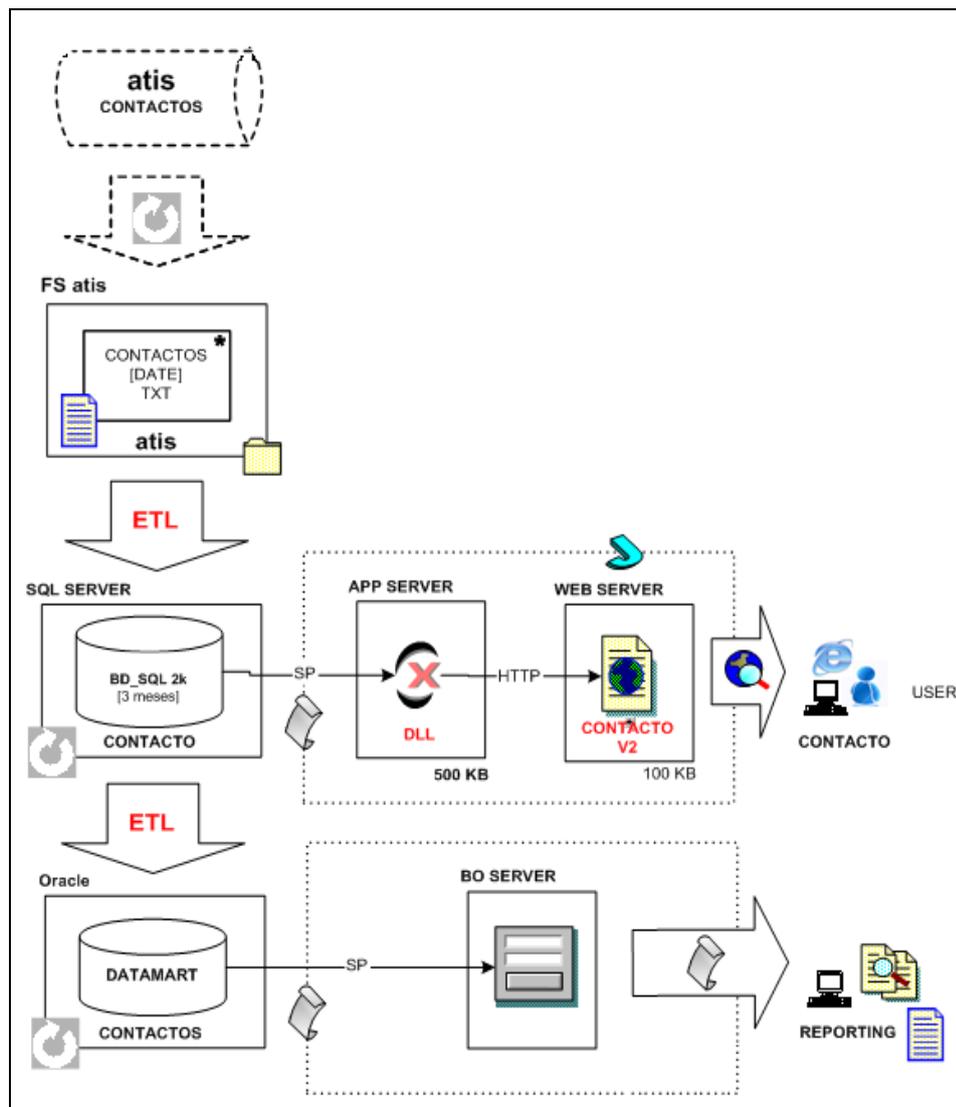


Figura IV.2 Diagrama de la arquitectura de aplicación
Fase 2

4.5. CONSTRUCCIÓN

La fase de construcción será realizada por los desarrolladores mediante la metodología de espiral. Para ello se seguirán los diseños de la fase anterior y los mismos serán implementados en los lenguajes y con las herramientas descritas anteriormente.

El desarrollo de las aplicaciones se llevará a cabo mediante fases, debido al presupuesto asignado y a la posibilidad de incrementar nuevos requerimientos durante la primera fase.

Para temas académicos se realizó una simulación de la construcción del proyecto según los diseños definidos, a menor escala, para poder contemplar el desarrollo de esta etapa, a continuación los detalles de esta simulación.

La construcción contempló la realización de los siguientes programas:

- Elaboración de la base de datos única: Se realizó siguiendo el diseño especificado. Para ello se creó el esquema de la base de datos llamado GCS_PERU, luego se crearon los tablespaces necesarios en la base de datos para posteriormente crear y ejecutar los scripts de creación de las tablas, índices, constraints, secuencias y demás objetos necesarios.
- Archivos de carga de datos de SQL Loader: Fueron elaborados para realizar la extracción de los datos de las filiales contenidos en archivos de texto plano, excel, mdf, mdb, sql, etc. Para llevar a cabo esta extracción se crearon dos archivos por cada tabla de carga, los cuales son:
 - Archivo Temporal de Carga: contiene la definición de los campos del archivo origen a extraer, de igual manera contiene instrucciones para el formato y filtrado de los datos.
 - Archivo de Carga Validado: contiene la definición de los campos de la base de datos destino con la información validada del archivo temporal de carga inicial. Almacena también información histórica y su información no será borrada automáticamente hasta que complete el año de información.

Para realizar la carga se debe ejecutar SQL Loader y colocar el archivo de parámetros, del cual deberá leer los parámetros.

- Procesos de comparación y análisis de clientes en SQL: los procesos ETL que realizan la comparación de las unidades se elaboraron en SQL siguiendo el diseño especificado. Los procedimientos, paquetes y funciones que realizan estos procesos se codificaron y probaron de forma independiente, cada uno como un componente único para luego realizar la integración de los mismos y armar el proceso completo.
- Finalmente, se utilizó la herramienta generadora de páginas web para crear el Sitio Web con código asp vbscript, el cual se alimenta de la información del Datamart de Contactos (Tabla resultante del proceso en el punto anterior) para que el usuario final realice las consultas por teléfono de los contactos en los últimos 3 meses de los clientes.
- Esta Tabla resultante será finalmente cargada en el Datawarehouse de la empresa considerada como el Datamart de Contactos de Telefónica.

4.5.1. Modelo de Base de Datos

El modelo de Base de Datos que se utilizó en la construcción del Simulador del proyecto, el cual será replicado a mayor escala cuando se inicie con la construcción del mismo es el siguiente:

Capítulo IV. Implementación

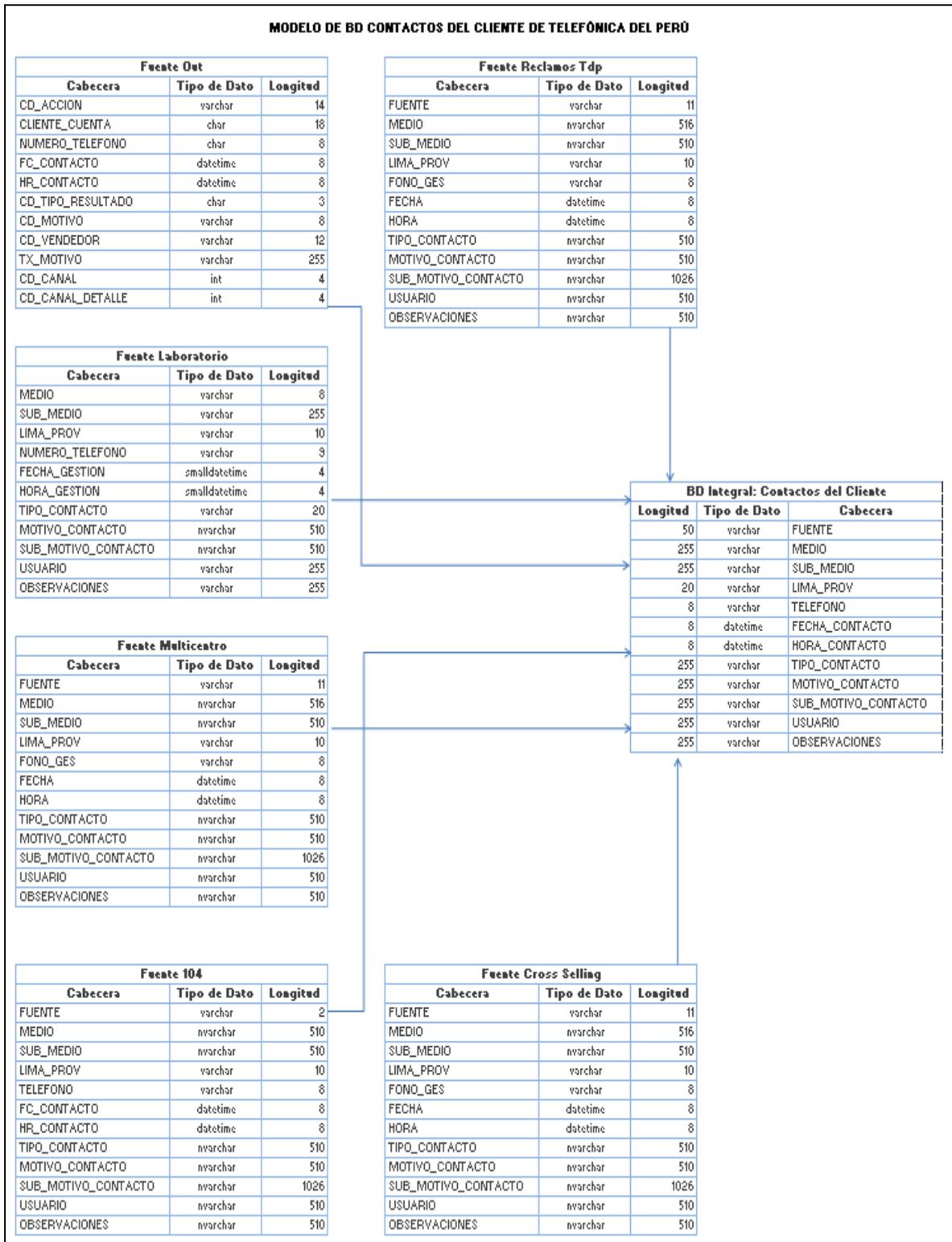


Figura IV.3 Modelo de Base de Datos del Simulador del Proyecto
Parte - I

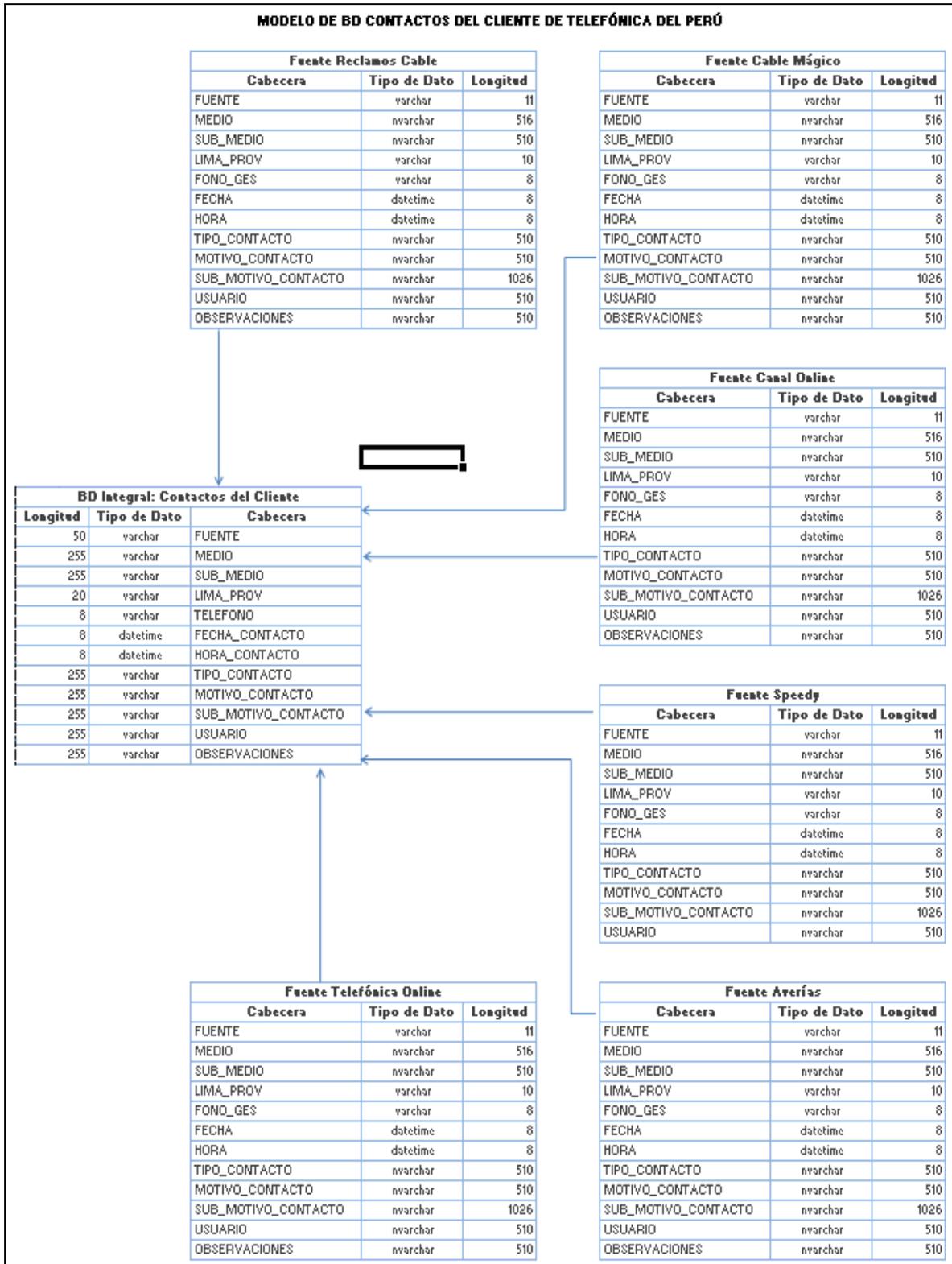


Figura IV.4 Modelo de Base de Datos del Simulador del Proyecto
Parte - II

4.5.2. Paquetes de Transformación

Para la construcción del Simulador del proyecto, se realizó una serie de consultas SQL reunidas en paquetes SQL que se encargan de extraer la información de cada una de las fuentes de contactos con el cliente, una vez que cada uno de estos paquetes culminan con la extracción total de la información necesaria, se ejecuta un paquete integrador final que es el responsable de generar la tabla resultante que servirá tanto para las consultas de los usuarios como para la exportación dentro del Datawarehouse de la empresa.

Los paquetes y consultas que se utilizaron durante la simulación son los siguientes:

- DTS Canal Averías
- DTS Canal Cable
- DTS Canal Online
- DTS Canal Cross Selling
- DTS Canal Multicentro
- DTS Canal Out
- DTS Canal Reclamos Cable
- DTS Canal Reclamos Tdp
- DTS Canal Speedy
- DTS Canal Telefónica Online
- DTS Canal Laboratorio
- DTS Integrar Contactos Cliente

- **DTS Canal 104**

Encargado de extraer la información de la Fuente del 104

Consulta SQL:

```
select distinct 'in' as fuente, lower(gestion) as medio, lower(negocio) as sub_medio,  
case when left(telefono,1)=1 then 'lima'  
else 'provincias'  
end as lima_prov,  
telefono,fc_contacto,hr_contacto,lower(accion) as tipo_contacto, lower(tematico) as  
motivo_contacto,  
(lower(producto) + ' - ' + lower(producto_detalle)) as sub_motivo_contacto,  
lower(nombre_rt) as usuario, lower(observacion) as observaciones  
from ag_in_feedback  
where fc_contacto >= (getdate() - 100)
```

Paquete SQL:

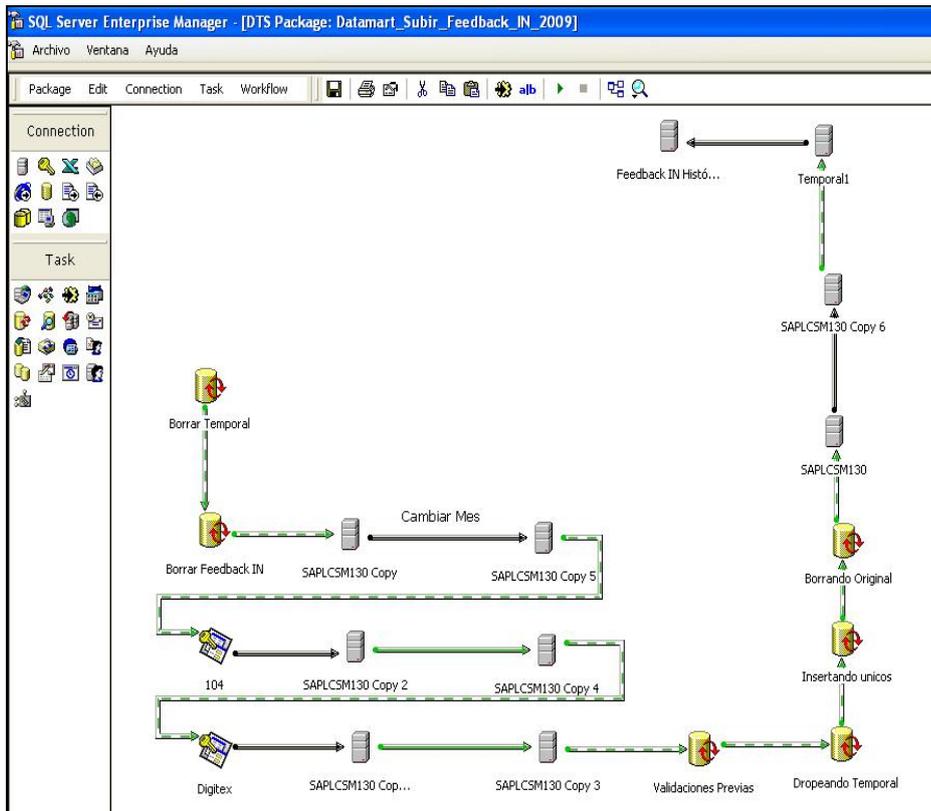


Figura IV.5 DTS de Carga de la Fuente del 104

- **DTS Canal Averías**

Encargado de extraer la información de la Fuente de Averías

Consulta SQL:

```
select distinct 'averias' as fuente, origen as medio, lower(medio_contacto) as sub_medio,  
case when left(telefono,1)=1 then 'lima'  
else 'provincias'  
end as lima_prov,  
telefono,fecha as fc_contacto," as hr_contacto,lower(tipo_averia) as tipo_contacto,  
lower(motivo) as motivo_contacto,  
lower(motivo2 + ' ' + motivo3) as sub_motivo_contacto,  
(estado + ' ' + convert(varchar(8),fecha_fin,112)) as usuario, lower(observaciones) as  
observaciones  
from ag_averias_feedback_bck_3  
where fecha >= (getdate() - 100)  
and telefono <> "
```

Paquete SQL:

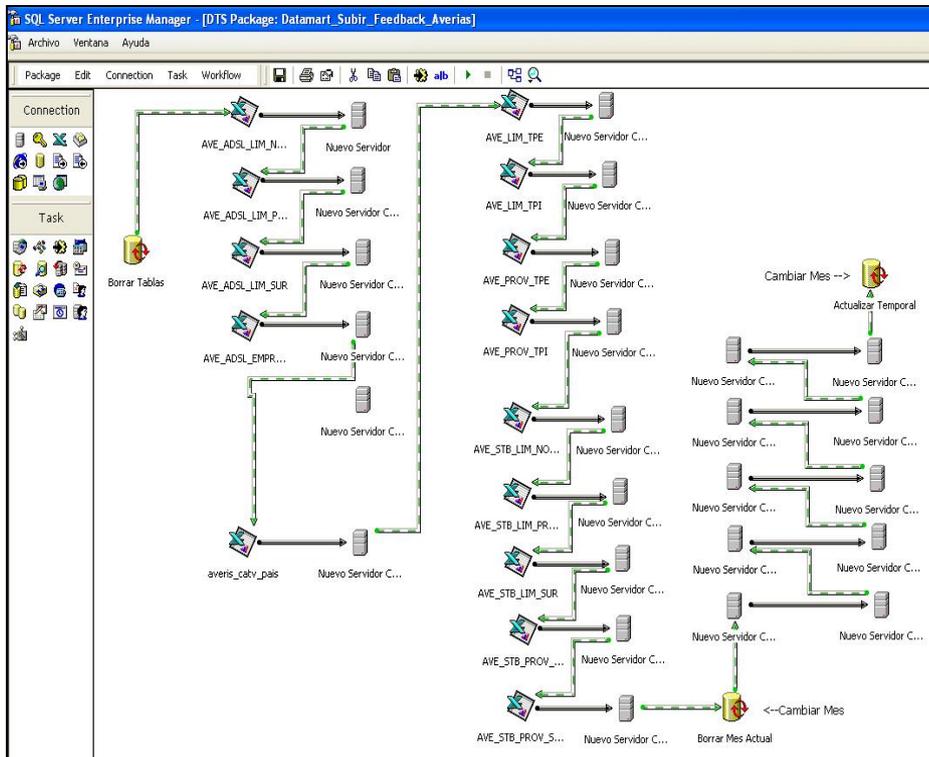


Figura IV.6 DTS de Carga de la Fuente de Averías

- **DTS Canal Cable**

Encargado de extraer la información de la Fuente de Cable Mágico

Consulta SQL:

```
select distinct 'cable magico' as fuente, lower(operador) as medio, 'dato no disponible' as sub_medio,  
    case when left(telefono,1)=1 then 'lima'  
        else 'provincias'  
    end as lima_prov,  
    telefono, fecha as fc_contacto, hora as hr_contacto, lower('llamada telefónica') as  
tipo_contacto, lower(tema_descrip) as motivo_contacto,  
    'dato no disponible' as sub_motivo_contacto,  
    'dato no disponible' as usuario, 'dato no disponible' as observaciones  
from ag_cms_feedback  
where fecha >= (getdate() - 100)
```

Paquete SQL:

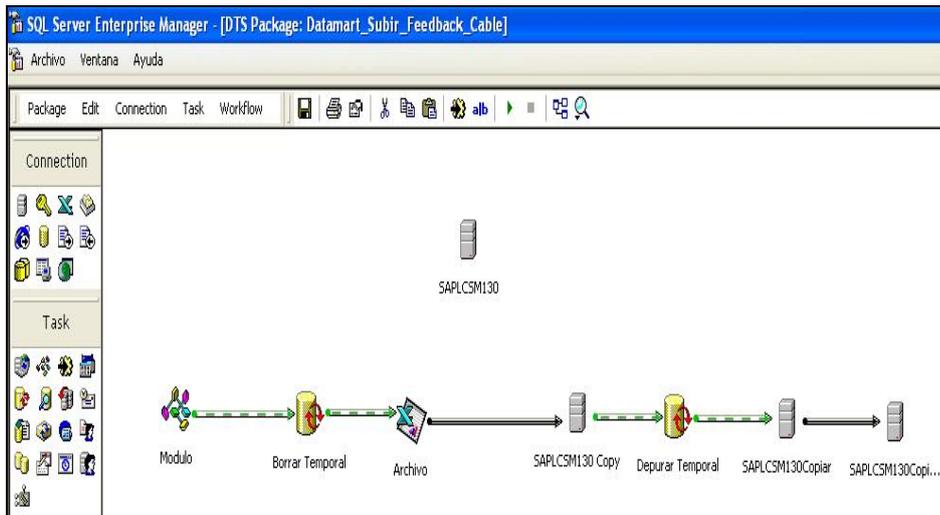


Figura IV.7 DTS de Carga de la Fuente de Cable Mágico

- **DTS Canal Online**

Encargado de extraer la información de la Fuente del Canal Online.

Consulta SQL:

```
select distinct 'canal online' as fuente, 'página web de telefónica' as medio, '' as sub_medio,
    case when left(telefono,1)=1 then 'lima'
        else 'provincias'
    end as lima_prov,
    telefono, fecha_consulta as fc_contacto, '' as hr_contacto, lower(producto_consulta) as
tipo_contacto, lower(consulta) as motivo_contacto,
    lower(resultado) as sub_motivo_contacto,
    '' as usuario, lower(observaciones) as observaciones
from ag_canal_online_feedback
where fecha_consulta >= (getdate() - 100)
```

Paquete SQL:

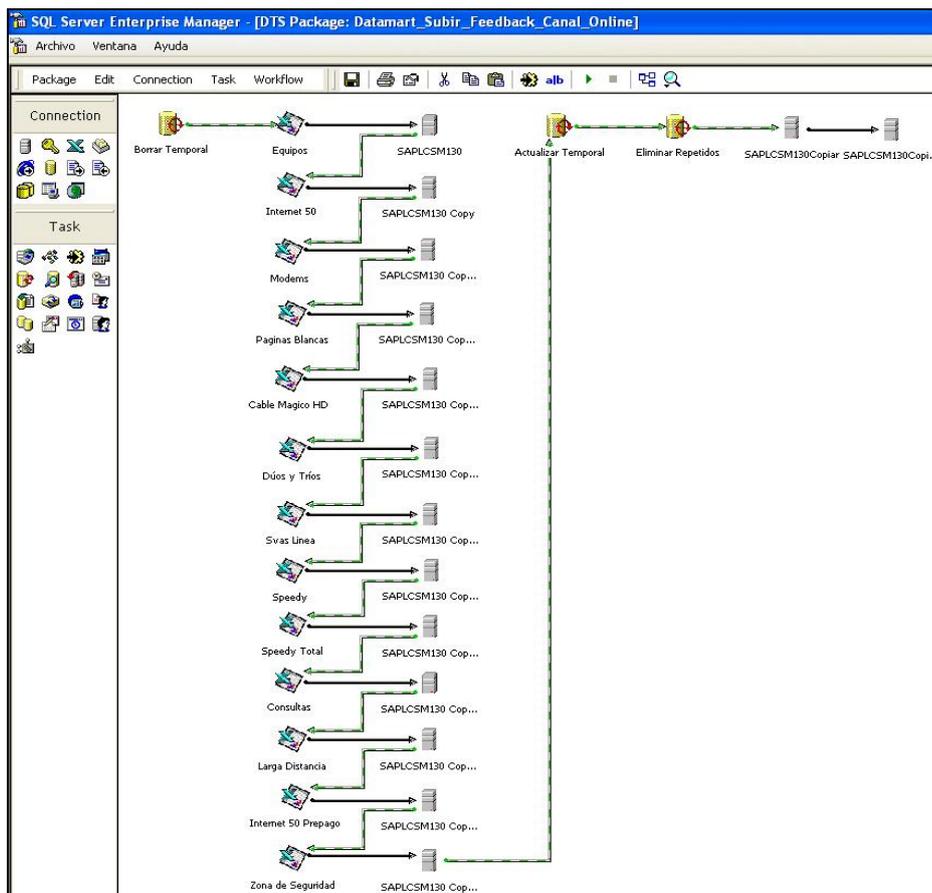


Figura IV.8 DTS de Carga de la Fuente del Canal Online

- **DTS Canal Cross Selling**

Encargado de extraer la información de la Fuente del Cross Selling.

Consulta SQL:

```
select distinct 'cross selling' as fuente, lower(motivo_llamada) as medio, sugerido as sub_medio,  
    case when left(telefono,1)=1 then 'lima'  
        else 'provincias'  
    end as lima_prov,  
    telefono, fecha as fc_contacto, " as hr_contacto, lower(prod_sugerido) as tipo_contacto,  
    lower(resultado) as motivo_contacto,  
    lower(motivo_resultado) as sub_motivo_contacto,  
    " as usuario, lower(observaciones) as observaciones  
from ag_cross_feedback_temp_4  
where fecha >= (getdate() - 100)
```

Paquete SQL:

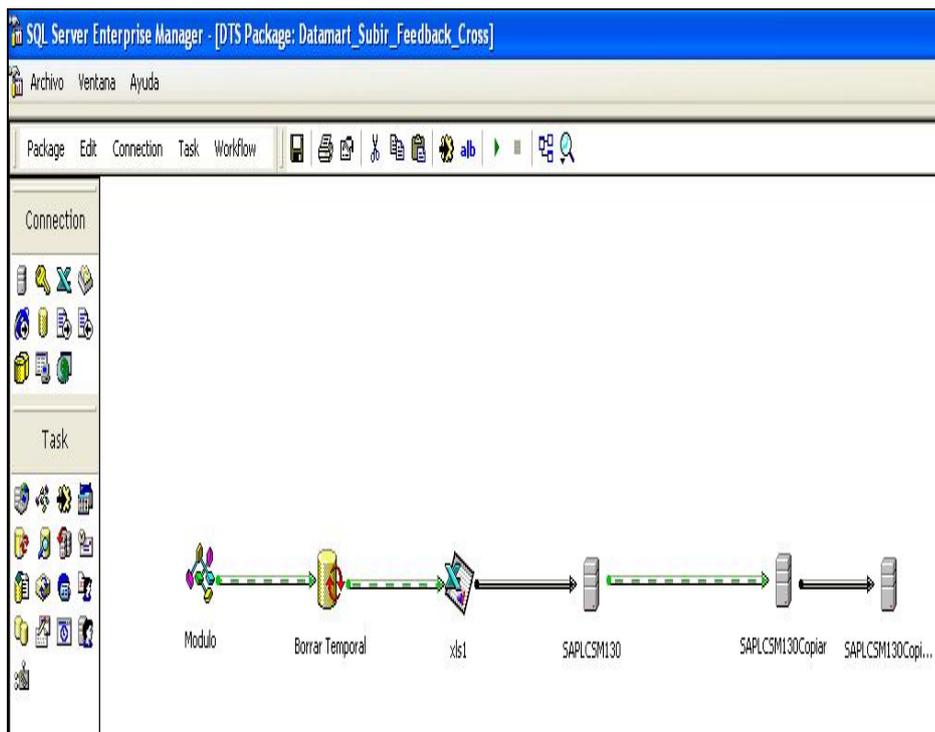


Figura IV.9 DTS de Carga de la Fuente del Cross Selling

- **DTS Canal Multicentro**

Encargado de extraer la información de la Fuente del Multicentro.

Consulta SQL:

```
select distinct 'multicentro' as fuente, ('mc ' + lower(mc)) as medio, lower(medio) as sub_medio,  
    case when left(fono_ges,1)=1 then 'lima'  
        else 'provincias'  
    end as lima_prov,  
    fono_ges,fecha,hora, lower(tipotra) as tipo_contacto,  
    lower(unidtra) as motivo_contacto,  
    (lower(clastra) + ' - ' + lower(accitra)) as sub_motivo_contacto,  
    lower(nomusu) as usuario, lower(obsge) as observaciones  
from ag_multicentro_feedback  
where fecha >= (getdate() - 100)
```

Paquete SQL:

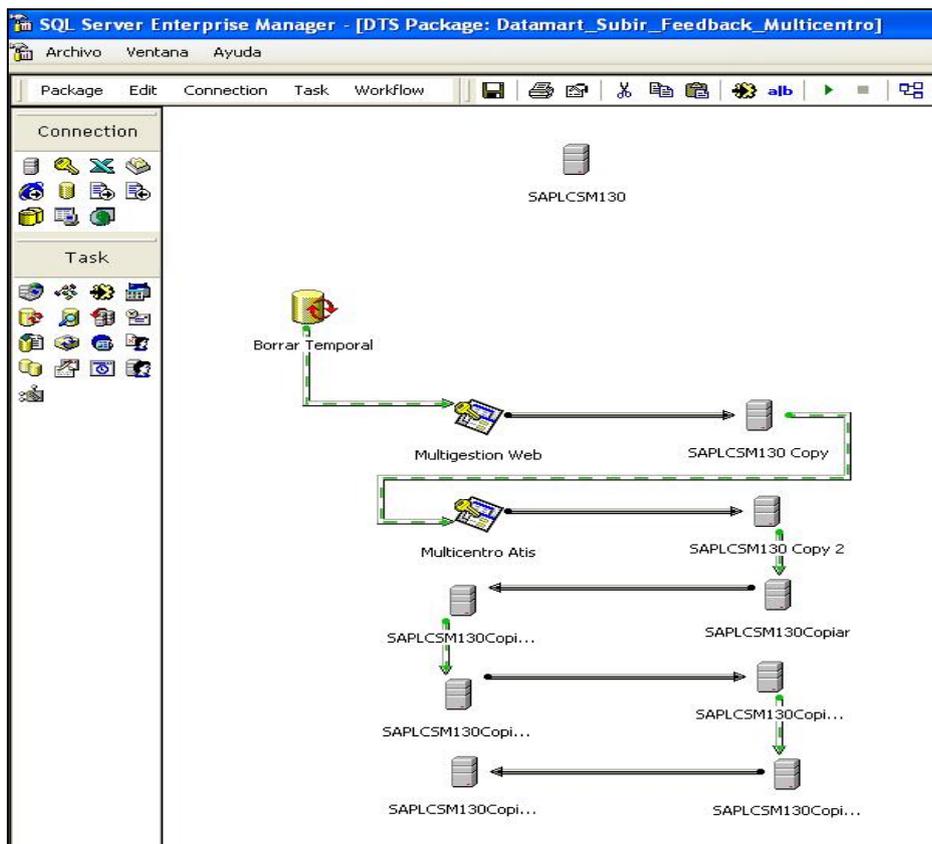


Figura IV.10 DTS de Carga de la Fuente del Multicentro

- **DTS Canal Out**

Encargado de extraer la información de la Fuente del Canal Out.

Consulta SQL:

```
select distinct 'out' as fuente,  
       case when cn.tx_detalle='hf' then 'succesful'  
       else lower(cn.tx_detalle)  
       end as medio,  
       lower(c.tx_campania) as sub_medio,  
       case when left(e.numero_telefono,1)='1' then 'lima'  
       else 'provincias'  
       end as lima_prov,e.numero_telefono,e.fc_contacto,e.hr_contacto,  
       case when e.cd_tipo_resultado in ('cef','cep') then 'contacto efectivo'  
       when e.cd_tipo_resultado='cen' then 'contacto efectivo negativo'  
       when e.cd_tipo_resultado='cne' then 'contacto no efectivo'  
       when e.cd_tipo_resultado='noc' then 'no contacto'  
       else 'otros'  
       end as tipo_contacto, lower(m.tx_motivo) as motivo_contacto, " as sub_motivo_contacto,  
       lower(e.cd_vendedor) as usuario, lower(e.tx_motivo) as observaciones  
from dbo.ag_resultado_2008_temporal e  
inner join canal_detalle cn on e.cd_canal=cn.cd_canal and  
e.cd_canal_detalle=cn.cd_canal_detalle  
inner join campania c on left(e.cd_accion,10)=c.cd_campania  
inner join motivos m on e.cd_motivo=m.cd_motivo  
where e.cd_canal is not null  
and e.cd_canal_detalle is not null
```

Paquete SQL:

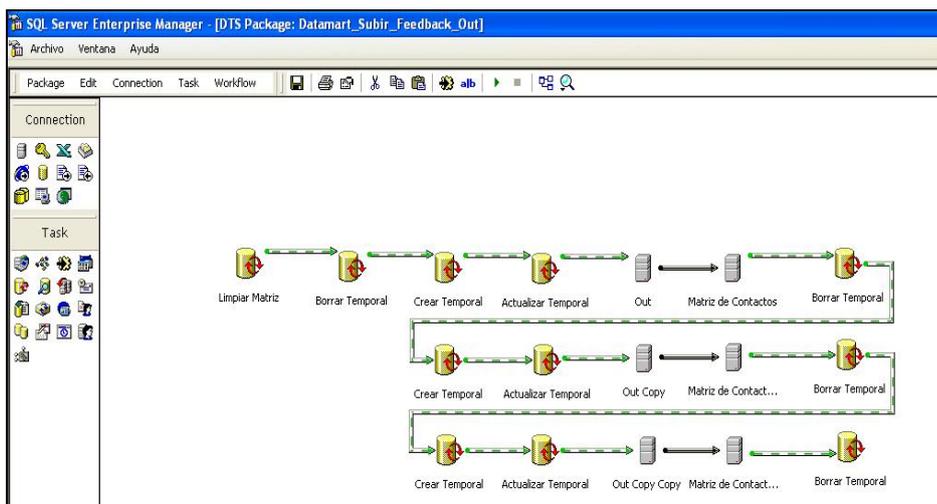


Figura IV.11 DTS de Carga de la Fuente del Canal Out

- **DTS Canal Reclamos Cable**

Encargado de extraer la información de la Fuente de Reclamos de Cable Mágico.

Consulta SQL:

```
select distinct 'reclamos' as fuente,lower(oficina) as medio, lower(tratamiento) as sub_medio,  
  case when left(telefono,1)=1 then 'lima'  
    else 'provincias'  
  end as lima_prov,  
  telefono,fecha as reclamo_fe,hora as hora_contacto,lower(resultado) as tipo_contacto,  
  lower(desmotv) as motivo_contacto,  
  lower(desobs) as sub_motivo_contacto, lower(codusrgen) as usuario, lower(desobs2) as  
  observaciones  
from ag_reclamos_cms
```

Paquete SQL:

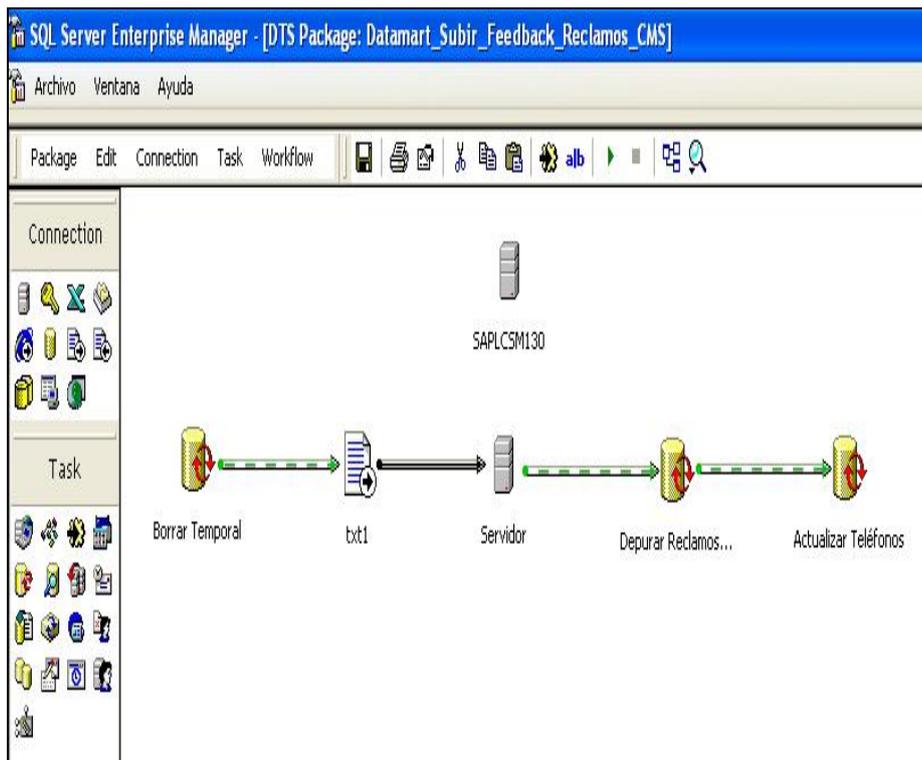


Figura IV.12 DTS de Carga de la Fuente de Reclamos de Cable Mágico

- **DTS Canal Reclamos Tdp**

Encargado de extraer la información de la Fuente de Reclamos de Tdp

Consulta SQL:

```
select distinct 'reclamos' as fuente,lower(oficina_co) as medio, lower(medio_util) as sub_medio,  
  case when left(telefono,1)=1 then 'lima'  
        else 'provincias'  
  end as lima_prov,  
  telefono,reclamo_fe," as hora_contacto,lower(estado_des) as tipo_contacto,  
  lower(reclamo_mo) as motivo_contacto,  
  '.' as sub_motivo_contacto, lower(usuario_ap) as usuario, lower(observacio) as  
observaciones  
from ag_reclamos_feedback  
where reclamo_fe >= (getdate() - 100)
```

Paquete SQL:

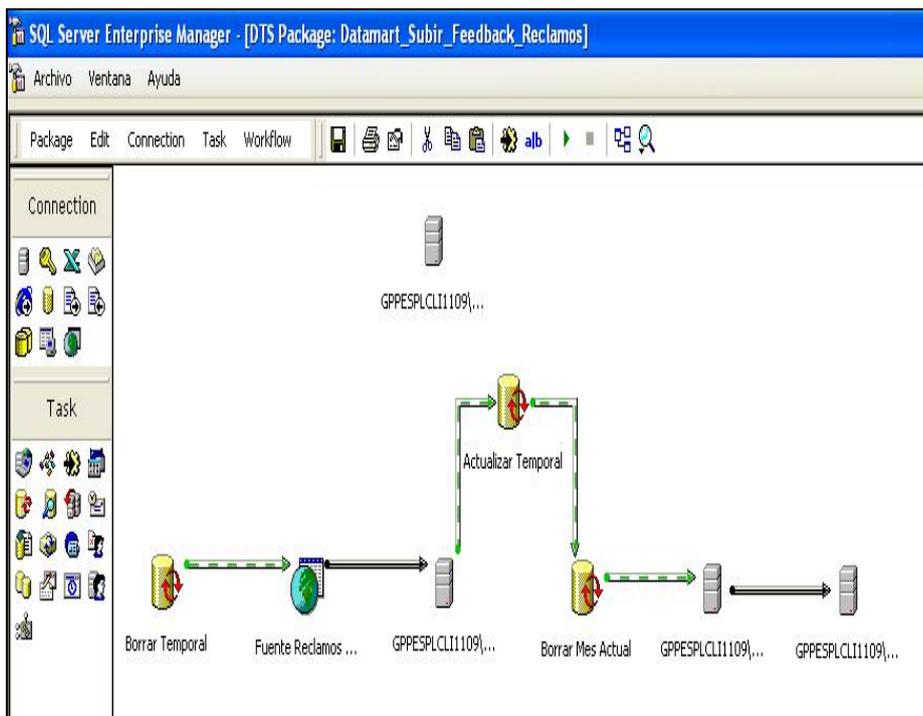


Figura IV.13 DTS de Carga de la Fuente de Reclamos Tdp

- **DTS Canal Speedy**

Encargado de extraer la información de la Fuente del Canal Speedy

Consulta SQL:

```
select distinct
telefono_entrante,telefono_speedy,cli_at_nombre_cliente,telefono_contacto,nombre_contacto,cli
_ult_ncalls,cli_at_email_1,cli_at_email_2,dominio,tipo_cliente,modalidad,equipo,[tipo
speedy]'tipo speedy',[sistema operativo]'sistema
operativo',cli_lugar_speedy,convert(varchar(8),cast(cnt_fini_video as
datetime),112)'cnt_fini_video',cnt_ffin_video,cnt_duracion,cnt_cpsa_puesto,grupo_atencion,cnt_
operador_id,oper_nombre,servicio_in,ges_fini_video,ges_ffin_video,ges_result_descripcion,ges
_gestion,ges_negocio,ges_producto,ges_servicio,ges_accion,codigo_tematico,ges_observacion,c
ampana,cd_incidencia_sigas_grupo,cd_incidencia_sigas_det,[cd_tipo de atencion ras]'cd_tipo de
atencion ras',[cd_afiliacion on line]'cd_afiliacion on
line',cd_scr,cd_asistente,cd_cros_codigo_sp,fecha,hora,segmento
from ag_vus_feedback
where convert(varchar(6),cast(cnt_fini_video as datetime),112) >= (getdate() - 100)
```

Paquete SQL:

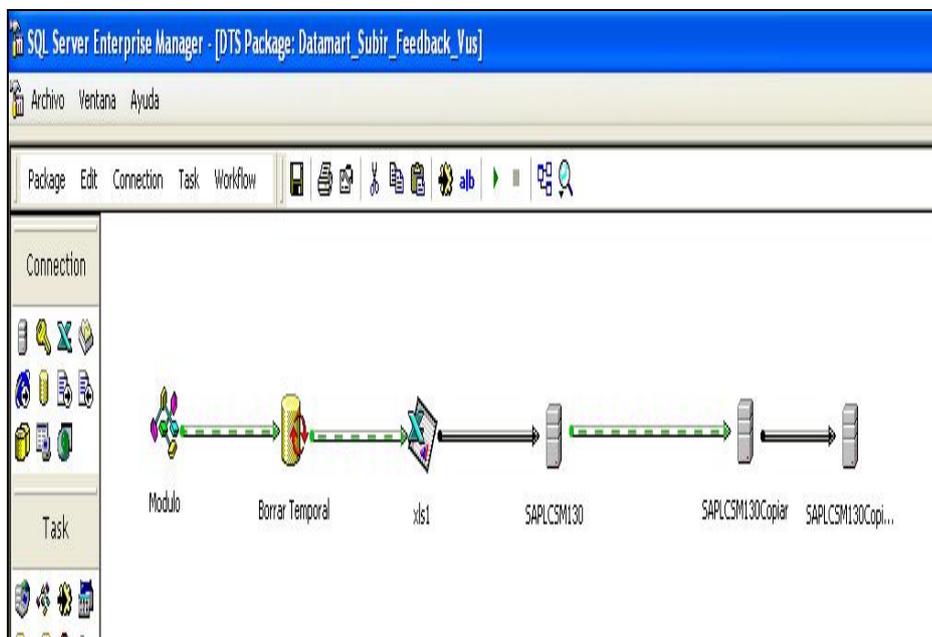


Figura IV.14 DTS de Carga de la Fuente del Canal Speedy

- **DTS Canal Telefónica Online**

Encargado de extraer la información de la Fuente de Telefónica Online

Consulta SQL:

```
select distinct 'canal online' as fuente, 'correo electrónico' as medio, 'revisó el correo' as sub_medio,
    case when left(telefono,1)=1 then 'lima'
        else 'provincias'
    end as lima_prov,
    telefono,fecha as fc_contacto," as hr_contacto,lower('correo electrónico') as tipo_contacto,
lower(producto) as motivo_contacto,
    " as sub_motivo_contacto,
    " as usuario," as observaciones
from ag_feedback_online
where fecha >= (getdate() - 100)
```

Paquete SQL:

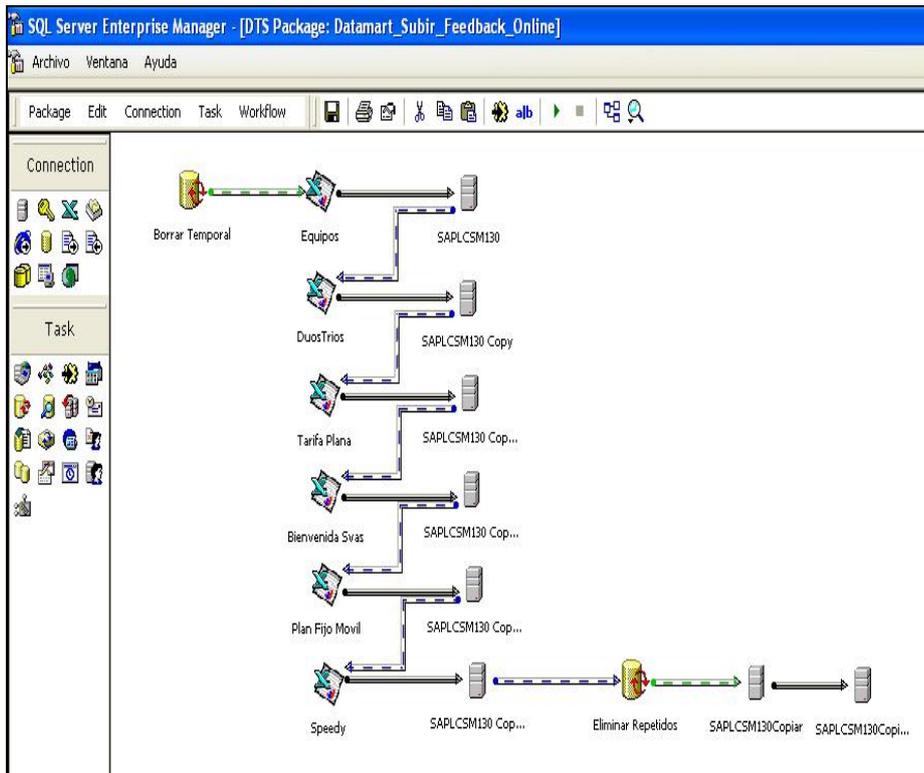


Figura IV.15 DTS de Carga de la Fuente de Telefónica Online

- **DTS Canal Laboratorio**

Encargado de extraer la información de la Fuente del Laboratorio.

Consulta SQL:

```
select distinct 'laboratorio' as fuente, 'encuesta' as medio, lower(l.tx_encuesta) as sub_medio,  
  case when left(l.numero_telefono,1)='1' then 'lima'  
    else 'provincias'  
  end as lima_prov,l.numero_telefono,l.fecha_gestion,l.hora_gestion,  
  case when m.tx_tipo_resultado='cef' then 'contacto efectivo'  
    when m.tx_tipo_resultado='cne' then 'contacto no efectivo'  
    when m.tx_tipo_resultado='noc' then 'no contacto'  
    else 'otros'  
  end as tipo_contacto, lower(m.tx2) as motivo_contacto, lower(m.tx_motivo) as  
sub_motivo_contacto,  
  lower(l.codempleado) as usuario, lower(l.observaciones) as observaciones  
from laboratorio_encuesta_historico l  
inner join motivos m on l.cd_motivo=m.cd_motivo  
where l.fecha_gestion >= (getdate() - 100)  
and m.tx_tipo_resultado in ('cef','cne','noc')
```

Paquete SQL:

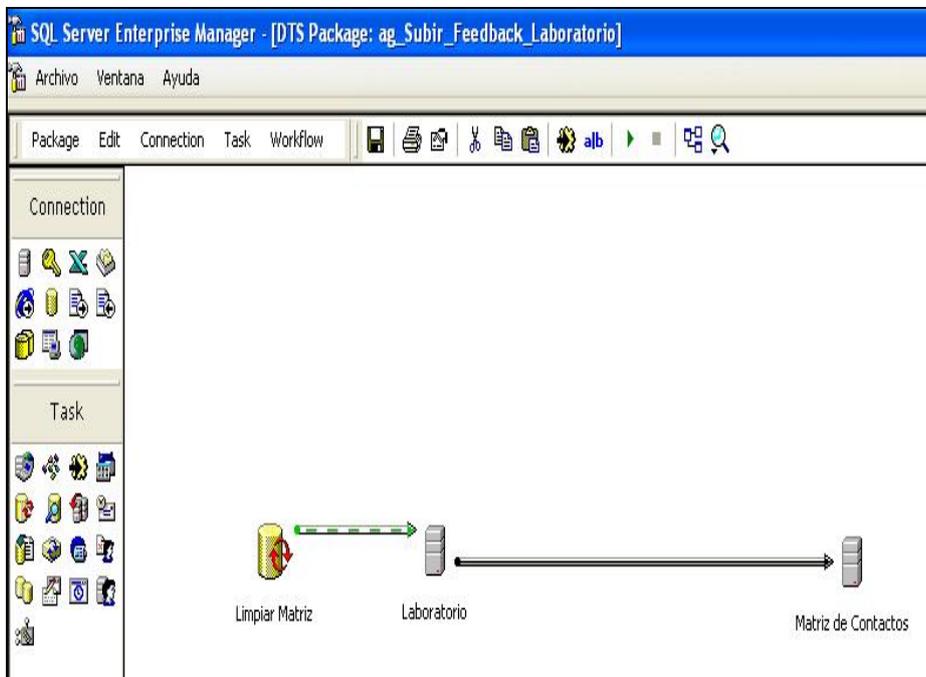


Figura IV.16 DTS de Carga de la Fuente del Laboratorio

- **DTS Integrar Contactos Cliente**

Encargado de integrar la información de las diferentes fuentes de contactos con el cliente, que han sido cargadas por cada uno de los DTS descritos previamente. De la tabla resultante en este paso, se realizarán las consultas y se generarán los reportes y dashboards que utilizarán los usuarios del Sistema.

Tabla Resultante:

AG_CONTACTOS_CLIENTE_FEEDBACK

Consulta SQL:

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[AG_CONTACTOS_CLIENTE_REPORTE0]') and OBJECTPROPERTY(id,
N'IsUserTable') = 1)
drop table [dbo].[AG_CONTACTOS_CLIENTE_REPORTE0]
GO
```

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[AG_CONTACTOS_CLIENTE_REPORTE1]') and OBJECTPROPERTY(id,
N'IsUserTable') = 1)
drop table [dbo].[AG_CONTACTOS_CLIENTE_REPORTE1]
GO
```

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[AG_CONTACTOS_CLIENTE_REPORTE2]') and OBJECTPROPERTY(id,
N'IsUserTable') = 1)
drop table [dbo].[AG_CONTACTOS_CLIENTE_REPORTE2]
GO
```

```
if exists (select * from dbo.sysobjects where id =
object_id(N'[dbo].[AG_CONTACTOS_CLIENTE_REPORTE3]') and OBJECTPROPERTY(id,
N'IsUserTable') = 1)
drop table [dbo].[AG_CONTACTOS_CLIENTE_REPORTE3]
GO
```

```
select fuente,convert(varchar(6),fecha_contacto,112)'mes',count(*)'total'
into ag_contactos_cliente_reporte0
from ag_contactos_cliente_feedback
group by fuente,convert(varchar(6),fecha_contacto,112)
order by 1
```

```
select fuente,sum(total)'total',(sum(total)/90)'dia'
into ag_contactos_cliente_reporte1
from ag_contactos_cliente_reporte0
group by fuente
order by 1
```

```
select motivo_contacto,convert(varchar(6),fecha_contacto,112)'mes',count(*)'total'
into ag_contactos_cliente_reporte3
from ag_contactos_cliente_feedback
where fuente='reclamos'
group by motivo_contacto,convert(varchar(6),fecha_contacto,112)
```

```
select motivo_contacto,sum(total)'total',(sum(total)/90)'dia'
into ag_contactos_cliente_reporte2
from ag_contactos_cliente_reporte3
group by motivo_contacto
order by 2 desc
```

Paquete SQL:

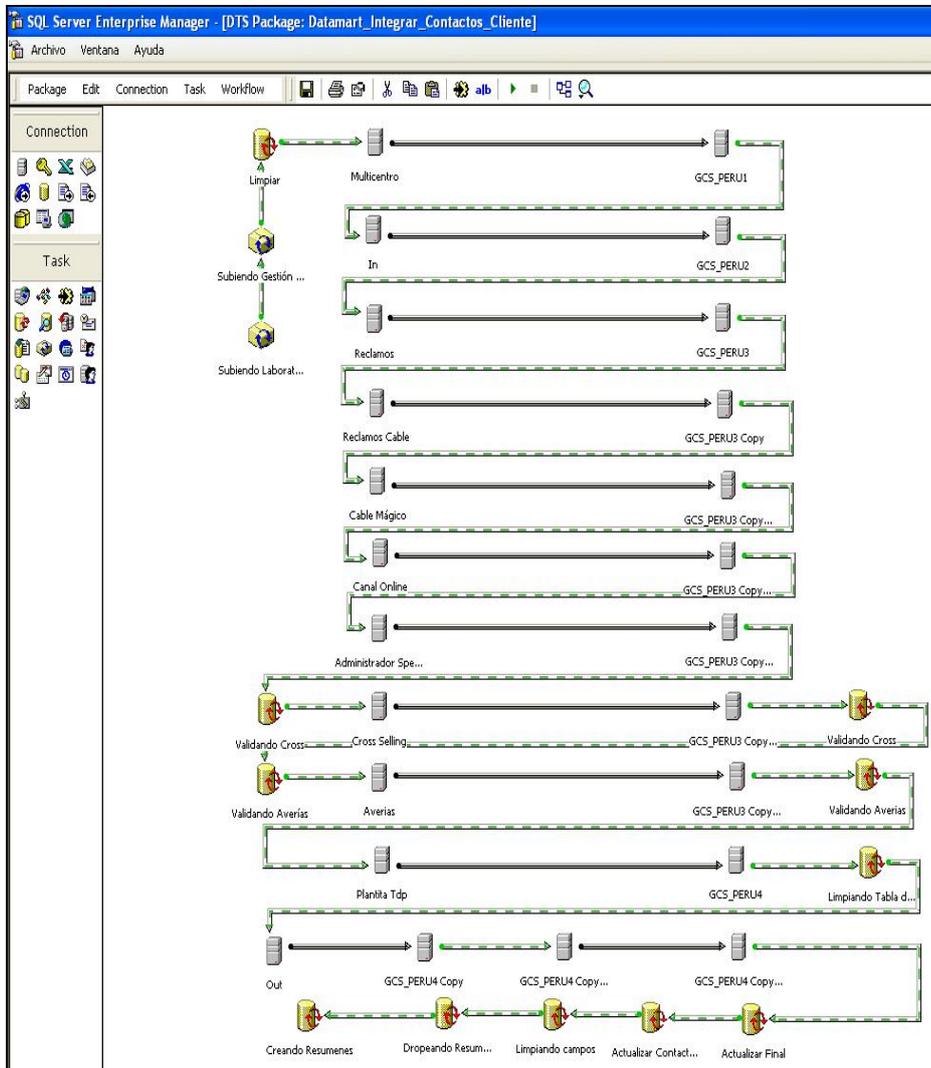


Figura IV.17 DTS de Integración de los Contactos con el Cliente de Telefónica

4.5.3. Construcción del Cubo del Datamart

Para la construcción del cubo multidimensional utilizado en la simulación del Datamart se utilizó el modelo copo de nieve, en el cual la estructura de las tablas de dimensiones es descompuesta o normalizada en una estructura de árbol que puede disponer de muchos niveles o jerarquías.

La construcción del cubo se realizó de la siguiente manera:

- **Utilizando el Analysis Manager, se inicia el asistente para generar cubos**

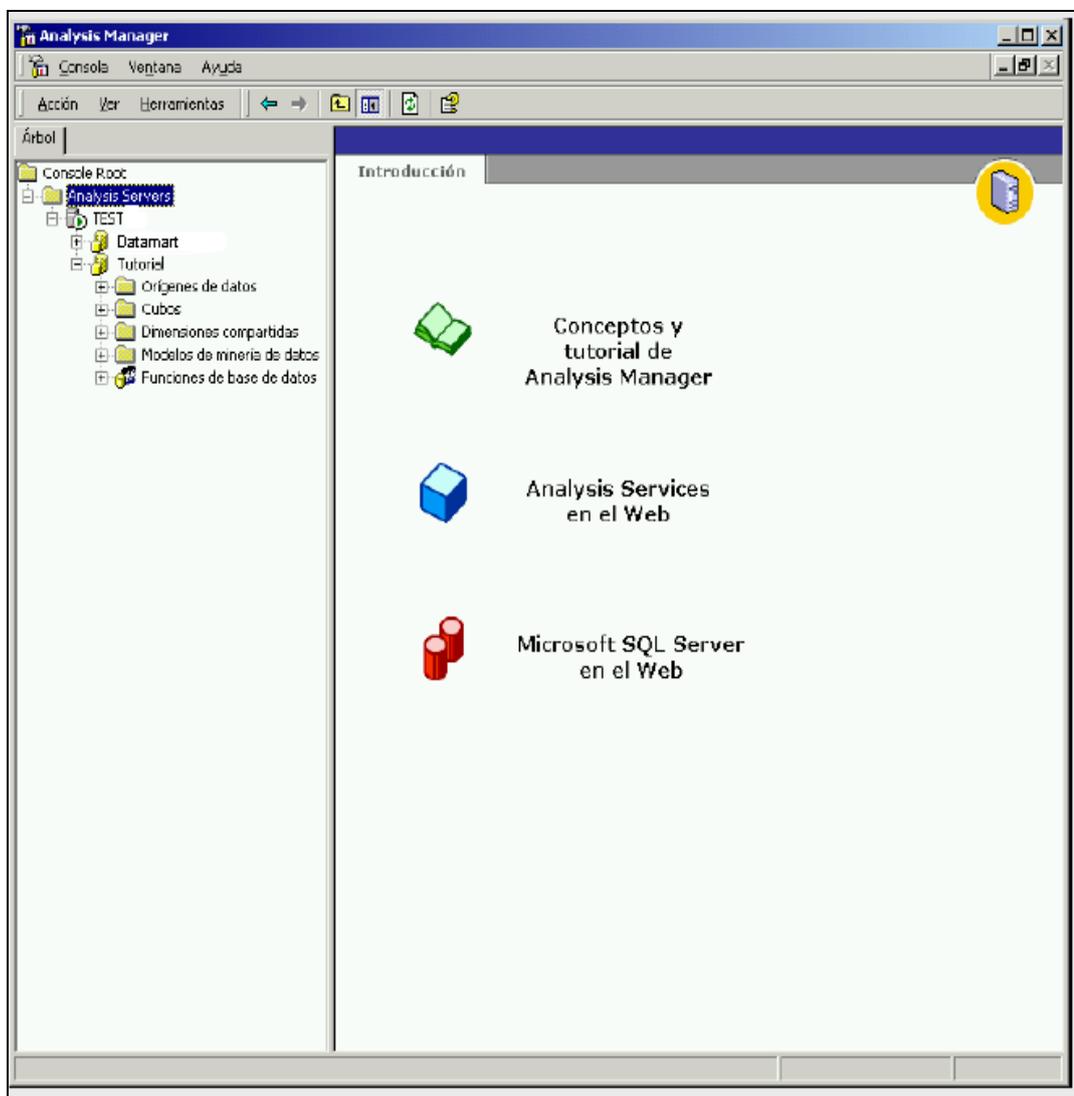


Figura IV.18 Asistente para generar cubos del Analysis Manager

- Se define la fuente de origen del cubo

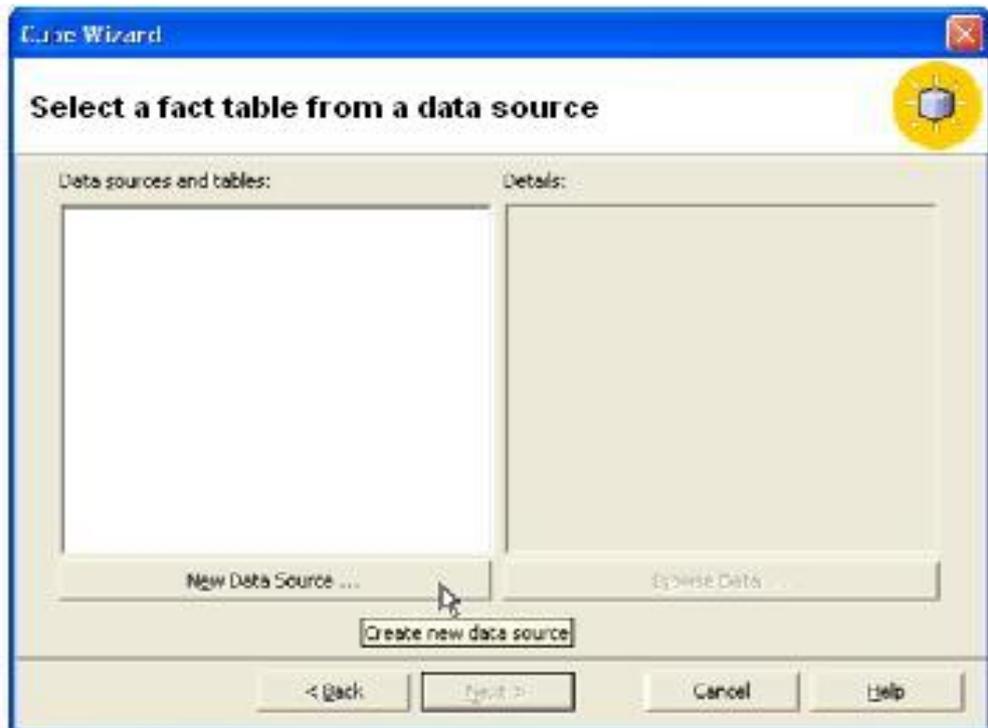


Figura IV.19 Selección de un nuevo origen de datos

- Se selecciona el proveedor del vínculo de datos



Figura IV.20 Selección del proveedor del vínculo de datos

- Se selecciona la base de datos del Datamart

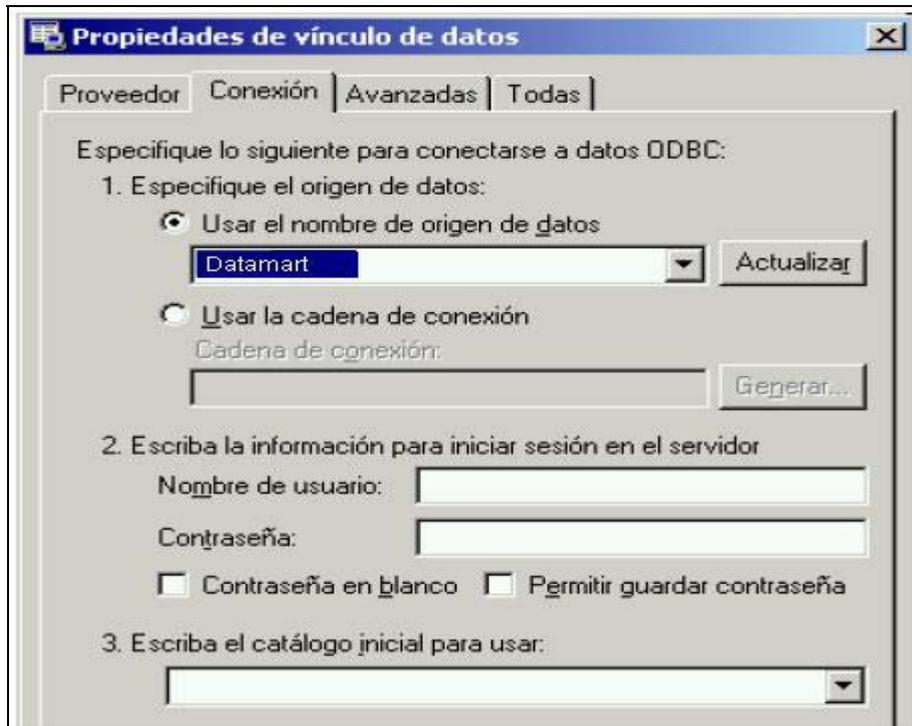


Figura IV.21 Selección de la base de datos del Datamart

- Definición de la Tabla de hechos y sus medidas

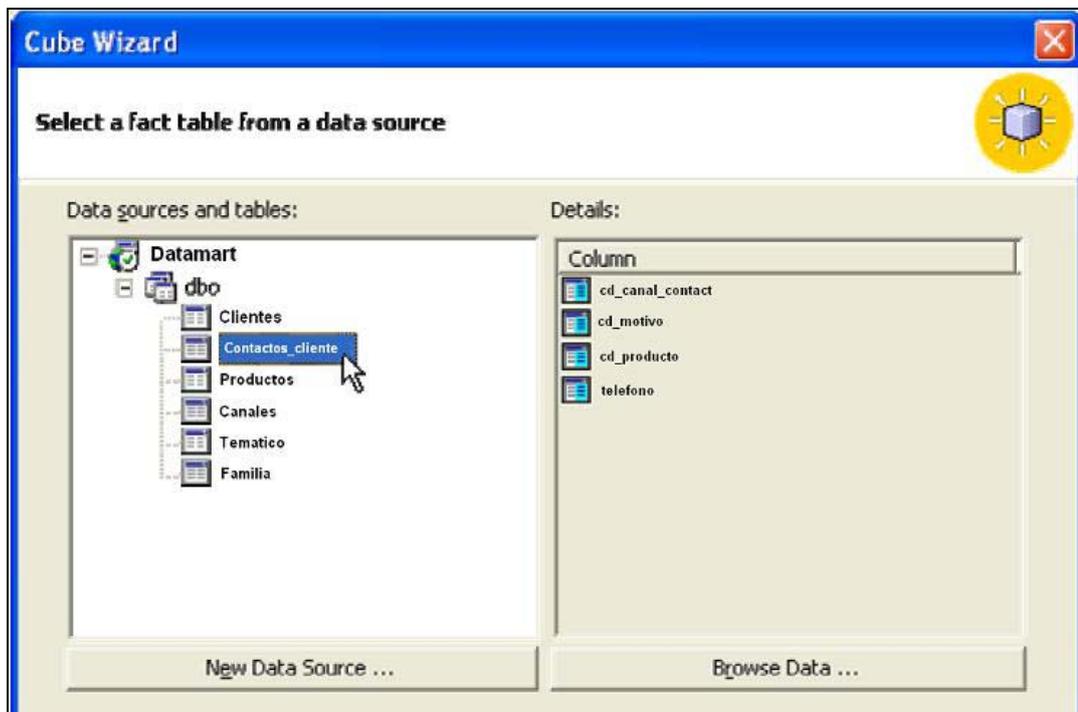


Figura IV.22 Definición de la Tabla de hechos y sus medidas

- **Definición de las dimensiones y sus jerarquías**

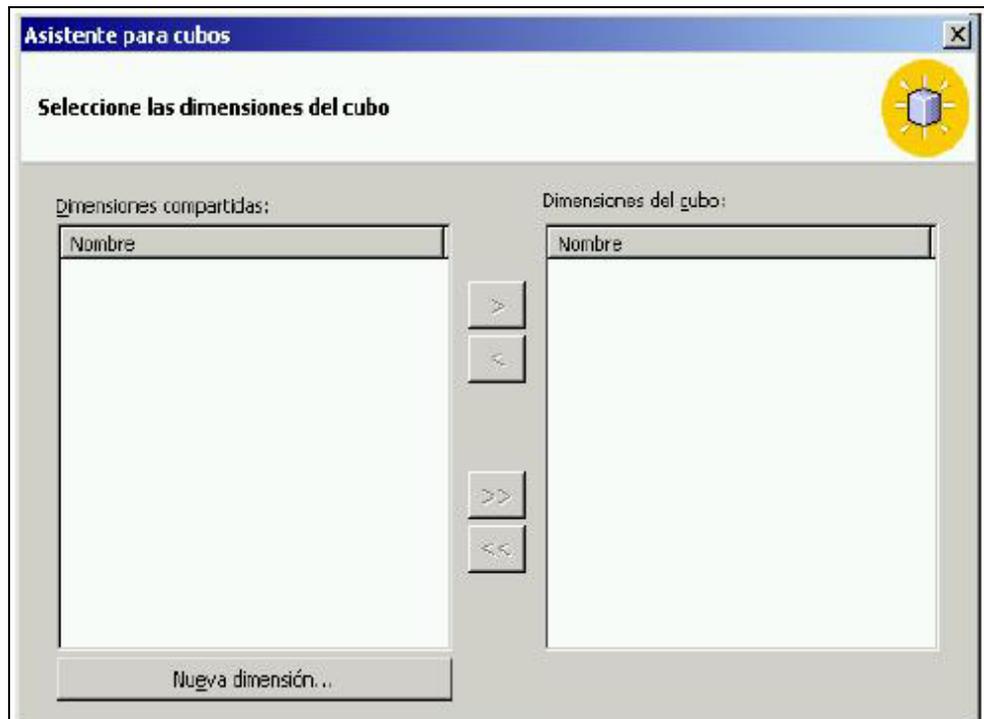


Figura IV.23 Definición de las dimensiones y jerarquías del cubo

- **Creación de la Dimensión Tiempo**



Figura IV.24 Creación de la Dimensión Tiempo

• Visualización del cubo realizado

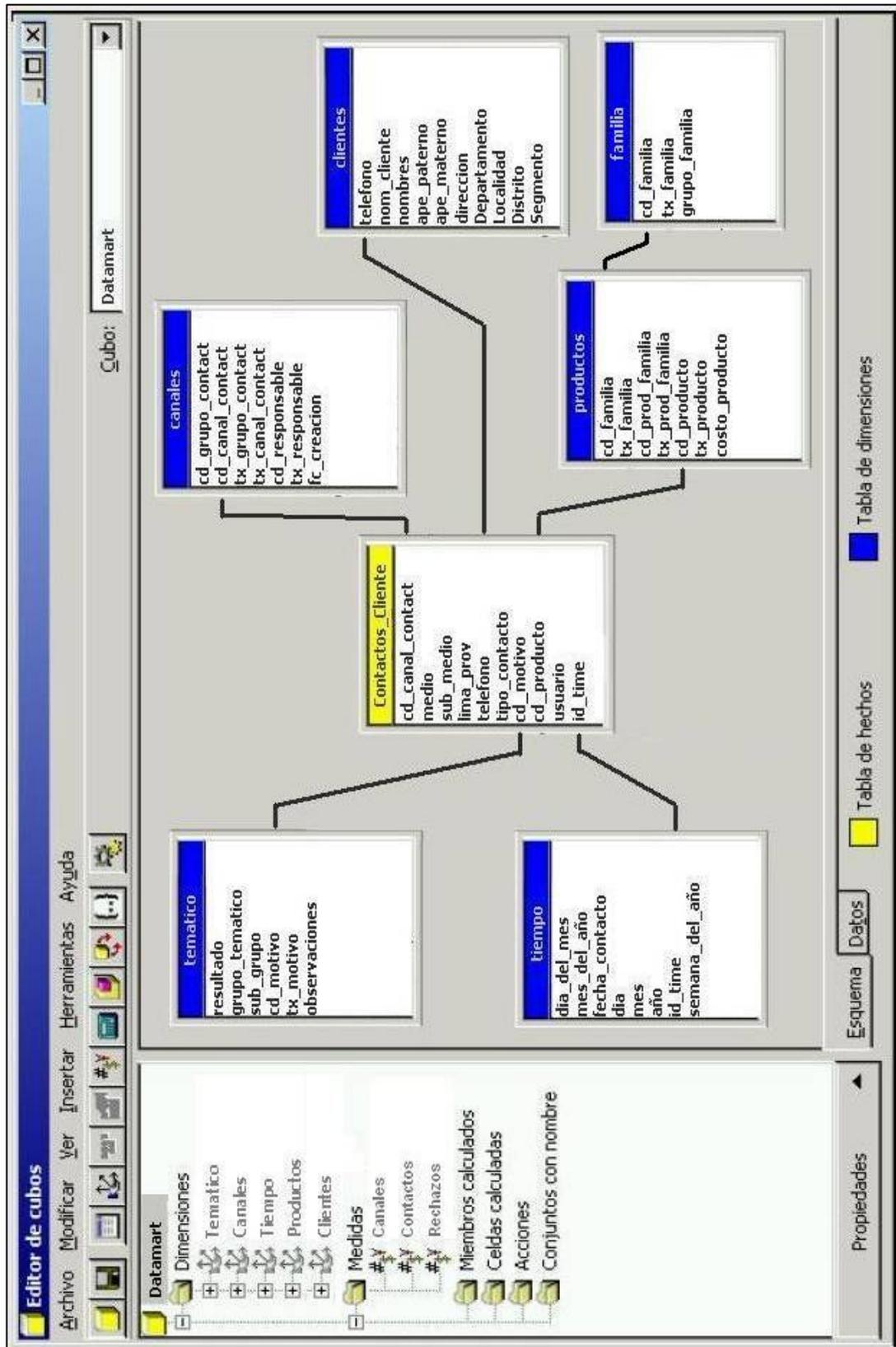


Figura IV.25 Vista del Cubo del Simulador del Datamart finalizado

4.5.4. Construcción de Web de Contactos del Simulador

Como resultado del Simulador se construyeron dos vistas web que muestran a menor escala uno de los beneficios del Datamart de Contactos que es la integración de la información de los contactos del cliente con Telefónica.

El cual puede ser explotado a través del teléfono del cliente y aprovechado para incrementar el conocimiento del asesor en el historial del cliente.

La Web de Contactos será muy similar a la construída durante la simulación del proyecto, la misma que puede ser apreciada en la Figura IV.26 y IV.27 que se muestran a continuación:

Inteligencia Comercial

sábado, 27 de noviembre de 2010 Bienvenido Alejandro Gamarra a nuestra Página Web Cerrar Sesión

Menu Principal

- Principal
- Quiénes somos
- Campañas
- Laboratorio
- Reporting
- Requerimientos
- Sugerencias

SEGUIMIENTO DE CLIENTES 360°

ESTADÍSTICAS DE CLIENTE

Contatos por medio de acción:

MEDIO	CONTACTOS	PROM.
CableMágico	367656	4085
In	3756201	41735
Multicentro	152650	1696
Out	7150402	79448

Cuadro que nos muestra la cantidad de contactos realizados en los últimos 3 meses y el promedio diario.

Ranking de Reclamos más frecuentes :

RECLAMOS	TOTAL	PROM.
Top 1	18268	202
Top 2	8777	97
Top 3	8655	96

Cuadro que nos muestra la relación de reclamos más frecuentes realizados en los últimos 3 meses y el promedio diario.

(c) 2010 - Telefónica del Perú S.A.A. - Vicepresidencia del Segmento Residencial

Figura IV.26 Web de Contactos con conexión al Datamart de Contactos
Página de Búsqueda

SEGUIMIENTO DE CLIENTES 360°

Ingrese el número deseado:

DATOS DEL CLIENTE		TENENCIA DE PC		DATOS ADICIONALES	
Cliete:	Gamarra Ramirez, Alejandro javier	Línea:	tarifa plana nacional	Top Premium:	No
Segmento:	oro - Servicio	Speedy:	convencional speedy 2000	Robinson:	No
Lima / Prov:	lima	Cable:	estelar tdp digital	Indecopi Mail:	SI - 2010/04/06
Departamento:	lima	Paquetización:	---	Indecopi Fijo :	SI - 2010/04/06
Distrito:	ate vitarte			Indecopi Celu :	---

Campañas enviadas al Cliente en el último mes						
Tipo de envío	Campaña	Acción	Destino	Oferta	Fecha de envío	
					Total de envíos:	0

Contactos con el Cliente en los últimos 3 meses									
Fuente	Medio	Sub medio	Fecha	Hora	Tipo Contacto	Motivo 1	Motivo 2	Observaciones	
Cable Magico	43463154	Dato no disponible	31/08/2010	00:00:00	llamada telefónica	conocer ofertas y requisitos	Dato no disponible	Dato no disponible	
Canal Online	Página Web de Telefónica		17/09/2010		módulo speedy	adquirir producto	ya tiene el producto		
In	conocer ofertas y requisitos		31/08/2010	13:52:22		conocer ofertas y requisitos			
								Total de envíos:	3

(c) 2010 - Telefónica del Perú S.A.A - Vicepresidencia del Segmento Residencial

Figura IV.27 Web de Contactos con conexión al Datamart de Contactos
Página de Resultados

4.6. DESPLIEGUE

Las actividades que están planificadas para llevarse a cabo en el despliegue son las siguientes:

- Carga la base de datos del Datamart con los datos de los Clientes Residenciales: se refiere a la creación de la tabla principal de la base de datos única con los datos de los clientes mediante consultas de SQL y uso de SQL Loader.
- Poblar las tablas de carga con los datos de las filiales: se refiere a la extracción de los datos de las filiales desde los archivos de texto plano, Excel, Access, Fox pro, etc suministrados, hacia las tablas de carga de la base de datos única mediante SQL Loader.
- Realización de las pruebas y ajustes necesarios a la solución: se refiere a las pruebas controladas que se realizarán sobre la solución a fin de verificar que se

llevan a cabo las tareas de forma correcta. El plan de pruebas realizado contempló las reglas de negocio establecidas de tal manera de garantizar su cumplimiento.

- Puesta en producción del Datamart y la Web de Contactos, para que tantos los usuarios que tendrán acceso al Reporting Services y los asesores comerciales que tendrán acceso a la Web de Contactos puedan empezar a utilizar la información disponible.

- Sobre éste último punto, una vez el Datamart se encuentre en producción será explotable a través de Reporting Services mediante Dashboards, el cual provee un entorno muy amigable y reportes dinámicos que ofrecen al usuario interno una información fácil de manejar.

A continuación, algunos ejemplos de los Dashboards que estarán disponibles en el Reporting Services:

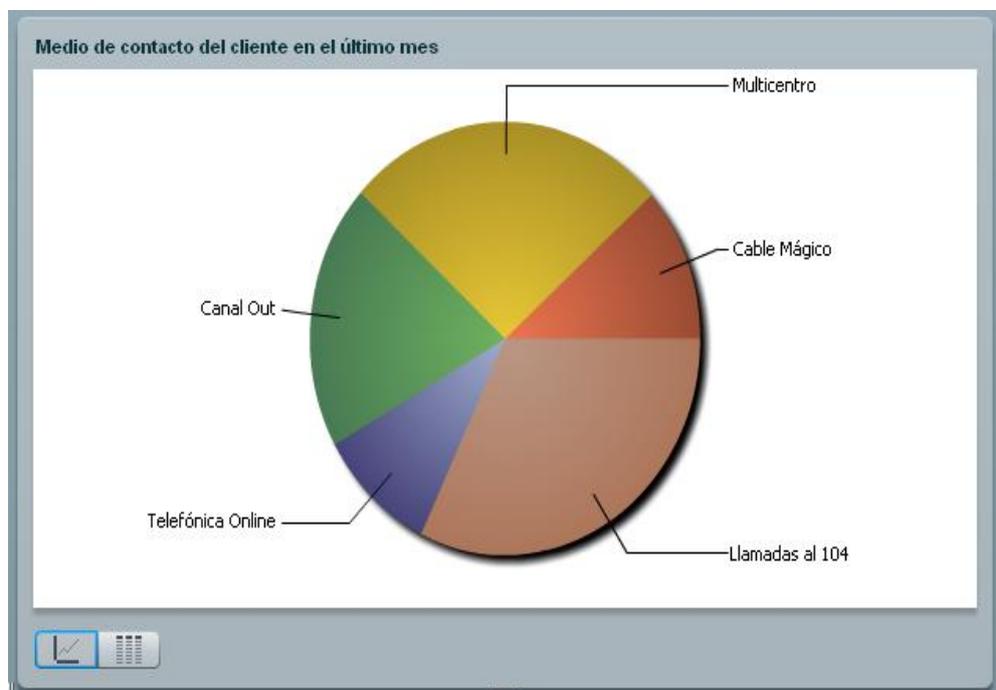


Figura IV.28 Dashboard – Reporte por Medio de Contactos en el último mes

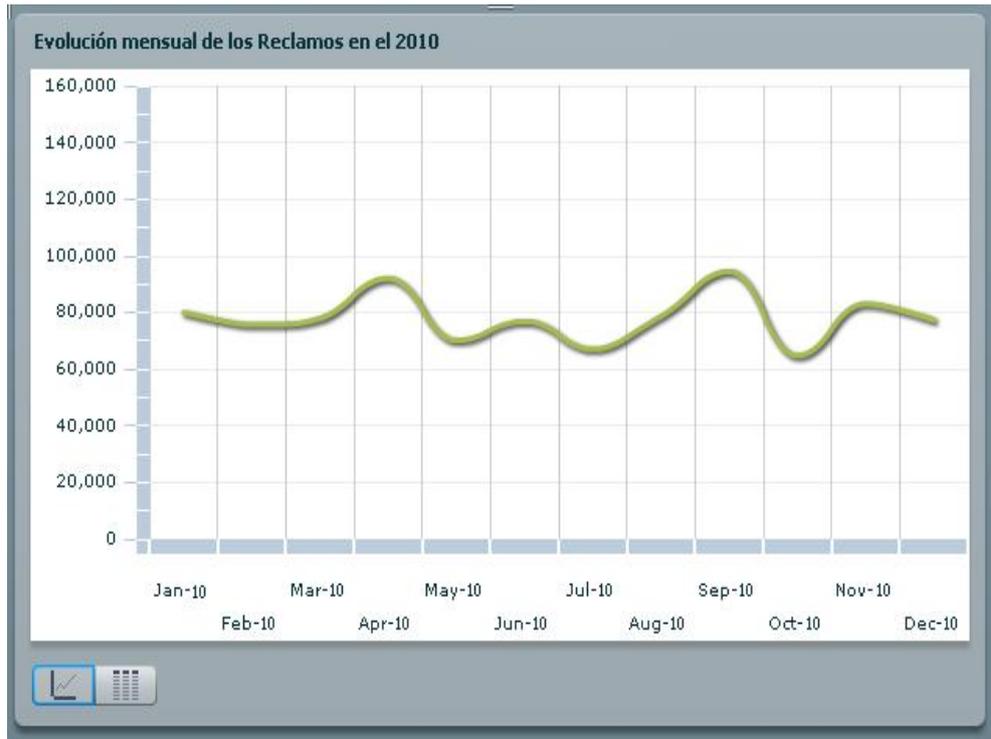


Figura IV.29 Dashboard – Reporte de la Evolución Mensual de los Reclamos en el 2010



Figura IV.30 Dashboard – Reporte de Satisfacción del Cliente por Medio de Contacto

Éste último Dashboard es la combinación de la Figura IV.27 y IV.28, el cual muestra la evolución de los Reclamos por el Canal Out durante el año 2010.

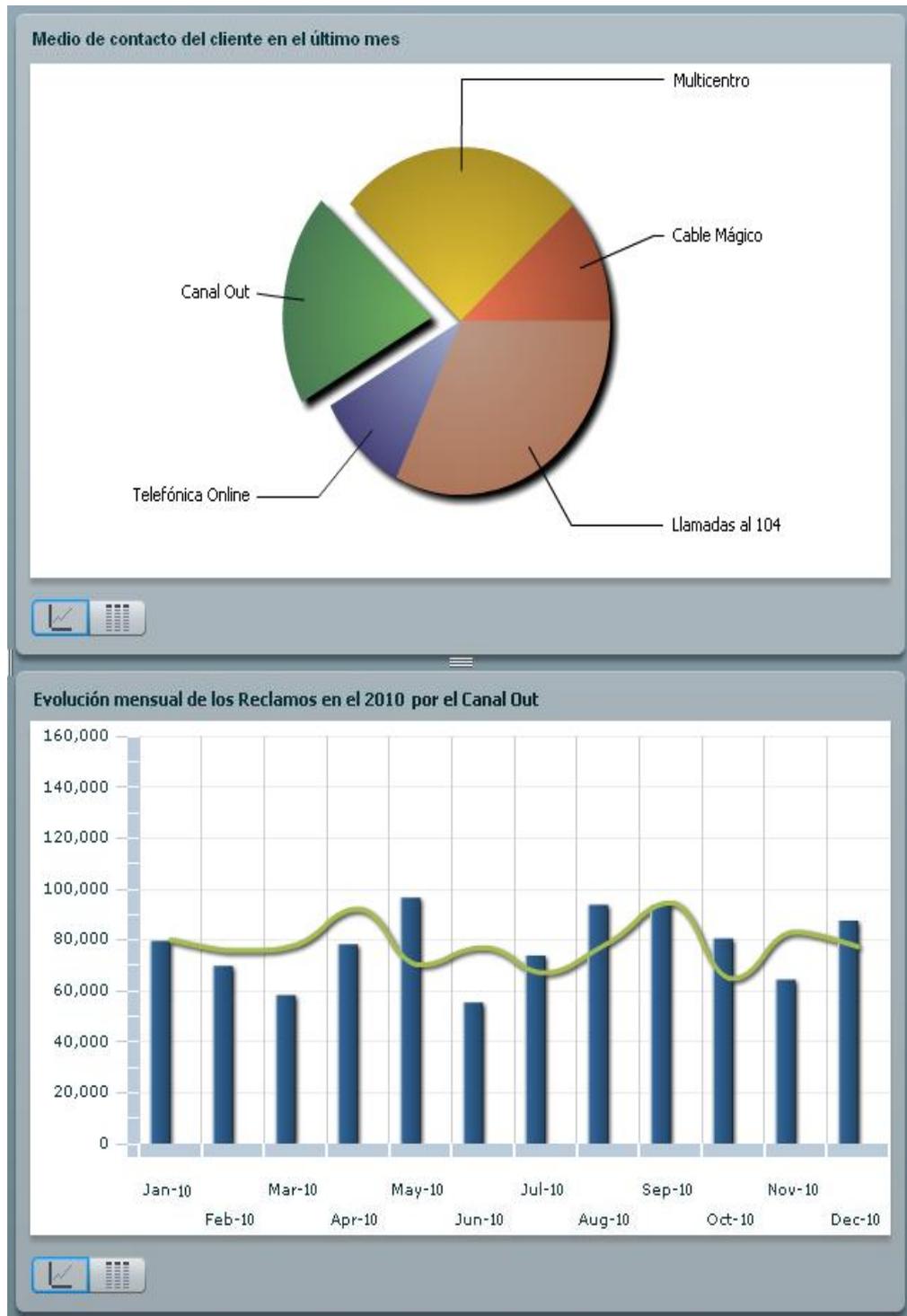


Figura IV.31 Dashboard – Evolución mensual de los Reclamos en el 2010 por el Canal Out

- La Web de Contactos como observamos en las Figuras IV.25 y IV.26 serán vinculadas al entorno de trabajo del asesor comercial para incrementar el conocimiento del cliente en tiempo real, el asesor comercial tendrá que usar esta herramienta ya que no posee privilegios para acceder al Datawarehouse de la empresa.

Como se puede apreciar de esta manera la solución ofrecida en esta tesina no sólo mejorará el trabajo de los usuarios internos de la empresa sino que también provee las herramientas al usuario externo para que realice mejor su trabajo teniendo como meta siempre la satisfacción del cliente.

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. CONCLUSIONES

- Los sistemas de soporte a las decisiones son una necesidad cada vez más apremiante en las empresas ante un mercado más competitivo. En el caso de Telefónica del Perú esta necesidad se ve acentuada ante el ingreso de nuevos actores en el mercado de telecomunicaciones, quienes buscan capturar clientes a toda costa.
- El Datamart de contactos es una respuesta a esta necesidad porque otorga a corto plazo una información valiosa de todos los contactos con el cliente que hasta ahora no era disponible: un sistema que pudiera registrar todas las interacciones entre el cliente y la empresa o viceversa, en un único repositorio, obteniendo así, la información más importante para crear nuevos productos y servicios, mejorar la atención al cliente, etc.
- Para el desarrollo de un Datamart, es fundamental tener en cuenta las necesidades de información que plantean los usuarios finales. De nada sirve una implementación que no dé respuesta a las preguntas definidas por los futuros usuarios.
- En cuanto a los objetivos planteados al iniciar este trabajo fueron cubiertos siguiendo los lineamientos teóricos de la investigación, al disponer en primer lugar del Datamart de contactos, el cual se alimenta de las fuentes de contactos del ATIS que funge como semilla y los datos de la página web de consultas de contactos, cargados mediante los procesos de extracción, transformación y carga que contemplan las reglas de negocio establecidas durante la fase de requerimientos. En segundo lugar, se dispone de la aplicación de análisis y revisión de datos, la cual cubre los aspectos de depuración y revisión de los datos suministrados por los

puntos de contactos de clientes. Para llevar a cabo todas las actividades descritas, fue necesario realizar la instalación de las herramientas de software y hardware necesarias.

- La investigación realizada en el marco teórico del presente trabajo dio sustento a las aplicaciones y estudios realizados debido a que aportó los conocimientos teóricos de Datamart, sistemas de soporte a las decisiones, procesos ETL, Datamining, entre otros.
- El marco metodológico utilizado, con base a la metodología de desarrollo de Datawarehouse planteada por Harjinder & Prakash, definió las fases y actividades necesarias para acometer las tareas requeridas en la elaboración del presente trabajo.
- La metodología utilizada para el desarrollo de Datawarehouse ofrece ventajas sustanciales con respecto a las metodologías clásicas de desarrollo de sistemas de información, en las etapas iniciales de recopilación y análisis de requerimientos.
- La aplicación Web desarrollada sólo responde a consultas fijas sobre la data cargada es por ello que el Datamart debe ser acompañado de una herramienta de consulta de nivel superior para toma de decisiones ya que la información almacenada es sólo el nivel base de almacenamiento.
- Los datos almacenados en el Datamart se convierten en información que aporta conocimiento a la empresa Telefónica del Perú referente a los contactos y necesidades diarias de sus clientes con la empresa.

5.2. RECOMENDACIONES

- Para mantener la precisión de los datos contenidos en el Datamart de Contactos es necesario realizar un proceso de calidad de datos que involucre la verificación y ajuste de la información disponible en la base de datos luego de la ejecución de los procesos. Una vez culminando este proceso deberán realizarse los ajustes necesarios en los procesos de validación de clientes para cargas posteriores, tomando en cuenta los resultados del proceso de calidad.
- Proporcionar adiestramientos a los usuarios finales en el uso de la herramienta final que se seleccione para trabajar sobre la plataforma de Datamart desarrollada y crear en ellos conciencia acerca de la importancia de las acciones que ellos realicen sobre los registros de la base de datos y las repercusiones que ocasionan en la calidad y precisión de la información.
- Entrenar a los administradores del repositorio de información para mantener la data con revisiones oportunas de los archivos log que se generan de la carga masiva de información.
- El sistema aportaría mayor valor al disponer de una herramienta de minería de datos debido a que mediante las técnicas de análisis de estas herramientas se podrán satisfacer las necesidades del departamento de mercadeo de forma más ágil y acertada que las consultas realizadas mediante SQL. Es por ello que adquirir e implantar la herramienta de Datamining, ofrecería una ventaja competitiva aun mayor para la empresa y a corto plazo.

BIBLIOGRAFÍA

1. Tesis

- Desarrollo de una Data Mart para el manejo de experiencia de proyectos de Inelectra S.A.C.A. haciendo uso de herramientas OLAP. Tesis de grado (Ing. Informática) -- Universidad Católica Andrés Bello, Facultad de Ingeniería, 2003. Realizado por Alcides Figueroa, Gabriela Reiner.
- Elaboración de un Data Mart de clientes PYME, a las unidades de mercadeo de CANTV Corporación. Tesis de grado (Ing. Informática) -- Universidad Católica Andrés Bello, Facultad de Ingeniería, Escuela de Ingeniería Informática, 2002. Realizado por Alvaro Enrique Trak Vásquez, Emma Carolina Zambrano Marcano.
- Diseño e implantación de un Datamart para el área de facturación y ventas de la empresa Makro Comercializadora, S.A. Sede La Urbina. Tesis de grado (Ing. Informática).-- Universidad Católica Andrés Bello, Facultad de Ingeniería, 2005. Régulo E. Guédez Flores, Yaranaís Zambrano Loyo.

2. Material Impreso

- Hanna, P. (2002). JSP Manual de referencia. España: Madrid
- Harjinder & Prakash (1996). Data Warehousing, la integración de información para la mejor toma de decisiones. México. Prentice Hall.
- Kimball Ralph; Ross Margy (2000). The Datawarehouse Toolkit. USA. John Wiley.
- Inmon, W. (2002). Building the Data Warehouse (3rd Edition). United States of America.

- Dyché, Jill (2001) E-Data: transformando datos en información con Data Warehousing (Primera ed.) Buenos Aires. Prentice Hall.

3. Páginas Web

- Nigro, Oscar. (2005). Estructura del Data Warehouse. Disponible en:
http://www.exa.unicen.edu.ar/catedras/dwhouse/Clase3_DW.zip
- Brotons, Antonio (2001) Datawarehousing.
Consultado el día 15 de Octubre del 2010 en la World Wide Web:
<http://comunidad.icemd.com/area-entrada/documentos/ver-documento.asp?REF=46>
- Palomar Sanz, Manuel (2001). Uso y diseño de Base de Datos Multidimensionales y almacenes de datos.
Consultado el día 15 de Octubre del 2010 en la World Wide Web:
www.dlsi.ua.es/~jtrujillo/doctoradotema2bnpro.pdf
- Gutierrez E, Damián. Data Warehouse, (Pagina web en línea). México:
Consultado el día 15 de Octubre del 2010 en la World Wide Web:
<http://www.monografias.com/trabajos17/data-warehouse/data-warehouse.shtml>
- Microsoft; “SQL Server 2000 Reporting Services”;
Consultado el día 15 de Octubre del 2010 en la World Wide Web:
http://download.microsoft.com/download/a/f/2/af2d3215-c3fb-4a88-89cf-3d55b8ddf17d/sql_reporting_services_es.pdf