Universidad Nacional Mayor de San Marcos

Universidad del Perú. Decana de América

Facultad de Ingeniería de Sistemas e Informática
Escuela Académico Profesional de Ingeniería de Sistemas

Gestión de la información basada en inteligencia de negocios para el área de establecimientos. Caso Diners Club S.A.

TESINA

Para optar el Título Profesional de Ingeniero de Sistemas

AUTOR

Teodoro Manuel ANDRADE MOGOLLÓN

ASESOR

Carlos Edmundo NAVARRO DEPAZ

Lima, Perú

2008

Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

https://creativecommons.org/licenses/by-nc-sa/4.0/

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Andrade, T. (2008). *Gestión de la información basada en inteligencia de negocios para el área de establecimientos. Caso Diners Club S.A.* Tesina para optar el título profesional de Ingeniero de Sistemas. Escuela Académico Profesional de Ingeniería de Sistemas, Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima, Perú.

Este trabajo esta dedicado a toda mi familia en especial a mi hijo Alessandro.

Con toda mi gratitud para todas aquellas personas que en el paso del tiempo estuvieron siempre apoyándome.

AGRADECIMIENTOS

La elaboración de este trabajo tomo tiempo y esfuerzo no solo de parte mía, sino de todas aquellas personas que colaboraron de una forma u otra. Es por esto que quiero agradecer a Milagros y Alessandro, mi hijo, quienes soportaron con paciencia las largas noches de trabajo y aquellos días en que no tenía mucho tiempo para dedicarle por trabajar en la tesina.

A toda mi familia en especial a mis padres que depositaron en mí toda su confianza y me dieron la posibilidad de estudiar, ellos que siempre se esforzaron para lograr junto conmigo los objetivos que hoy cumplimos. A mi prima Paty que me apoyo en toda la carrera, a mi hermano David por apoyarme y compartir mi alegría y a mis abuelos que desde el cielo me entregaron toda la ayuda espiritual, gracias.

Al decano de la Facultad de Ingeniería de Sistemas e Informática Carlos Navarro de Paz por su orientación y dedicación para que este trabajo cumpla con los objetivos trazados.

A mis colegas y amigos de la UNMSM en especial a la base 98 un excelente grupo de personas que siempre luchamos para ver a nuestra escuela en lo que es hoy, la Facultad de Ingeniería de Sistemas e Informática.

A los profesores de la UNMSM, principalmente al profesor Luis Roig y Virginia Vera grandes personas a las cuales siempre recuerdo con mucho cariño.

A todas aquellas personas que indirectamente me ayudaron para culminar este trabajo y que muchas veces constituyen un invalorable apoyo.

Y por encima de todo doy gracias a Dios.

GESTIÓN DE LA INFORMACIÓN BASADA EN INTELIGENCIA DE NEGOCIOS PARA EL AREA DE ESTABLECIMIENTOS

CASO DINERS CLUB SA

RESUMEN

Este proyecto busca diseñar e implementar una solución de inteligencia de negocios que permita mejorar la capacidad de la organización en la toma de decisiones a través del análisis de la información, en este caso el proyecto se enfocará en el área de establecimientos de la empresa Diners Club SA.

Palabras Claves: Inteligencia de negocios, análisis multidimensional, datamart, datawarehouse, datawarehousing

INFORMATION MANAGEMENT BASED ON BUSINESS INTELLIGENCE FOR THE AREA OF ESTABLISHMENT CASO DINERS CLUB SA

ABSTRACT

This project seeks to design and implement a business intelligence solution to improve the organization's capacity in decision-making through information analysis, in this case the project will focus on the the area of the business establishments Diners Club SA

.

Key words: Business Intelligence, multidimensional analysis, datamart, datawarehouse, datawarehousing

ÍNDICE

Lista de Figuras x

Lista de Tablas xiv

CAPÍTULO 1: INTRODUCCIÓN 1

1.1 Antecedentes 1

1.2 Definición del problema 3

1.3 Objetivos 3

1.4 Justificación 4

1.5 Propuesta 4

1.6 Organización de la tesina 4

CAPÍTULO 2: MARCO TEÓRICO 6

2.1 Niveles de la Organización y Necesidad de Información 7

2.1.1 Sistemas Estratégicos 7

2.1.2 Sistemas Tácticos 7

2.1.3 Sistemas Técnico – Operativos 8

2.2 Gerentes de la Organización 8

2.3 Data Warehouse 9

2.3.1 Características 10

2.3.2 Problemas y Ventajas del Data Warehouse 13

2.3.3 Diferencias entre Sistema Tradicional (OLTP) y Data Warehouse 14

2.3.4 Arquitectura de un Data Warehouse 14

2.4 Modelo de Datos 17

2.4.1 Modelo Relacional 18

2.4.2 Modelo Multidimensional 19

2.5 OLAP (Online Analytical Processing) 27

2.5.1 Herramientas OLAP 28

2.5.2 Características de las Herramientas OLAP 29
2.5.3 Modos de Almacenamiento 30
2.5.4 Data Warehouse y OLAP 32

CAPÍTULO 3: ESTADO DEL ARTE CICLO DE VIDA Y METODOLOGÍA DE IMPLEMENTACIÓN PARA SOLUCIONES DE INTELIGENCIA DE NEGOCIOS 34

3.1 Planificación del Proyecto 35

3.2 Definición de los Requerimientos del Negocio 36

3.3 Modelado Dimensional 38
3.3.1 Componentes básicos 39
3.3.2 Comparación con modelo ER 44

3.4 Diseño Físico 45
3.4.1 Mapeo desde Modelo Lógico hacia Esquema Físico 46
3.4.2 Construcción de tablas Lookup 48
3.4.3 Relaciones de atributos dentro de una dimensión 49
3.4.4 Definición de tablas Base 49
3.4.5 Tipos de esquemas 50

3.5 Diseño y Desarrollo de Presentación de Datos 57

3.6 Diseño de la Arquitectura Técnica 59

3.7 Selección de Productos e Instalación 60

3.8 Especificación de Aplicaciones para Usuarios Finales 61

3.9 Desarrollo de Aplicaciones para Usuarios Finales 63

3.10 Implementación 64

3.11 Mantenimiento y crecimiento 67

3.12 Gerenciamiento del Proyecto 69

3.13 Puntos problemáticos de un DW 70

CAPÍTULO 4: METODOLOGÍA DE DESARROLLO PARA LA CONSTRUCCIÓN DEL DATAMART ESTABLECIMIENTOS CASO DINERS CLUB…….. 71

4.1 Organización del Proyecto - Anteproyecto 72

4.2 Análisis de Requerimientos de Información 73

4.3 Arquitectura, Tecnología y Modelamiento Dimensional 77

4.3.1 Dimensiones y Tabla de Hechos 78

4.3.2 Matriz de Indicadores Tabla Nº 4.15 88

4.3.3 Modelo Lógico 89

4.3.4 Modelo Dimensional 92

4.4 Diseño e Implementación de Procesos de Extracción, Transformación y Carga (ETL) 94

Extracción y Transformación de Datos 96

4.4.1 Proceso Central AS400 96

4.4.2 Proceso Central SQL 100

4.4.3 Creación de Tablas 111

4.4.4 Integration Services 119

4.5 Explotación de la Información 122

CAPÍTULO 5: SOLUCIÓN TECNOLÓGICA 126

CAPÍTULO 6: CONCLUSIONES Y TRABAJOS FUTUROS 132

6.1 Conclusiones 132

6.2 Trabajos Futuros 132

Índice de figuras

2.1 Niveles de la Organización y Necesidades de Información 7

2.2 Clasificación de los Gerentes en las Organizaciones 8

2.3 Data Warehouse Integrado 11

2.4 Data Warehouse Temático 11

2.5 Data Warehouse No Volátil 12

2.6 Arquitectura real de un Data Warehouse 16

2.7 Esquema Estrella 20

2.8 Esquema Copo de Nieve 21

2.9 Constelación de Hechos 21

2.10 Ejemplo de un Cubo con tres dimensiones 22

2.11 Slice & Dice 23

2.12 Rotación 24

2.13 Rotación Drill-Down y Drill-Up 24

2.14 Roll-Up 25

2.15 Drill-Across 25

2.16 Drill-Trough 26

2.17 Arquitectura Sistema ROLAP 31

2.18 Arquitectura Sistema MOLAP 32

2.19 Interacción entre Data Warehouse y OLAP 33

3.1 Business Dimensional Lifecycle propuesto por Ralph Kimball 34

3.2 Interacción entre Planificación y Requerimientos 35

3.3 Interacción entre Requerimientos del Negocio y etapas subsiguientes 37

3.4 Diagramas para Modelado Dimensional 38

3.5 Esquema de un modelo dimensional 39

3.6 Dimensión 40

3.7 Atributos 41

3.8 Elementos o instancias de atributos 41

3.9 Relaciones y Jerarquías 43

3.10 DER vs.BDM 44

3.11 Proceso de diseño físico a alto nivel 45

3.12 Mapeo del BDM a Esquema Físico Relacional 46

3.13 Tablas Lookup 48

3.14 Implementación de relaciones 48

3.15 Tabla Base 49

3.16 Esquema Estrella Consolidado I 50

3.17 Esquema Estrella Consolidado II 51

3.18 Esquema Estrella Consolidado III 51

3.19 Esquema Estrella Normalizado I 53

3.20 Esquema Estrella Normalizado II 54

3.21 Esquema Estrella Normalizado III 55

3.22 Proceso ETL 57

3.23 Modelo de Arquitectura Técnica a alto nivel 58

3.24 Ejemplo de matriz de evaluación de productos 60

3.25 Variedad de interfaces por perfil 60

3.26 Configuración de Metadata y construcción de reportes 62

3.27 Convergencia de la tecnología, los datos y las aplicaciones 64

3.28 Data Warehousing como proceso de naturaleza espiral 67

3.29 Gerenciamiento del proyecto 69

4.1 Organigrama Diners Club SA 70

4.2 Metodología Implementación 71

4.3 Análisis de Requerimientos 72

4.4 Arquitectura, tecnología y modelo dimensional 76

4.5 Modelo Lógico 90

4.6 Modelo Dimensional 93

4.7 Etapas Procesos ETL I 93

4.8 Etapas Procesos ETL II 94

4.9 AS400 ETL Dimensión ESTABLECIMIENTOS 96

4.10 AS400 ETL Dimensiones VARIAS 97

4.11 AS400 Fact Tables VISITAS Y ENTREGA DE MATERIALES 98

4.12 AS400 Fact Tables AJUSTES 98

4.13 AS400 Fact Tables PAGOS 99

4.14 SQL Dimensión ARTÍCULO 100

4.15 SQL Dimensión REPRESENTANTE 101

4.16 SQL Dimensión FORMA PAGO ESTABLECIMIENTO 102

4.17 SQL Dimensión TIPO VISITA 103

4.18 SQL Dimensión ESTABLECIMIENTO 104

4.19 SQL Fact Table AJUSTES 105

4.20 SQL Fact Table PAGOS 106

4.21 SQL Fact Table VISITAS 107

4.22 SQL Fact Table PROYECCIÓN DE AFILIACIONES 108

4.23 SQL Fact Table PROYECCIÓN DE VISITAS 109

4.24 SQL SERVER 2005 117

4.25 Base de datos SQL 117

4.26 Integration Services Actualiza Dimensión Establecimiento 118

4.27 Integration Services Actualiza Tabla de Hechos Visita 119

4.28 Integration Services Actualiza Tabla de Hechos Ajustes 119

4.29 Integration Services Actualiza Tabla de Hechos Pagos 120

4.30 Integration Services Actualiza Proyección de Visitas 120

4.31 Integration Services Actualiza Proyección de Afiliaciones 121

4.32 Explotación de Información 121

4.33 Ingreso al Diseñador 123

4.34 Creación de Universo 123

4.35 Generación de Reporte 124

5.1 Entorno Web 126

5.2 Ingreso a la Aplicación Business Objects 127

5.3 Lista de Datamarts y Tableros 127

5.4 Construcción de Consultas 128

5.5 Reporte I 129

5.5 Reporte I I 129

5.5 Reporte III 130

Índice de tablas

2.1 Diferencias entre Sistema Tradicional versus Data Warehouse 4

4.1 Dimensión Establecimiento 80

4.2 Dimensión Artículo 80

4.3 Dimensión Tipo Visita 80

4.4 Dimensión Representante 80

4.5 Dimensión Ejecutivo 81

4.6 Dimensión Tiempo 81

4.7 Dimensión Forma Pago Establecimiento 82

4.8 Dimensión Geografía 82

4.9 Tabla de Hechos Visita 83

4.10 Tabla de Hechos Entrega de Materiales 83

4.11 Tabla de Hechos Ajustes 84

4.12 Tabla de Hechos Pagos 86

4.13 Tabla de Hechos Proyección de Afiliaciones 86

4.14 Tabla de Hechos Proyección de Visitas 86

4.15 Matriz de Indicadores 87

Capítulo 1: Introducción

El escenario en el cual se desenvuelven las organizaciones hoy en día está marcado por situaciones como: la globalización, la fuerte competencia, el ritmo acelerado de cambio, la lealtad transitoria del cliente y la variabilidad de la información, por tales motivos es de vital importancia la comprensión del negocio, el enfoque en el cliente, la información como arma competitiva y sobre todo, buenas decisiones en todos los niveles de la organización esto último se apoya en lo que hoy en día se le conoce como Inteligencia de Negocios.

Para ello se construirá el Datamart Establecimientos como parte del Datawarehouse Diners, el cual permitirá a los gerentes y mandos medios tener una herramienta que les permita el análisis de la información y en base a ello tomar las mejores decisiones. Para tal efecto se revisarán conceptos, metodologías y experiencias anteriores las cuales servirán como base para que al término del proyecto se obtenga una solución de calidad.

1.1 Antecedentes

Desde el momento en que se da inicio a la era de las computadoras, las organizaciones han utilizado sus datos desde los sistemas operacionales para atender sus necesidades de información, para ello se utilizaron procesos de carácter típicamente repetitivo o administrativo. Entendiendo por aplicaciones operacionales aquellas que resuelven las necesidades de funcionamiento de la empresa. En este tipo de sistemas, los conceptos más importantes son la actualización y el tiempo de respuesta.

Al tener resueltas las necesidades operacionales más importantes, surge un nuevo tipo de necesidad, aquella que tiene por objetivo, obtener la información necesaria que sirva de base para la toma de decisiones tanto a escala estratégica como táctica.

Es importante, además, una visión histórica de todas las variables analizadas y definidas en la organización, así como de los datos del entorno. Estos requerimientos no son difíciles de obtener, dado que la información necesaria está efectivamente en alguna parte de los sistemas operacionales. Cualquier actividad que realice la empresa, está reflejada minuciosamente en sus bases de datos.

A la hora de atender las necesidades de información para la toma de decisiones, los responsables de sistemas, tropiezan con múltiples problemas teniendo entre los más importantes los siguientes:

• Para tomar decisiones es necesario hacer gestión sobre los datos de la organización, que por lo general, los sistemas operacionales, la presentan en bruto.

• Al realizar consultas masivas de información, se puede ver perjudicado el nivel de servicio del resto del sistema, dado a que tales consultas suelen ser bastante costosas en recursos.

• Las necesidades se ven insatisfechas por la limitada flexibilidad a la hora de navegar por la información y a su inconsistencia debido a la falta de una visión global.

• Encontrar múltiples fuentes de datos, diferentes formatos para la misma información e inconsistencias en los mismos sistemas transaccionales.

Tras las dificultades de los sistemas tradicionales, en dar soporte al proceso de toma de decisiones, surge el concepto de Data Warehouse (DW). El Data Warehouse, provee un ambiente para que las organizaciones hagan un mejor uso de la información que está siendo administrada por diversas aplicaciones operacionales.

A grandes rasgos, un Data Warehouse es una colección o almacén de datos orientado a temas, integrado, no volátil, de tiempo variante, en la cual se encuentra integrada la información de la organización y que se usa como soporte para el proceso de toma de decisiones gerenciales.

Tanto para la eficiencia operativa, como para la planificación a futuro, se deben analizar muchos datos empresariales interrelacionados. Esta necesidad empresarial se aborda mediante el procesamiento analítico. En éste, el enfoque está en el análisis de los datos, específicamente en el análisis multidimensional. En un análisis multidimensional los datos se representan mediante dimensiones como: producto, establecimiento y cliente.

Para que el procesamiento analítico sea óptimo, se utiliza una tecnología llamada OLAP

(Online Analytical Processing) la que crea información empresarial a partir de datos existentes, por medio de un conjunto de transformaciones empresariales y cálculos numéricos.

Usando como base todos estos conceptos, la empresa Diners Club S.A., particularmente el departamento de Establecimientos de tal entidad, quien tiene las tareas de afiliación, capacitación y mantenimiento de establecimientos y comercios, tiene a consideración la implementación del DataMart Establecimientos como parte del Data Warehouse Diners pues en la actualidad no cuenta con una herramienta que le brinde la calidad de información necesaria para tomar buenas decisiones.

1.2 Definición del problema

El departamento de Establecimientos de la empresa Diners Club S.A. cuenta con un gran volumen de datos operacionales, que hace casi imposible no solo el análisis de información que permita el soporte de una mejor toma de decisiones, si no también tener un panorama global que permita conocer la situación del departamento como parte de la empresa y el posicionamiento de la misma frente a los clientes externos, en este caso los clientes y comercios.

1.3 Objetivos

El objetivo general de este proyecto, es brindar una herramienta al área de Establecimientos que ponga a su disposición, información de calidad en forma rápida y eficiente, logrando crear un entorno que permita a los usuarios acceder a la información requerida para el análisis y toma de decisiones de forma fácil, oportuna y flexible.

El objetivo general se apoya de los siguientes objetivos específicos

- Entrega del Plan de Trabajo, en base a la organización del proyecto.
- Entrega del Alcance Funcional, en base al análisis de requerimiento de información.
- Entrega del Modelo Multidimensional y Matriz de Indicadores y Dimensiones, en base al modelamiento multidimensional.
- Documentación de la Construcción de los procesos de extracción, transformación y carga.
- Entrega de reportes y capacitación de explotación de información.
- Cierre del proyecto.

1.4 Justificación

El principal reto de muchas empresas es proporcionar la información adecuada a las personas adecuadas en el momento adecuado, necesidades que el presente proyecto tiene por finalidad. Si a esto le sumamos el hecho que dicha información servirá como soporte a la toma de decisiones tanto estratégicas como tácticas y que estas afectan directa y positivamente a la organización, aumentando el potencial de mejorar la comprensión del negocio y obtener ventajas competitivas, incrementar su habilidad para identificar o generar soluciones posibles a problemas de decisión, y de efectuar oportunamente formulaciones tácticas o estratégicas alineadas con los objetivos de la organización hace que la viabilidad del proyecto quede justificada.

1.5 Propuesta

La propuesta para el desarrollo de la tesina se basa en la investigación de la bibliografía existente acerca del tema para de esta manera ofrecer un amplio marco teórico que permita un mejor entendimiento del objetivo del proyecto. Revisar los conceptos que involucran Soluciones de Inteligencia de Negocios centrándose en la metodología de implementación de desarrollo adaptándola a la realidad de la empresa, así como recoger experiencias de anteriores desarrollos.

1.6 Organización de la tesina

En el Capítulo 2 se aborda el marco teórico que servirá de base para el desarrollo del proyecto.

En el capítulo 3 se expone el estado del arte centrándose en el marco metodológico definido por Ralph Kimball para la implementación de una Solución de Inteligencia de Negocios.

En el capítulo 4 se introduce el caso de estudio, detallando los requerimientos y análisis de los mismos, se procede a los diseños: conceptual y físico. Se definen los lineamientos y mapeos entre el modelo conceptual y las fuentes, además se realiza la implementación de los diseños mencionados, se muestran los procesos de extracción transformación y carga de la información. También se presenta la metodología a utilizar para el diseño e implementación del proyecto tomando como referencia el marco metodológico definido en el capítulo anterior.

En el capítulo 5 se refiere a la plataforma tecnológica y a la solución de inteligencia de negocios.

Finalmente en el capítulo 6 se cometa las conclusiones y los trabajos futuros relacionados con la tesina.

Capítulo 2: Marco Teórico

2. FUNDAMENTO TEÓRICO

Actualmente las empresas están inmersas en entornos cada vez más competitivos, en los cuales es fundamental el disponer de información valiosa con la que se puedan adoptar estrategias empresariales que lleven a situar el negocio por delante de sus competidores. Para este propósito es indispensable disponer de información que permita a los directivos de las empresas tomar decisiones a nivel estratégico de su negocio.

Para ello, las empresas, independiente de su tamaño, cuentan con un conjunto de aplicaciones de procesamiento transaccional que mecanizan los procesos operativos, muy estructurados y repetitivos, que constituyen las funcionalidades básicas de la entidad. En este conjunto de aplicaciones se procesan, de manera automática, grandes volúmenes de datos referentes a las actividades rutinarias, que se almacenan en bases de datos operativas. De ellas se puede extraer información, fundamentalmente válidas para transacciones del día a día, es decir, sirven para apoyar y ejecutar las decisiones operativas que conducen a las actividades básicas, pero no sirven para realizar análisis más avanzados, incluso de tipo estratégico, ya que no están diseñadas para apoyar este tipo de tareas.

A la hora de atender las necesidades de información para la toma de decisiones a nivel estratégico, los sistemas operacionales no son de mucha utilidad. Esto es debido a que estas necesidades se basan en el análisis de una gran cantidad de datos, internos, externos e históricos, en el cual es importante un valor muy detallado del negocio como valor totalizado.

Como solución a los problemas de información de las empresas, a partir de los datos almacenados en las bases de datos operacionales, es posible extraer un cúmulo de datos que aporten un valor añadido a la gestión de la empresa, lo que constituirá el Data Warehouse.

El ingreso de datos en el Data Warehouse viene desde el ambiente operacional en casi todos los casos. El Data Warehouse es siempre un almacén de datos transformados y separados físicamente de la aplicación donde se encontraron los datos en el ambiente operacional. Como apoyo al entendimiento del origen de los datos para la

transformación, es necesario hacer una referencia a la estructuración de las organizaciones con respecto a sus sistemas de información.

2.1 Niveles de la Organización y Necesidad de Información

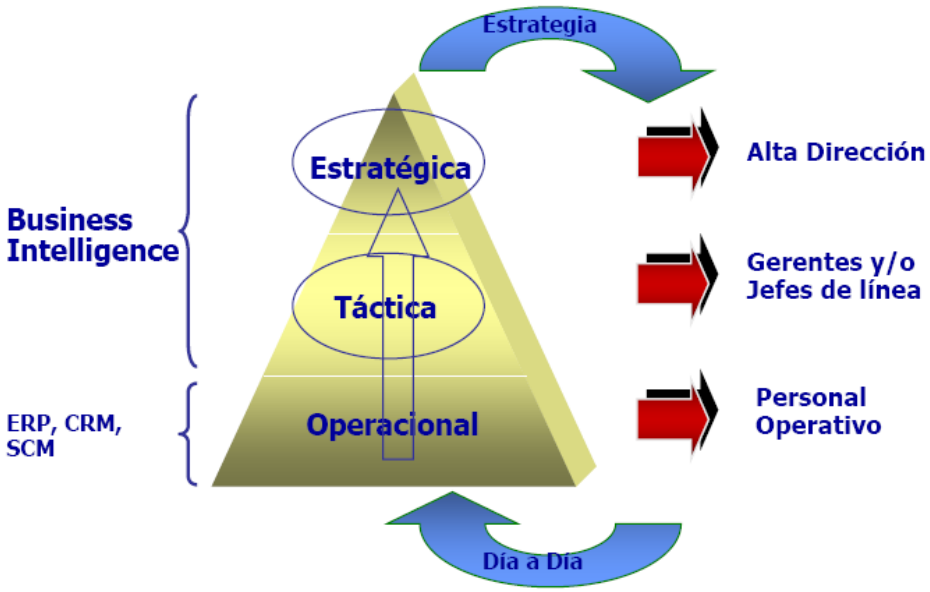


Figura Nº 2.1:"Niveles de la Organización y Necesidades de Información"

2.1.1 Sistemas Estratégicos

Están orientados a soportar la toma de decisiones, facilitan la labor de la dirección, proporcionándole un soporte básico, en forma de mejor información para la toma de decisiones. Se caracterizan porque son sistemas sin carga periódica de trabajo, es decir, su utilización no es predecible, al contrario de los casos posteriores, cuya utilización es periódica.

Destacan entre estos sistemas: los Sistemas de Información Gerencial (MIS), Sistemas de Información Ejecutivos (EIS), Sistemas de Información Georeferencial (GIS), Sistemas de Simulación de Negocios (BIS y que en la práctica son sistemas expertos o de Inteligencia Artificial - AI).

2.1.2 Sistemas Tácticos

Diseñados para soportar las actividades de coordinación y manejo de documentación, definidos para facilitar consultas sobre información almacenada en el sistema,

proporcionar informes y, en resumen, facilitar la gestión independiente de la información por parte de los niveles intermedios de la organización.

Destacan entre ellos: los Sistemas Ofimáticos (OA), Sistemas de Transmisión de Mensajería (Correo electrónico y Servidor de fax), coordinación y control de tareas (Work Flow) y tratamiento de documentos (Imagen, Trámite y Bases de Datos Documentales).

2.1.3 Sistemas Técnico – Operativos

Cubren el núcleo de operaciones tradicionales de captura masiva de datos (Data Entry) y servicios básicos de tratamiento de datos, con tareas predefinidas (contabilidad, facturación, almacén, presupuesto, personal y otros sistemas administrativos). Estos sistemas están evolucionando con la irrupción de censores, autómatas, sistemas multimedia, bases de datos relacionales más avanzadas y data warehousing.

2.2 Gerentes de la Organización

Asimismo, podemos clasificar dentro de una organización, a los gerentes o los cargos de alto mando, encargados de tomar decisiones. La forma de clasificar a los gerentes en las organizaciones estructuradas tradicionalmente, es decir, las que cuentan con estructuras de trabajo deliberadas o estructuradas configuradas como una pirámide, se muestra en la figura Nº 2.2. [ROB2000].

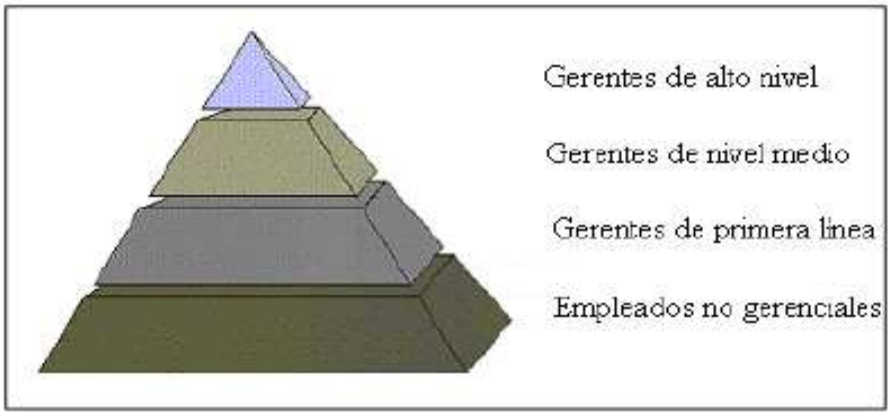


Figura Nº 2.2:"Clasificación de los Gerentes en las Organizaciones" [ROB2001]

Aquí se distinguen los gerentes de primera línea, que ocupan el nivel más bajo de la gerencia, y con frecuencia cumplen con labores como la de supervisión.

Los gerentes de nivel medio, pueden tener diversos títulos, como jefe de departamento o agencia, líder de proyecto, gerente de planta, jefe de unidad, decano, obispo o gerente de división. En la cumbre de la organización, o cerca de ella, están los gerentes de alto nivel, responsables de tomar las decisiones y establecer las políticas y estrategias que afectan a toda la organización. Es precisamente a estos individuos, y a los gerentes del nivel medio, a los cuales está dirigido el propósito de este proyecto: facilitarles el proceso de toma de decisiones.

2.3 Data Warehouse

Los Data Warehousing, nacen debido a la necesidad de contar con información de apoyo a la toma de decisiones, dado que los datos operacionales no cumplen eficientemente con este objetivo. Esto conlleva a desarrollar Data Warehouses para poner el uso estratégico de la información como generador de ventajas competitivas.

Para entender el concepto Data Warehouse, daremos tres definiciones de autores diferentes:

"Un Data Warehouse es una copia de los datos de las transacciones de una organización, estructurados específicamente para la pregunta y el análisis". [KIM1996]

"Un Data Warehouse es algo que provee dos beneficios empresariales reales: Integración y Acceso de Datos. El Data Warehouse elimina una gran cantidad de datos inútiles y no deseados, como también el procesamiento desde el ambiente operacional clásico" [OST1993]

"Un Data Warehouse es una colección de datos integrados, temáticos, no volátiles y variantes en el tiempo, organizados para soportar necesidades empresariales orientadas a la toma de decisiones". [INM2002]

La primera definición, no deja muy claro el concepto de Data Warehouse, ya que da una idea no muy clara ni muy global del concepto.

La segunda definición refleja claramente el principal beneficio que el Data Warehouse aporta a una organización, eliminar aquellos datos que obstaculizan la labor de análisis

y entregar la información que se requiere en la forma más apropiada, facilitando así el proceso de gestión.

Para este proyecto, profundizaremos la tercera definición dada por Inmon, considerado el padre del Data Warehouse, por su sencillez, claridad y fácil comprensión.

Podemos concluir, que un Data Warehouse, es el proceso de extraer y filtrar datos de las operaciones comunes de las empresas, procedentes de los distintos subsistemas operacionales, para transformarlos, integrarlos, sumariarlos y almacenarlos en un depósito o almacén de datos, para poder acceder a ellos cada vez que se necesite mediante mecanismos flexibles para el usuario

2.3.1 Características

Siguiendo con la definición de Inmon, un Data Warehouse se caracteriza por ser [INM2002]:

2.3.1.1 Integrado

El aspecto más importante en los Data Warehouses es que la información encontrada siempre está integrada. La integración de los datos se muestra de muchas maneras: en conversiones de nombres consistentes, en la medida uniforme de las variables, en la conversión de estructuras consistentes, en atributos físicos de los datos consistentes, fuentes múltiples, entre otros. Los datos almacenados en el Data Warehouse, deben integrarse en una estructura consistente, por lo que las inconsistencias existentes en los diversos sistemas operacionales, deben ser eliminadas. La información suele estructurarse también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

Como se muestra en la Figura Nº 2.3 un Data Warehouse integra datos recogidos de diferentes sistemas operacionales de la organización y/o fuentes externas.

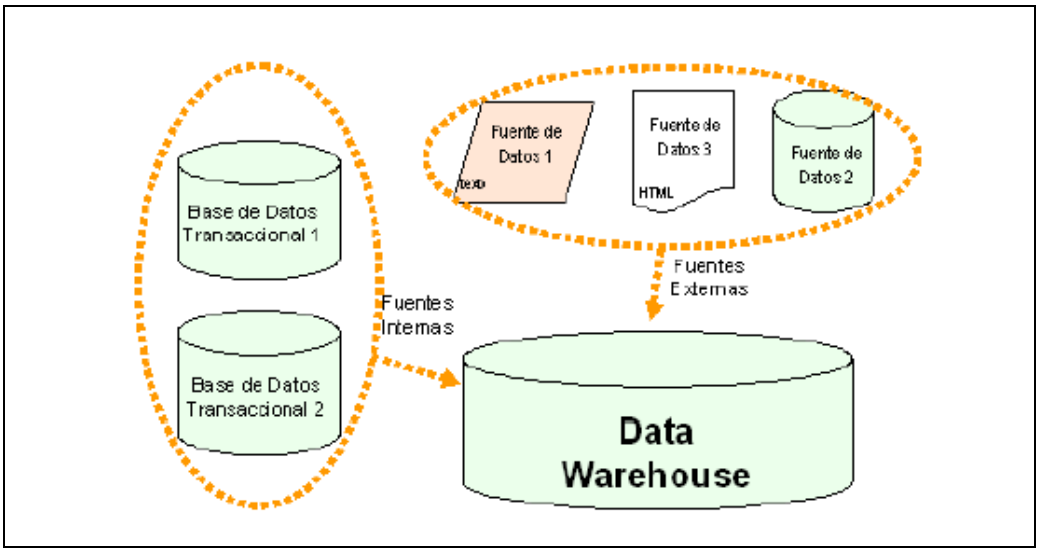


Figura Nº 2.3: "Data Warehouse Integrado"

2.3.1.2 Temáticos

Sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, como se aprecia en la figura Nº 2.4, todos los datos sobre *establecimientos* pueden ser consolidados en una única tabla del Data Warehouse. Así, las peticiones de información sobre establecimientos serán más fáciles de responder dado a que toda la información está en el mismo lugar.

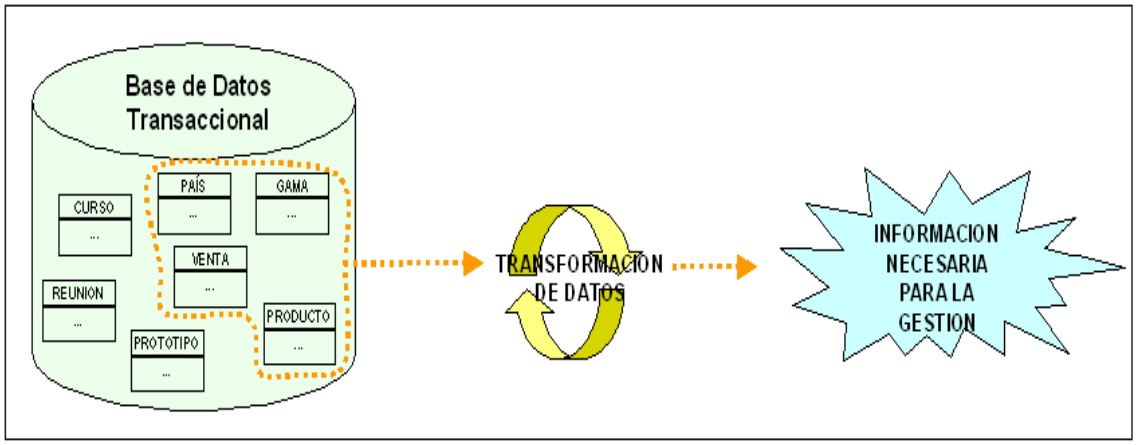


Figura Nº 2.4: "Data Warehouse Temático"

2.3.1.3 No Volátil

La información es útil solo cuando es estable. Los datos operacionales cambian sobre una base momento a momento. La perspectiva más grande, esencial para el análisis y la toma de decisiones, requiere una base de datos estable. En la figura Nº 2.5, se puede ver que la actualización (insertar, actualizar, borrar), se hace en el ambiente operacional. Pero la manipulación básica de los datos que ocurre en el Data Warehouse son: la carga inicial de los datos y el acceso a los mismos. Por lo tanto, la información de un Data Warehouse existe para ser leída, y no modificada, es decir mas datos son agregados pero los datos existentes no son removidos. Es importante mencionar que el periodo de tiempo cubierto por un Data Warehouse varía entre 2 y 10 años. [HER2003]

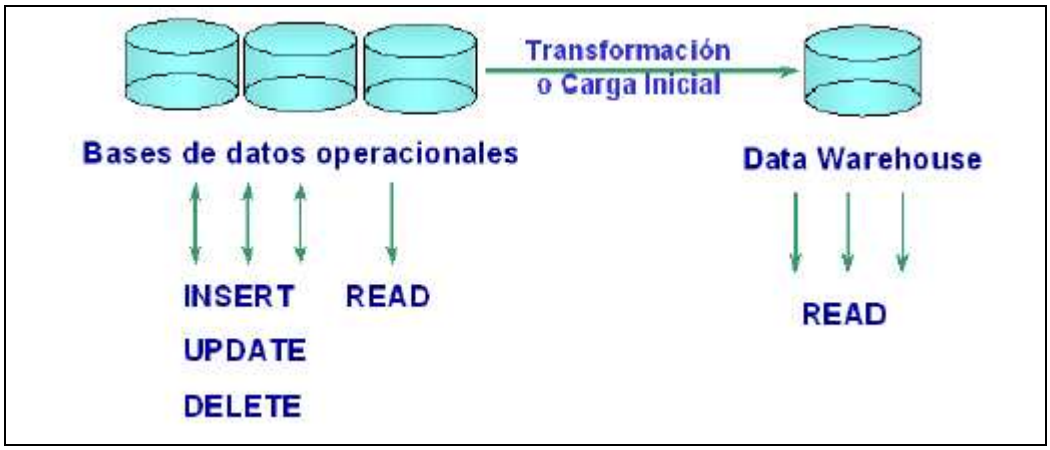


Figura Nº 2.5: "Data Warehouse No Volátil"

2.3.1.4 Variante en el Tiempo

El tiempo es la parte implícita de la información en un Data Warehouse. En los sistemas operacionales, los datos siempre reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información contenida en el Data Warehouse sirve, entre otras cosas, para realizar análisis de tendencias. Por lo tanto, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.

2.3.2 Problemas y Ventajas del Data Warehouse

Los Data Warehouse convierten los datos operacionales de una organización en una herramienta competitiva, que permite a los usuarios finales (gerentes de alto nivel, o nivel medio) examinar los datos de modo más estratégico, realizar análisis y detección de tendencias, seguimiento de medidas críticas, producir informes con mayor rapidez, un acceso más fácil, más flexible y más intuitivo a la información que se necesite en cada momento. Frecuentemente, datos que son difíciles de interpretar, desde varias fuentes, se convierten en información lista para el usuario final.

Estas ventajas competitivas para el negocio, se pueden traducir en: [SAS2003]

• Proporciona una herramienta para la toma de decisiones en cualquier área funcional, basándose en información integrada y global del negocio.

• Facilita la aplicación de técnicas estadísticas de análisis y modelización 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.

• Simplifica dentro de la empresa la implantación de sistemas de gestión integral de la relación con el cliente.

• Supone una optimización tecnológica y económica en entornos de Centro de Información, estadística o de generación de informes con retornos de la inversión espectaculares.

Aun con los beneficios que se incorporan a la organización, junto con la implantación de un Data Warehouse, también se pueden presentar ciertos problemas, que según José Hernández Orallo, son: [HER2003]

• Infravaloración del esfuerzo necesario para su diseño y creación.

• Infravaloración de los recursos necesarios para la captura, carga y almacenamiento de los datos.

• Incremento continuo de los requisitos de los usuarios.

• Privacidad de los datos.

2.3.3 Diferencias entre Sistema Tradicional (OLTP) y Data Warehouse

En cuanto a los requerimientos de diseño y características de implementación, los Sistemas Tradicionales (OLTP, Online Transactional Processing) y los Data Warehouse poseen muchas diferencias, las que posemos resumir en la tabla Nº 2.1

SISTEMA TRADICIONAL	DATA WAREHOUSE
Predomina la actualización	Predomina la consulta
La actividad más importante es del tipo operativo(día a día)	La actividad más importante es el análisis y la decisión estratégica
Predomina el proceso puntual	Predomina el proceso masivo
Mayor importancia a la estabilidad	Mayor importancia al dinamismo
Datos en general desagregados	Datos en distintos niveles de detalle y agregación
Importancia del dato actual	Importancia del dato histórico
Importancia del tiempo de respuesta de la transacción instantánea	Importancia de la respuesta masiva
Estructura relacional	Visión multidimencional
Usuarios de perfiles medios o bajos	Usuarios de perfiles altos
Explotación de la información relacionada con la operativa de cada aplicación	Explotación de toda la información interna y externa del negocio

Tabla Nº 2.1. "Diferencias entre Sistema Tradicional versus Data Warehouse. [SAS2003]

2.3.4 Arquitectura de un Data Warehouse

Una de las razones por las que el desarrollo de un Data Warehouse crece rápidamente, es que es una tecnología muy entendible. Un Data Warehouse puede representar de mejor forma la amplia estructura de una empresa, para administrar los datos que serán útiles para la toma de decisiones dentro de la organización. A fin de comprender cómo se relacionan todos los componentes involucrados en un Data Warehouse, es esencial analizar su arquitectura.

• Arquitectura real: Es generalmente la arquitectura elegida para los sistemas de decisión. El almacenamiento de los Data Warehouses se realiza en un SGBD (Sistema Gestor de Base de Datos) separado del sistema de producción. Este SGBD se alimenta

por extracciones periódicas. Antes de la carga, los datos sufren importantes procesos de integración, limpieza y transformación. La ventaja de esta solución es que se dispone de datos preparados para las necesidades de la decisión que responden bien a los objetivos del Data Warehouse. La principal razón para justificar la arquitectura real es la inadaptación de los datos de producción a las necesidades de los sistemas de decisión. Las estructuras de datos en un sistema de producción son complejas en cuanto a almacenamiento e implican herramientas de programación para acceder a ellas. Los datos también están codificados. En un contexto de ayuda a la decisión, el dato debe ser comprensible por el usuario. Es necesario transformar todos los códigos legibles y comprensibles. Las cargas relacionadas con los accesos son también incompatibles con los objetivos de rendimiento del sistema de producción. Los datos están dispersos. El Data Warehouse debe integrar los datos unos con los otros a fin de asegurar una coherencia global. [FRA1997]

• Arquitectura virtual: En una arquitectura virtual, los datos del Data Warehouse residen en el sistema tradicional. Se hacen visibles por productos middleware o pasarelas. En esta arquitectura no hay coste de almacenamiento suplementario y el acceso se hace en tiempo real. Sin embargo, las numerosas desventajas de este tipo de arquitectura impiden frecuentemente su elección. Los datos no son preparados. El apartado anterior muestra la dificultad real que presenta la utilización de los datos de la empresa. Los accesos de decisión pueden perturbar el rendimiento del sistema de la empresa, tanto más cuanto que los procesos de transformación y de integración están aquí relacionados forzosamente con los procesos de acceso. Según los antecedentes recopilados en diferentes fuentes, no existe ninguna implementación de Data Warehouse virtual. [FRA1997]

• Arquitectura remota: La arquitectura remota es una combinación de los dos tipos de arquitectura descritas anteriormente. El objetivo es implementar físicamente los niveles agregados (los niveles de datos utilizados más a menudo) a fin de facilitar el acceso y conservar el nivel de detalle en los sistemas de producción dando acceso por medio de

middleware o de gateways. Esta arquitectura se utiliza también muy raramente.

[FRA1997]

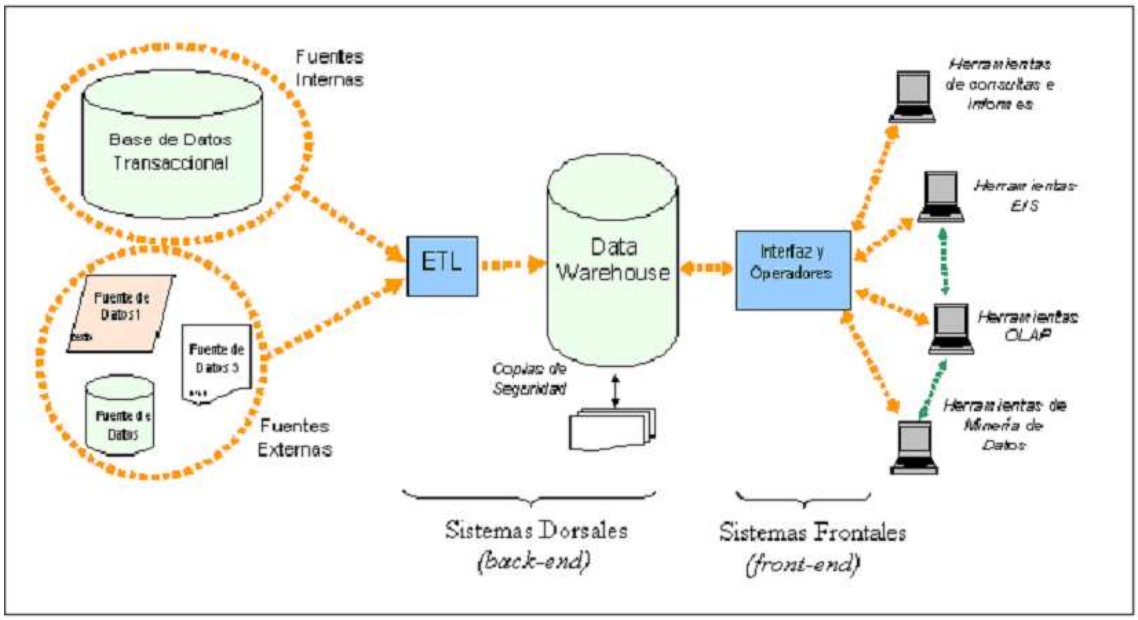


Figura Nº 2.6: "Arquitectura real de un Data Warehouse"

En la figura Nº 2.6, se muestra la arquitectura de un Data Warehouse, que viene determinada por su situación central como fuente de información para las herramientas de análisis. En la figura se pueden distinguir dos tipos de sistemas: Sistemas Dorsales (conocidos como backend) y los Sistemas Frontales (o front-end tools).

2.3.4.1 Sistemas Dorsales (back-end)

Los Sistemas Dorsales, que son herramientas de adquisición de datos de fuentes externas, se encargan principalmente de extracción, limpieza, transformación, integración, carga y refresco de los datos. Dentro de la clasificación de sistemas dorsales, podemos mencionar los sistemas ETL y el Data Warehouse mismo, que se describen a continuación. Los Sistemas ETL (de las siglas en inglés Extraction, Transformation, Load), que se encargan de las funciones de extracción de las distintas fuentes de datos, sean estas transaccionales o externas; transformación, realizando las tareas de limpieza y consolidación de los datos y la carga del Data Warehouse.

Entre las principales funciones de los sistemas ETL, se pueden enunciar: [HER2003]

• La extracción de los datos.

• El filtrado de los datos: limpieza, consolidación, etc.

• La carga inicial del Data Warehouse: ordenación, agregaciones, etc.

• Refresco del Data Warehouse: operación periódica que propaga los cambios de las fuentes externas al Data Warehouse mismo Repositorio Propio de Datos (Data Warehouse), que posee información relevante de los datos o metadatos.

Los metadatos contienen información relativa a los datos. Comúnmente, los metadatos permiten mantener información sobre: [SAS2003] [INC2001]

• La procedencia de los datos

• La periodicidad de refresco de los datos

• La fiabilidad y forma de cálculo de los datos

• Semántica de los datos y su localización en el DW.

• Localización de los datos en los sistemas de producción y reglas de transformación.

• Especificación de fórmulas de cálculo de agregados.

• Información sobre frecuencias de carga, mecanismo de historización, etc.

2.3.4.2 Sistemas Frontales (front-end)

Los sistemas frontales, o herramientas front-end, ofrecen al usuario final mecanismos eficaces de acceso a la información mediante mecanismos simples y potentes, y así poder tener una conexión eficaz con el Data Warehouse. Cumplen con las tareas de realizar consultas, informes, análisis y extracción del conocimiento (data mining). Dentro de esta clasificación podemos mencionar las Interfaces y Gestores de Consultas y los Sistemas de Integridad y Seguridad.

Las Interfaces y Gestores de Consulta permiten acceder a los datos y sobre ellos conectar herramientas más sofisticadas como OLAP (Online Analytical Processing, se profundizará el estudio de OLAP en una sección posterior), EIS, minería de datos (data mining).

2.4 Modelo de Datos

Un modelo de datos es un sistema formal y abstracto que permite describir los datos de acuerdo con reglas y convenios predefinidos. Es formal pues los objetos del sistema se manipulan siguiendo reglas perfectamente definidas y utilizando exclusivamente los

operadores definidos en el sistema, independientemente de lo que estos objetos y operadores puedan significar [ULL1999]. Para entender una de las partes principales de la arquitectura del Data Warehouse, como es el modelo de datos, es necesario conocer primero la diferencia que existe entre lo modelos existentes: Modelo Relacional y Modelo multidimensional.

2.4.1 Modelo Relacional

El modelo relacional puede considerarse como un lenguaje de programación más bien abstracto, orientado de manera específica hacia las aplicaciones de bases de datos. En términos tradicionales una relación se asemeja a un archivo, una tupla a un registro, y un atributo a un campo. Pero estas correspondencias son aproximadas, en el mejor de los casos. Una relación no debe considerarse como "solo un archivo", sino más bien como un archivo disciplinado, siendo el resultado de esta disciplina una simplificación considerable de las estructuras de datos con las cuales debe interactuar el usuario, lo cual a su vez simplifica los operadores requeridos para manejar esas estructuras.

El modelo relacional es el pilar fundamental para el diseño de la mayoría de las bases de datos que existen hoy en las grandes y pequeñas empresas. La composición de estas bases de datos son decenas de tablas relacionadas a través de una compleja tela de araña de uniones.

La implementación de bases de datos con un modelo relacional da lugar a escenarios como los siguientes [FER2001]:

• Legibilidad limitada: Los usuarios finales no son capaces de entender el modelo entidad relación. Evidentemente no pueden navegar por dicho modelo en busca de información.

• Dificultad para las herramientas de consulta en el acceso a un modelo entidad relación general: Los optimizadores de las bases de datos no siempre realizan la selección correcta y a menudo adolecen de prestaciones mediocres o inaceptables cuando trabajan en entornos de grandes volúmenes de información.

• La utilización de la técnica del modelado relacional frustra el principal atractivo de Data Warehouse, que es, la recuperación de información intuitiva y con alto rendimiento.

2.4.2 Modelo Multidimensional

El modelo multidimensional es una técnica para modelar bases de datos simples y entendibles al usuario final, ya sea, por presentar la información en un marco estándar e intuitivo que permitan un acceso de alto rendimiento.

Los objetivos del modelo multidimensional son [PER2001]:

• Representar los datos en forma cercana a la intuición del usuario.

• Resolver problemas planteados en sistemas relacionales.

Este tipo de modelado es independiente de las tecnologías y permite el empleo de cualquier base de datos, ya sea multidimensional, MOLAP, ésta almacena y manipula los datos en estructuras especiales (matrices multidimensionales); Base Datos relacional, ROLAP, la que almacena los datos en un SGBD relacional extendido y transforma operaciones sobre datos multidimensionales en operaciones relacionales en SQL; Base de datos objetos, etc.

2.4.2.1 Componentes del Modelo Multidimensional

El Modelo Multidimensional está compuesto principalmente por dos elementos: esquemas y tablas.

Las tablas en el Modelo Multidimensional se dividen en dos tipos:

• Tablas Hechos: Es el objeto a analizar, además posee atributos llamados atributos de hechos o síntesis, estos atributos son de tipo cuantitativo cuyos valores (cantidades) se obtienen generalmente por aplicación de una función estadística que resume un conjunto de valores en un único valor. Además contienen funciones resumen, funciones de tipo estadístico que se aplican a los atributos de hecho.

• Tablas Dimensiones: Representan cada uno de los ejes en un espacio multidimensional. Como todas las tablas también poseen atributos llamados de dimensión o de clasificación, estos son de tipo cualitativo (sus valores son modalidades) que suministran el contexto en el que se obtienen las medidas en un esquema de hecho. Las dimensiones poseen jerarquías, que son varios atributos unidos mediante una

relación de tipo jerárquico. Un esquema multidimensional se puede representar como un esquema estrella (star join), copo de nieve (snowflake) o Constelación de hechos.

Esquema estrella: Formado por una tabla de hechos con una única tabla para cada dimensión, como lo muestra la figura Nº 2.7. El esquema estrella deriva su nombre del hecho que su diagrama forma una estrella, con puntos radiales desde el centro. El centro de la estrella consiste de una o más tablas fact, y las puntas de la estrella son las tablas lock_up.

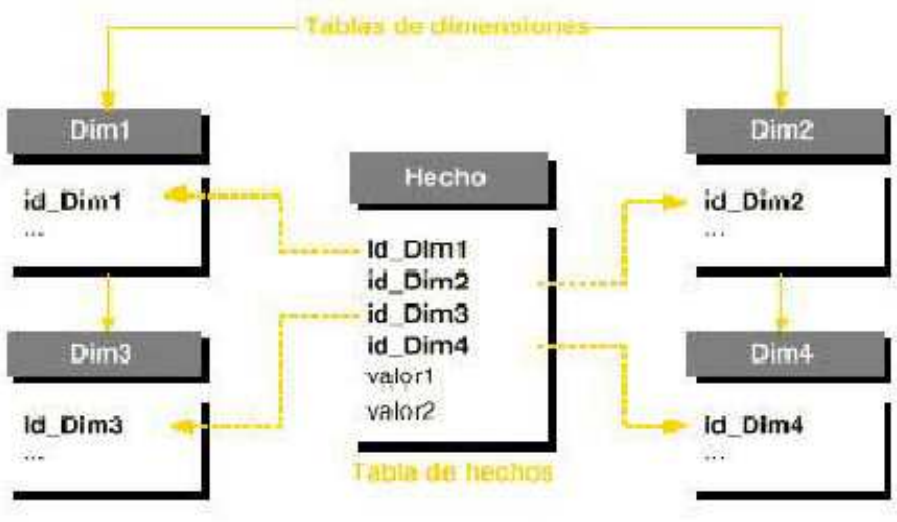


Figura Nº 2.7: "Esquema Estrella". [DIA2002]

Esquema copo de nieve: La diferencia del esquema snowflake comparado con el esquema estrella, está en la estructura de las tablas lock_up: las tablas lock_up en el esquema snowflake están normalizadas. Cada tabla lock_up contiene sólo el nivel que es clave primaria en la tabla y la foreign key de su parentesco del nivel más cercano del diagrama como se puede observar en la figura Nº 2.8.

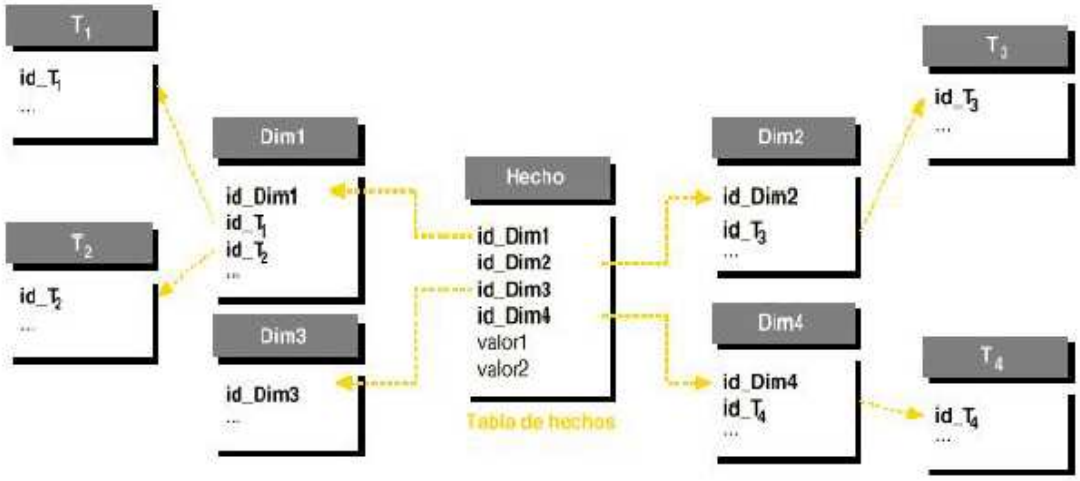


Figura Nº 2.8: "Esquema Copo de Nieve". [DIA2002]

Constelación de Hechos: es un conjunto de tablas de hechos que comparten algunas tablas de dimensiones, como lo muestra la figura Nº 2.9.

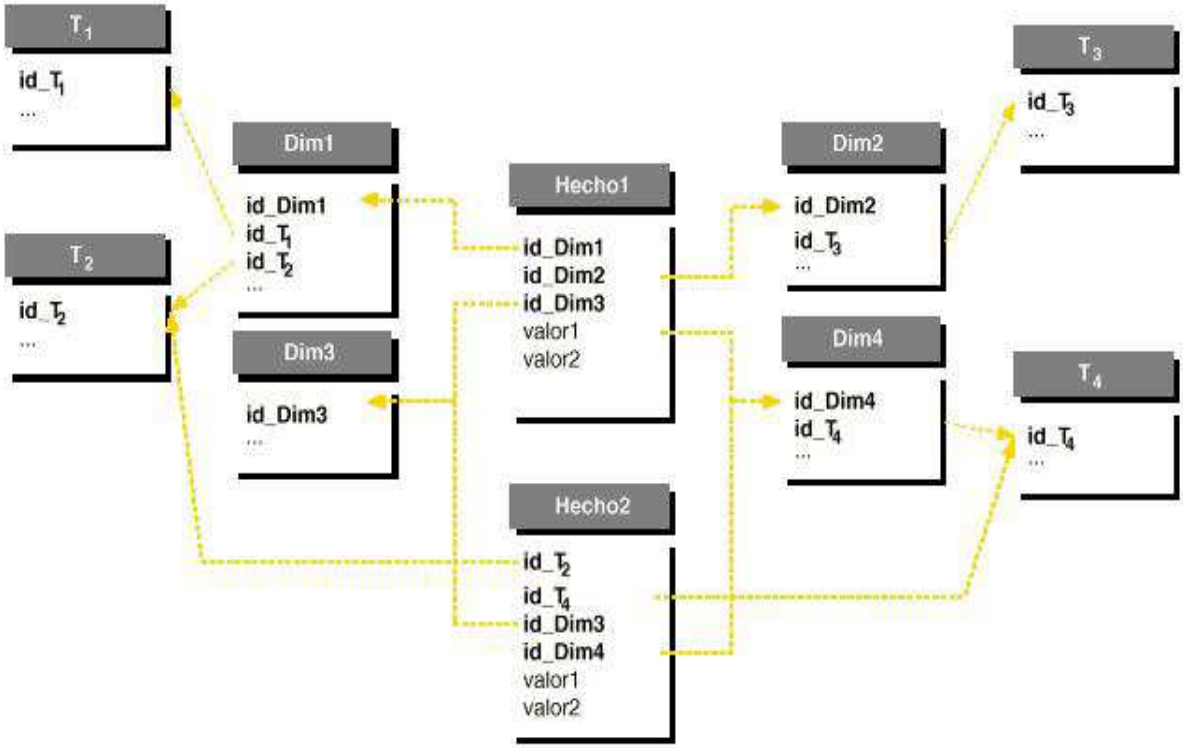


Figura Nº 2.9: "Constelación de Hechos" [DIA2002]

2.4.2.2 Estructura Básica de las Bases de Datos Multidimensionales.

Un modelo multidimensional soporta el manejo de una vasta cantidad de datos empresariales y temporales. De esta forma surge la instancia del modelo multidimensional, también conocido como Cubo o Hipercubo, en la cual podemos encontrar en interacción los conceptos mencionados en estos párrafos.

En una base de datos multidimensional la información se representa como una matriz multidimensional, en los ejes de esta matriz se encuentran los criterios de análisis, en los cruces están los valores a analizar. Como ejemplo de Cubo encontramos el que nos muestra la figura Nº 2.10, la cual contiene tres dimensiones: modelo, vendedor y color; donde cada dimensión tiene tres diferentes niveles o hechos, para finalmente encontrar la medida al interceptar los valores.

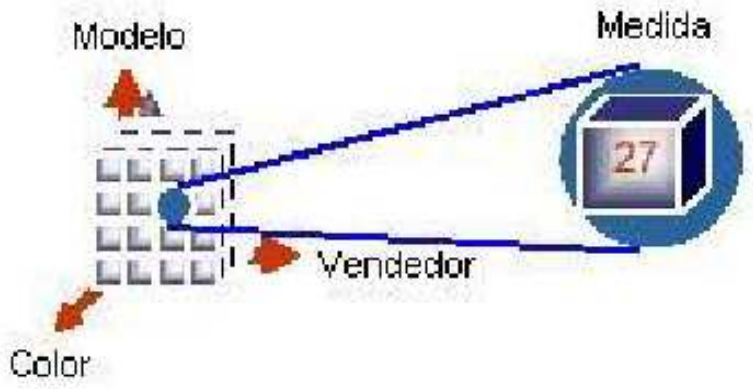


Figura Nº 2.10: *"Ejemplo de un Cubo con tres dimensiones"*. [YDI2004]

Los Cubos o Hipercubos están formados por:

• Dimensiones: Las dimensiones son un concepto esencial de bases de datos multidimensionales. Las dimensiones representan los criterios de análisis de los datos, macro-objetos del problema y variables independientes. Si una dimensión tiene más de un nivel entonces los miembros de la dimensión pueden ser organizados en una o más jerarquías.

• Medida: Las medidas son un dato numérico que representa una actividad específica de un negocio, mientras que una dimensión representa una perspectiva de los datos. Cada dimensión está descrita por un conjunto de atributos (datos agregados). A su vez se pueden interceptar estas dimensiones para obtener un valor, llamado medida.

Una medida contiene una propiedad numérica y una fórmula. Existen tres clases de medidas:

• Medidas aditivas: Pueden ser combinadas a lo largo de cualquier dimensión.

• Medidas semi-aditivas: No pueden ser combinadas a lo largo de una o más dimensiones.

• Medidas no aditivas: N er combinadas a lo largo de ninguna dimensión

2.4.2.3 Operaciones en Modelos Multidimensionales

Las principales operaciones en los modelos multidimensionales, son:

• Slice & Dice: Éste nos permite hacer una selección de valores de las dimensiones que uno requiere. En la figura Nº 2.11 se puede observar un ejemplo de este tipo en el cual el usuario solo necesita regiones y productos en un tiempo específico.

Figura Nº 2.11: "Slice & Dice". *[YDI2004]*

• Rotación: Este operador selecciona el orden de visualización de las dimensiones. En la figura Nº 2.12 se muestra un ejemplo de ello.

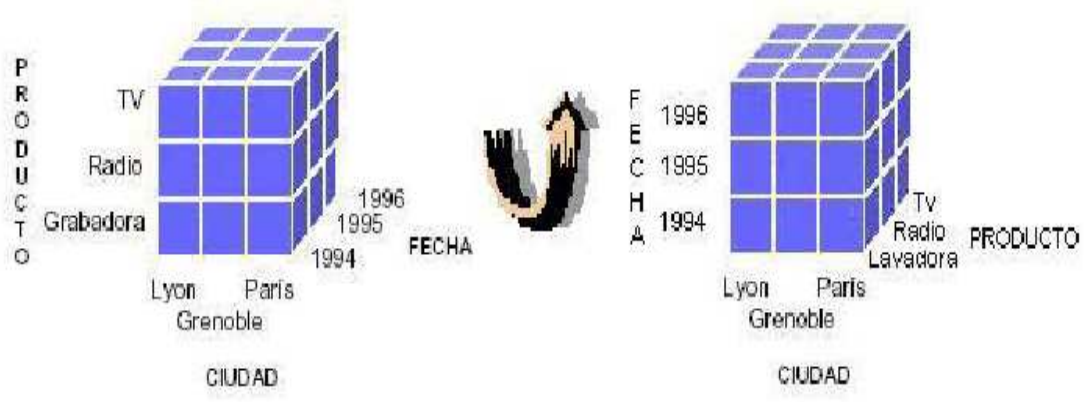


Figura Nº 2.12: "Rotación". *[YDI2004]*

• Movimientos en la Jerarquía de una dimensión (Drill Down, Drill Up): Este operador nos permite bajar o subir a los niveles más atómicos de nuestro esquema multidimensional. En la figura Nº 2.13 se puede ver como se ajustan las escalas de los ejes, produciendo movimientos en la jerarquía.

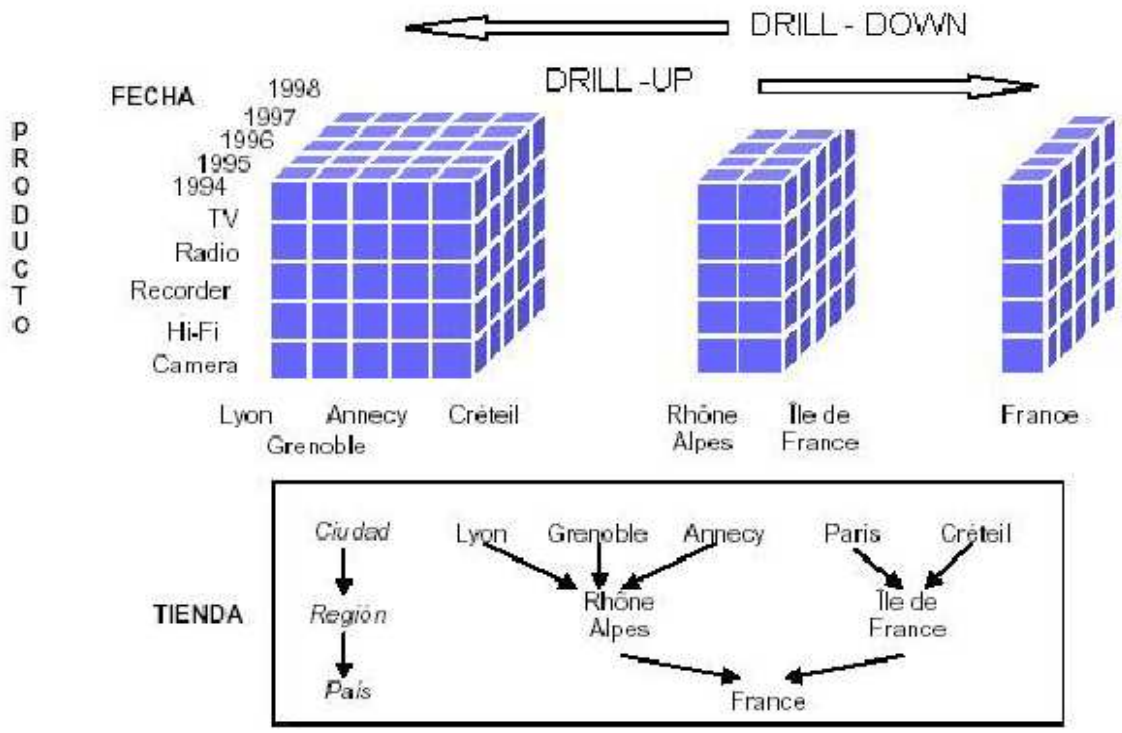


Figura Nº 2.13: "Drill-Down y Drill-Up". *[YDI2004]*

• Consolidación (Roll-Up): Este operador nos permite hacer cálculos a las medidas en función de agrupamientos. Realiza el recálculo de la medida de acuerdo a los ajustes de escala, además se debe especificar cuál es la operación que calcula el nuevo valor de la medida. Esta operación pueden ser: suma, promedio, etc. También puede haber medidas con comportamientos diferentes. Por ejemplo en la figura Nº 2.14 se realiza un Roll – Up suma sobre la dimensión Tienda.

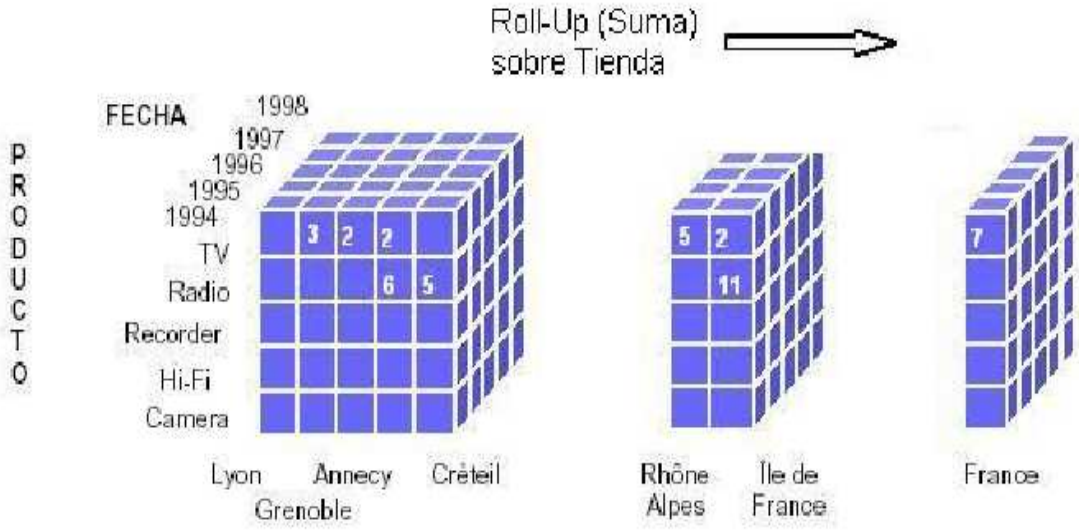


Figura Nº 2.14: "Roll-Up". [YDI2004]

• Drill-Across: Operador que relaciona dos cubos, como lo muestra la figura Nº 2.15.

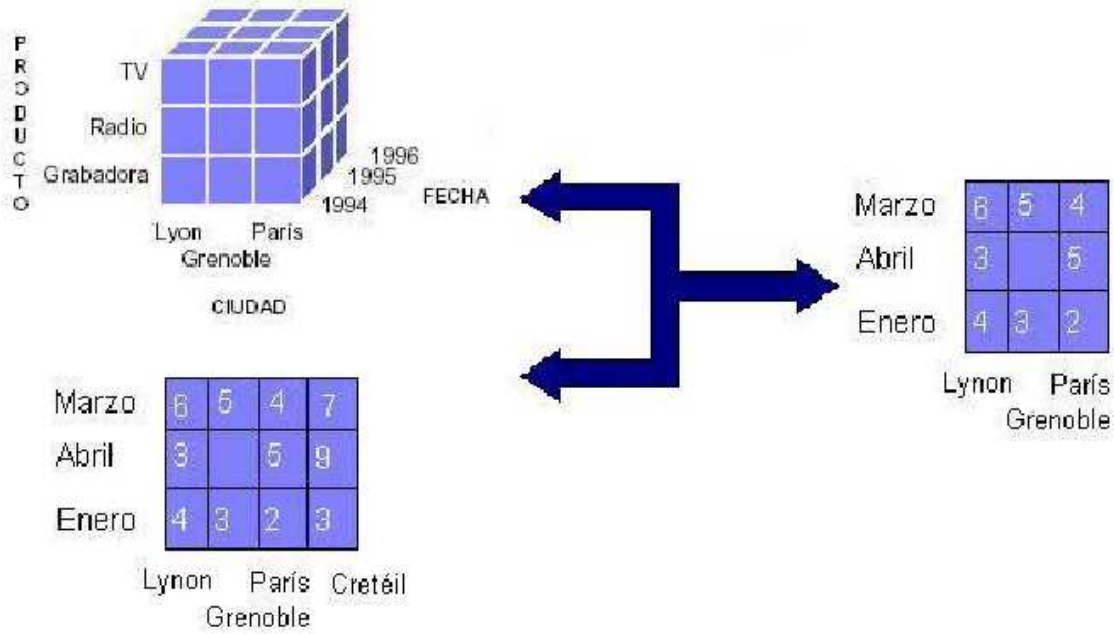


Figura Nº 2.15: "Drill-Across". [YDI2004]

• Drill-Through: Operador que accede a datos descriptivos. Un ejemplo se muestra en la figura Nº 2.16.

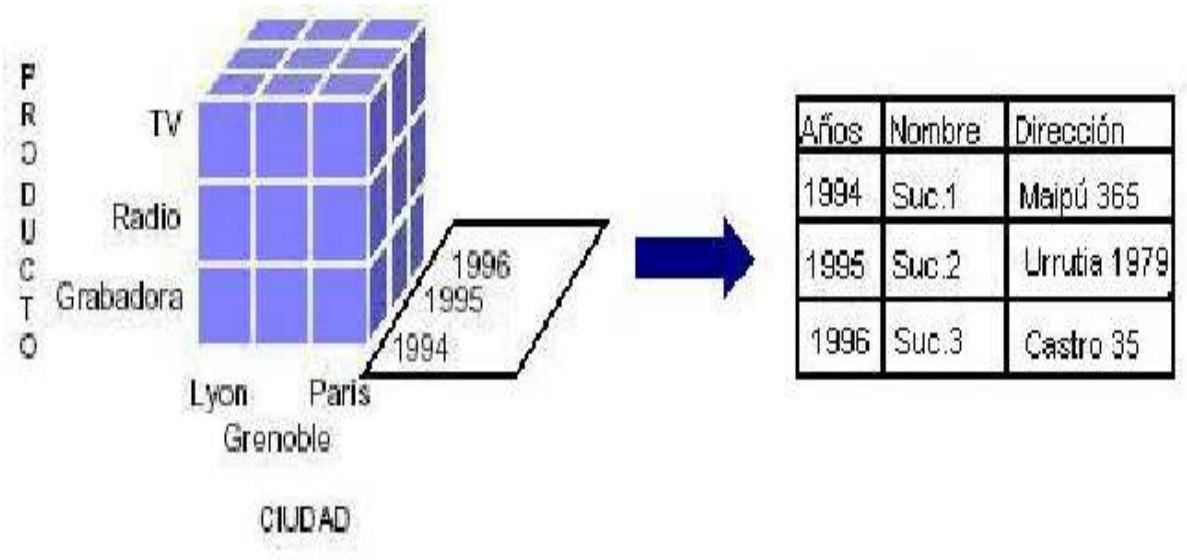


Figura Nº 2.16: "Drill-Trough*"*

2.4.2.4 Ventajas del Modelo Multidimensional

Implementar un Data Warehouse a través del modelo multidimensional tiene un número importante de ventajas:

• Primero, es un modelo es una estructura homogénea y predecible. Reportes escritos, herramientas de consulta e interfaces de usuario, todas pueden hacer suposiciones acerca del modelo multidimensional para que las interfaces del usuario sean más entendibles y los procesos, más eficientes.

• Una segunda ventaja del modelo multidimensional reside en que es una estructura predecible. Cada dimensión es equivalente. Todas las dimensiones pueden ser vistas como un conjunto de puntos igualmente simétricos dentro de una tabla de hechos.

• Una tercera ventaja del modelo multidimensional, es que es extensible para acomodar nuevos elementos de datos inesperados y nuevas decisiones de diseño. Esto se lleva a cabo añadiendo nuevas dimensiones, nuevos atributos dimensiónales y cambiando datos de una cierta granularidad para pasarlos a otra.

• Una última ventaja del modelo multidimensional es la creciente cantidad de herramientas administrativas y de software que manejan y usan la granularidad.

2.5 OLAP (Online Analytical Processing)

OLAP (Proceso Analítico en Línea) desarrolla un análisis multidimensional de los datos, que consiste en organizar la información según los parámetros que la alta dirección considere oportuno para darles un sentido y llevar a cabo el análisis de los mismos.

Para un concepto más formal de OLAP, citamos a dos autores, quienes definen OLAP como:

"Es una categoría de tecnología de software que permite a los ejecutivos y analistas, una vista rápida, consistente, con un acceso interactivo a una completa variedad de vistas posibles de la información que ha sido transformada de los registros de una empresa".

"Se trata de un término inventado para describir una aproximación dimensional interactiva al soporte de toma de decisiones (análisis desde la perspectiva de sus componentes o dimensiones, contemplando también los distintos niveles o jerarquías que éstas poseen): Proceso, Dimensiones, Interacción, Análisis, Toma de decisiones. Este tipo de Análisis, se soporta mediante una visión dimensional: ***Modelo Multidimensional o Cubo".***

Como se describe en la definición anterior, comúnmente los sistemas OLAP ofrecen a los empleados calificados una estructura de datos conocida como un "cubo", que virtualmente contiene niveles de reportes resumidos actualizados. Este cubo es multidimensional y ofrece a los usuarios finales diferentes alternativas y perspectivas de ver los datos, a partir del mismo cubo.

Por ejemplo, un cubo que contiene datos de ventas puede producir reportes tan diversos como: Ventas totales por trimestre, porcentaje de ventas por vendedor, ventas promedio por región geográfica, etc.

Las herramientas OLAP se caracterizan por una rápida recuperación de los datos, ya que la recuperación se puede realizar en una sola entrada y salida, los índices son pequeños y los datos se encuentran modelados en un esquema multidimensional. Además, el

almacenamiento es muy eficiente, ya que los bloques solo contienen datos y los índices son simples.

2.5.1 Herramientas OLAP

Las herramientas de OLAP presentan al usuario una visión multidimensional de los datos (esquema multidimensional) para cada actividad que es objeto de análisis. El usuario formula consultas a la herramienta OLAP seleccionando atributos de este esquema multidimensional sin conocer la estructura interna (esquema físico) del almacén de datos. La herramienta OLAP genera la correspondiente consulta y la envía al gestor de consultas del sistema.

Una consulta a un Data Warehouse consiste generalmente en la obtención de medidas sobre los hechos parametrizadas por atributos de las dimensiones y restringidas por condiciones impuestas sobre las dimensiones [HER2003].

Utilizando herramientas OLAP, los usuarios pueden acceder al DW, brindando a los responsables de las tomas de decisiones de las organizaciones, el potencial de mejorar su comprensión del negocio y los cambios que lo afectan, de incrementar su habilidad para identificar o generar soluciones posibles a problemas de decisión, y de efectuar oportunamente formulaciones tácticas o estratégicas alineadas con los objetivos de la organización. Esta característica de analizar y sintetizar la información a partir de OLAP, surge de la elaboración de múltiples escenarios que contestan preguntas tales como “qué pasaría si...” o “por qué...”.

En estos verdaderos modelos de simulación, y a partir de la modificación de ciertas variables clave, se analiza el comportamiento del resto de las variables.

Lo realmente interesante de las herramientas OLAP son sus operadores de refinamiento o manipulación de consultas:

• Drill

• Roll

• Slice & Dice

• Pívot

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

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

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

Las operaciones de agregación (roll) y disgregación (drill) se pueden hacer sobre atributos de una dimensión sobre los que se ha definido una jerarquía: (drill-down, roll-up) como por ejemplo: Departamento - Categoría - Producto (Producto), Año - Trimestre - Mes – Día (Tiempo). O sobre dimensiones independientes (drill-across, roll-across). Por ejemplo: Producto - Almacén – Tiempo.

Slice & Dice: permite seleccionar y proyectar datos en el informe.

Pivot: permite la reorientación de las dimensiones en el informe.

2.5.2 Características de las Herramientas OLAP

En 1994, Codd y Codd introdujeron reglas sobre el modelo OLAP, pensadas para eliminar interpretaciones erróneas sobre las características de esta tecnología. En forma resumida, las doce reglas de Codd para los sistemas OLAP son las siguientes [MAR2001]

1. Vistas Multidimensionales. Manejo y organización conceptual y física de la información en forma multidimensional.

2. Transparencia. Capacidad para acceder a datos de otras fuentes de manera sencilla y transparente.

3. Accesibilidad. Habilidad para obtener información completa y estructurada de fuentes externas de datos, tales como bases de datos relacionales, archivos planos, etc.

4. Desempeño y consistencia. El número de dimensiones utilizadas en el sistema no debe degradar el desempeño del sistema, ni tampoco afectar la consistencia de la información.

5. Cliente/servidor. Las herramientas deben poder operar en ambientes cliente/servidor.

6. Dimensionalidad genérica. Cada dimensión deberá ser tratada de igual manera.

7. Uso eficiente del almacenamiento. Manejo eficiente de la porosidad (sparseness) de la base multidimensional, para ocupar la misma cantidad de espacio. Por porosidad se entiende la manera en la que herramienta maneja el espacio requerido para almacenar la información multidimensional; este punto es muy importante ya que, debido a la estructura de los datos en las bases multidimensionales se cuenta con muchas celdas o campos vacíos. Un buen manejo de la porosidad implica que la herramienta es capaz de detectar las celdas vacías, y hacer eficiente el espacio que éstos requieren.

8. Soporte o múltiples usuarios. Permitir el acceso de múltiples usuarios al mismo tiempo al modelo.

9. Operaciones entre dimensiones sin límite. Capacidad para realizar operaciones entre varias dimensiones sin ningún tipo de restricción.

10. Manipulación intuitiva de datos. Capacidad de navegación a través de los datos, dimensiones y jerarquías de la base mediante una interfaz de usuario fácil de usar.

11. Producción flexible de reportes. Utilitarios para la creación rápida de reportes, consultas y gráficos.

12. Capacidad ilimitada para dimensiones y relaciones (jerarquías). Capacidad para manejar un número ilimitado de jerarquías, relaciones y dimensiones de datos.

Lo interesante no es poder realizar consultas que, en cierto modo, se pueden hacer con selecciones, proyecciones, concatenaciones y agrupamientos tradicionales.

2.5.3 Modos de Almacenamiento

OLAP soportas tres tipos de almacenamiento:

• ROLAP

• MOLAP

• HOLAP

2.5.3.1 Almacenamiento ROLAP (OLAP Relacional)

La arquitectura ROLAP, accede a los datos de un Data Warehouse para proporcionar los análisis OLAP. La premisa de los sistemas ROLAP es que las capacidades OLAP se soportan mejor contra las bases de datos relacionales.

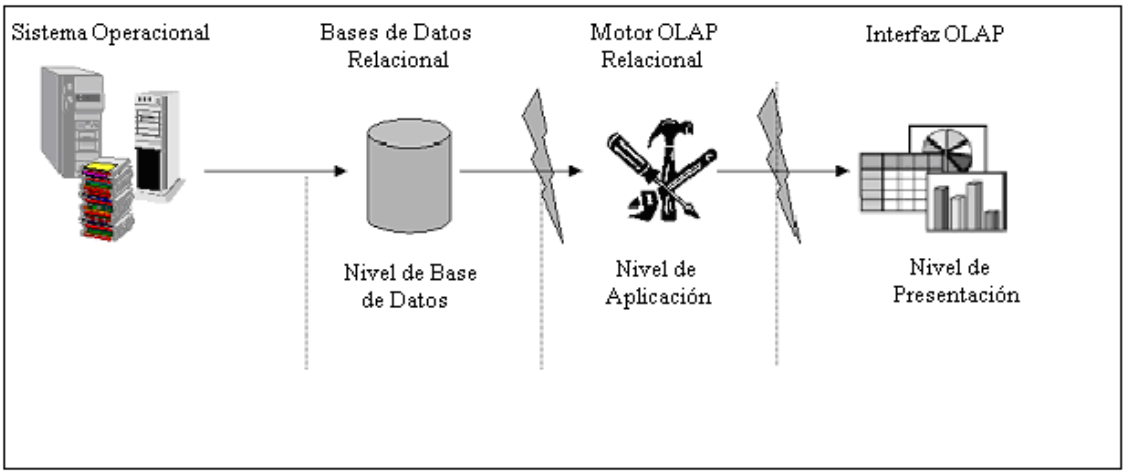


Figura Nº 2.17: "Arquitectura Sistema ROLAP". [SAS2003]

Como se muestra en la figura Nº 2.17, el sistema ROLAP utiliza una arquitectura de tres niveles:[SAS2003]

• El Nivel de Bases de Datos usa bases de datos relacionales para el manejo, acceso y obtención del dato, para realizar operaciones como: consultas, tablas hash, indexación paralela, consolidaciones batch, particionamiento de datos, joins y optimizaciones.

• El Nivel de Aplicación es el motor que ejecuta las consultas multidimensionales de los usuarios, como ratios, rangos, transformaciones, predicciones, manejo de excepciones, planificación, etc.

• El motor ROLAP se integra con niveles de presentación, a través de los cuales los usuarios realizan los análisis OLAP, a través de gráficos, alertas, drill-down, slice & dice.

Esta arquitectura accede directamente a los datos del Data Warehouse, y soporta técnicas de optimización de accesos para acelerar las consultas. Estas optimizaciones son, entre otras particionado de los datos a nivel de aplicación, soporte a la desnormalización y joins múltiples.

2.5.3.2 Almacenamiento MOLAP (OLAP Multidimensional)

La arquitectura MOLAP usa bases de datos multidimensionales para proporcionar el análisis. Su premisa es que el OLAP esta mejor implementado almacenando los datos multidimensionalmente. El sistema MOLAP utiliza una arquitectura de dos niveles tal como se muestra en la figura Nº 2.18:

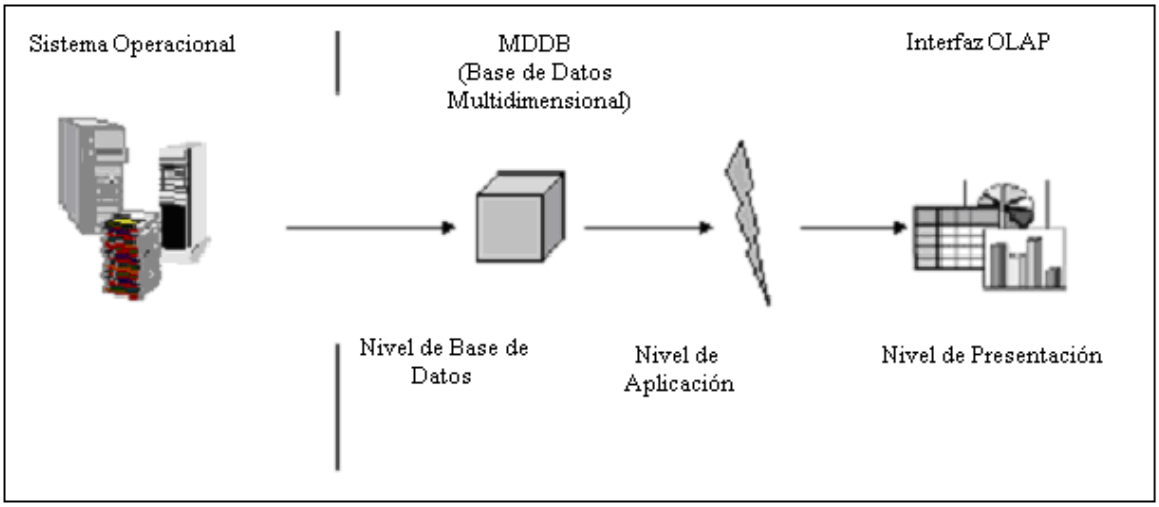


Figura Nº 2.18: *"Arquitectura Sistema MOLAP"*. [SAS2003]

• La Base de Datos multidimensional es la encargada del manejo, acceso y obtención del dato.

• El nivel de aplicación es el responsable de la ejecución de los requerimientos OLAP. El nivel de presentación se integra con el de aplicación y proporciona una interfaz a través del cual los usuarios finales visualizan los análisis OLAP. Una arquitectura cliente-servidor permite a varios usuarios acceder a la misma base de datos multidimensional.

La arquitectura MOLAP requiere unos cálculos intensivos de compilación. Lee datos precompilados, y tiene capacidades limitadas de crear agregaciones dinámicamente o de hallar ratios que no se hayan precalculado y almacenado previamente.

2.5.4 Data Warehouse y OLAP

Aunque los términos Data Warehouse y OLAP en ocasiones se utilizan indistintamente, se aplican a diferentes componentes de sistemas conocidos como sistemas de ayuda a la toma de decisiones o sistemas de inteligencia empresarial.

Los datos históricos de un Data Warehouse se utilizan para realizar análisis que apoyen las decisiones empresariales a diferentes niveles, desde el diseño estratégico a la evaluación del rendimiento de una unidad determinada de la organización. Los datos contenidos en un Data Warehouse se encuentran organizados para permitir el análisis más que para procesar transacciones en tiempo real como ocurre en los sistemas de proceso de transacciones en línea (OLTP).

La tecnología OLAP permite un uso más eficaz de los Data Warehouses para el análisis en línea, lo que proporciona respuestas rápidas a consultas analíticas complejas e iterativas. Los modelos de datos multidimensionales de OLAP y las técnicas de agregados de datos organizan y resumen grandes cantidades de datos para que puedan ser evaluados con rapidez mediante el análisis en línea y las herramientas gráficas. La respuesta a una consulta realizada sobre datos históricos a menudo suele conducir a consultas posteriores en las que el analista busca respuestas más concretas o explora posibilidades. Los sistemas OLAP proporcionan la velocidad y la flexibilidad necesarias para dar apoyo al analista en tiempo real.[MIC2003b]

La figura Nº 2.19 muestra como interacciona un Data Warehouse con la tecnología OLAP.

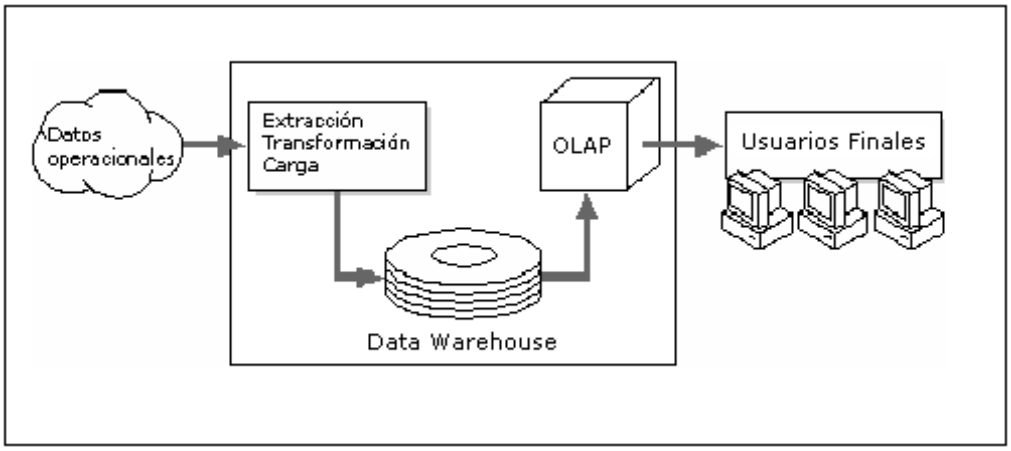


Figura Nº 2.19: "Interacción entre Data Warehouse y OLAP". [MIC2003b]

Capítulo 3: Estado del Arte Ciclo de Vida y Metodología de Implementación para Soluciones de Inteligencia de Negocios

3. EL CICLO DE VIDA DIMENSIONAL DEL NEGOCIO

El ciclo de vida de los Data Warehouses fue definido por Ralph Kimball [Kim98]. El marco presentado por Ralph Kimball con el nombre de Business Dimensional Lifecycle (BDL) ilustra las diferentes etapas por las que debe pasar todo proceso de Data Warehousing. Este enfoque de implementación de data warehouses es ilustrado en la siguiente figura. Este diagrama ilustra la secuencialidad de tareas de alto nivel requeridas para el efectivo diseño, desarrollo e implementación de Data Warehouses. El diagrama muestra una vista general del mapa de ruta de un proyecto en el cual cada rectángulo es un mojón que nos indica dónde estamos parados, por dónde pasamos y hacia dónde debemos dirigirnos.

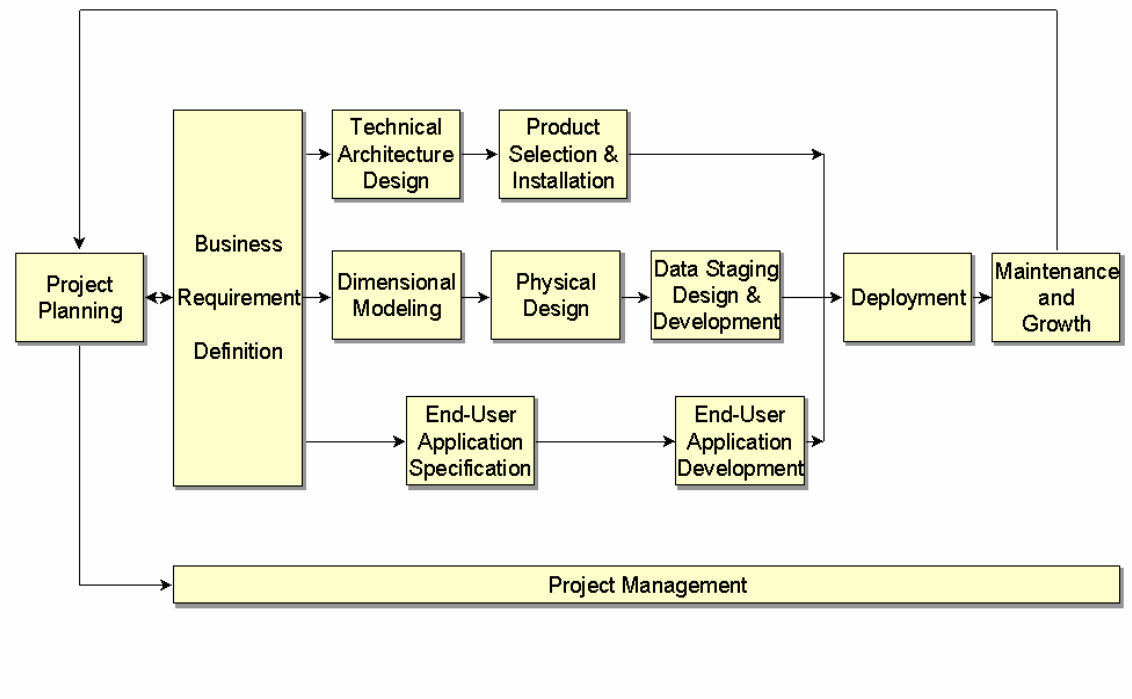


Figura Nº 3.1:"Business Dimensional Lifecycle propuesto por Ralph Kimball"

Es importante aclarar, como lo hacen los autores, que el BDL no intenta reflejar un proyecto en término de tiempos y plazos. Como se puede notar cada rectángulo del diagrama tiene el mismo ancho, con la excepción del gerenciamiento del proyecto. Cualquiera que haya pasado por algún proyecto de Data Warehousing sabe que la magnitud de recursos y tiempo requerido para cada rectángulo del ciclo de vida no es igual. El BDL se focaliza en secuencialidad y concurrencia no en tiempos y plazos.

A continuación pasaremos de describir cada una de las etapas del BDL.

3.1 Planificación del Proyecto

La planificación busca identificar la definición y el alcance del proyecto de Data Warehouse, incluyendo justificaciones del negocio y evaluaciones de factibilidad.

La planificación del proyecto se focaliza sobre recursos, perfiles, tareas, duraciones y secuencialidad. El plan de proyecto resultante identifica todas las tareas asociadas con el BDL e identifica las partes involucradas.

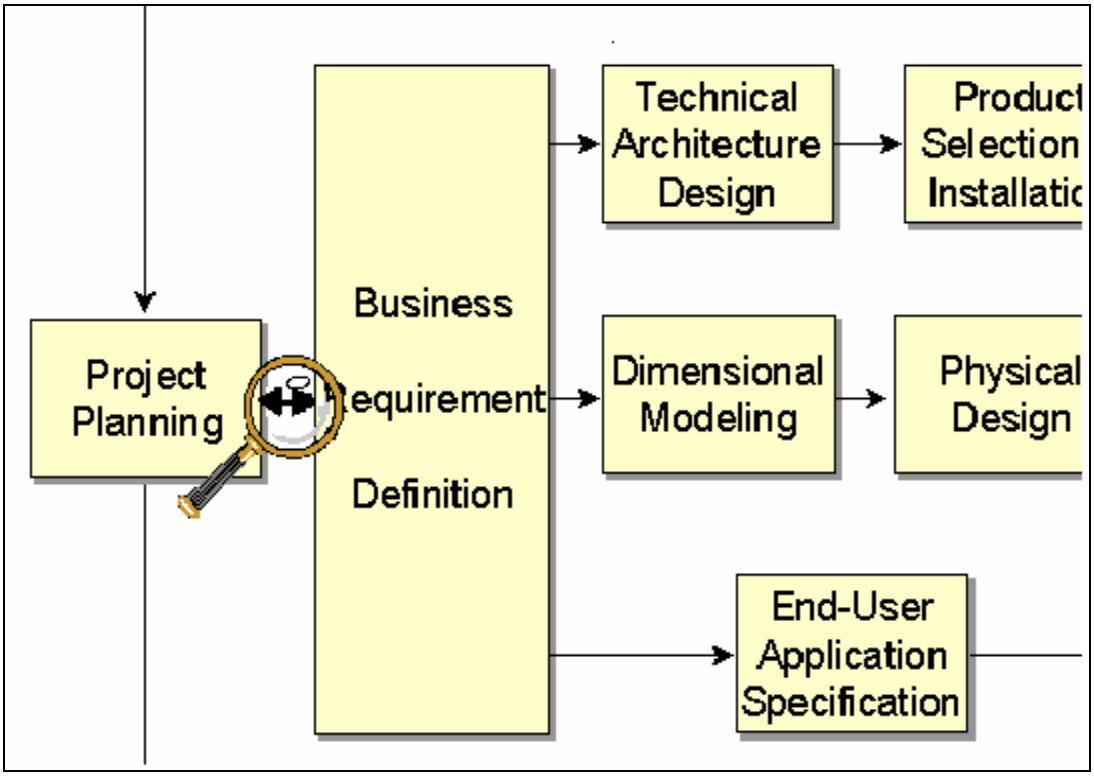


Figura Nº 3.2:"Interacción entre Planificación y Requerimientos"

La planificación del proyecto es dependiente de los requerimientos del negocio, como es denotado en el diagrama del BDL, ya que los requerimientos del negocio determinan el alcance del proyecto,

definen los recursos necesarios, etc. y la planificación acotará los requerimientos ya sea por cuestiones de recursos y/o tiempo.

Esta etapa se concentra sobre la definición del proyecto (identificación del escenario del proyecto para saber de dónde surge la necesidad del data warehouse). Según sentencia Kimball, "*Antes de comenzar un proyecto de data warehouse o data mart, hay que estar seguro si existe la demanda y de dónde proviene. Si no se tiene un sólido usuario sponsor y no hay usuarios entusiasmados, posponga el proyecto*".

Factores asociados con estas etapas incluyen: identificación de los usuarios sponsors, convincentes motivaciones del negocio, cooperación entre áreas de sistemas y negocios, cultura analítica de la organización y análisis de factibilidad (tanto tecnológica como de disponibilidad de datos).

Para medir estos factores propone un test de buena disposición del proyecto dónde describe diferentes escenarios posibles [Kim98]. Adicionalmente propone técnicas (relevamientos de alto nivel, priorización de requerimientos y pruebas de concepto) para mitigar las deficiencias que el proyecto pudiera tener en algunos de los factores mencionados anteriormente.

Cómo metodología de estas etapas propone identificar el alcance preliminar basándose en los requerimientos del negocio y no en fechas límites (deadlines) construyendo la justificación del proyecto en términos del negocio con indicadores como el ROI (retorno de inversión), NPV (valor presente neto) y el IRR (índice de retorno interno).

A nivel de planificación del proyecto, establece la identidad del mismo, el personal (staff): los usuarios sponsors, lideres, gerentes del proyecto (tanto de sistemas como del sector usuarios), equipo "corazón" del proyecto (analistas, arquitectos, DBAs, diseñadores, responsables de extracción, desarrolladores, instructores, etc.), equipo "especial" del proyecto (soporte, seguridad informática, programadores, analistas QA), el desarrollo del plan del proyecto, el seguimiento y monitoreo.

3.2 Definición de los Requerimientos del Negocio

Un factor determinante en el éxito de un proceso de Data Warehousing es la interpretación correcta de los diferentes niveles de requerimientos expresados por los

diferentes niveles de usuarios. La técnica utilizada para relevar los requerimientos de los analistas del negocio difiere de los enfoques tradicionales guiados por los datos [Inm92] [Gol99]. Los diseñadores de los data warehouses deben entender los factores claves que guían al negocio para determinar efectivamente los requerimientos y traducirlos en consideraciones de diseño apropiadas.

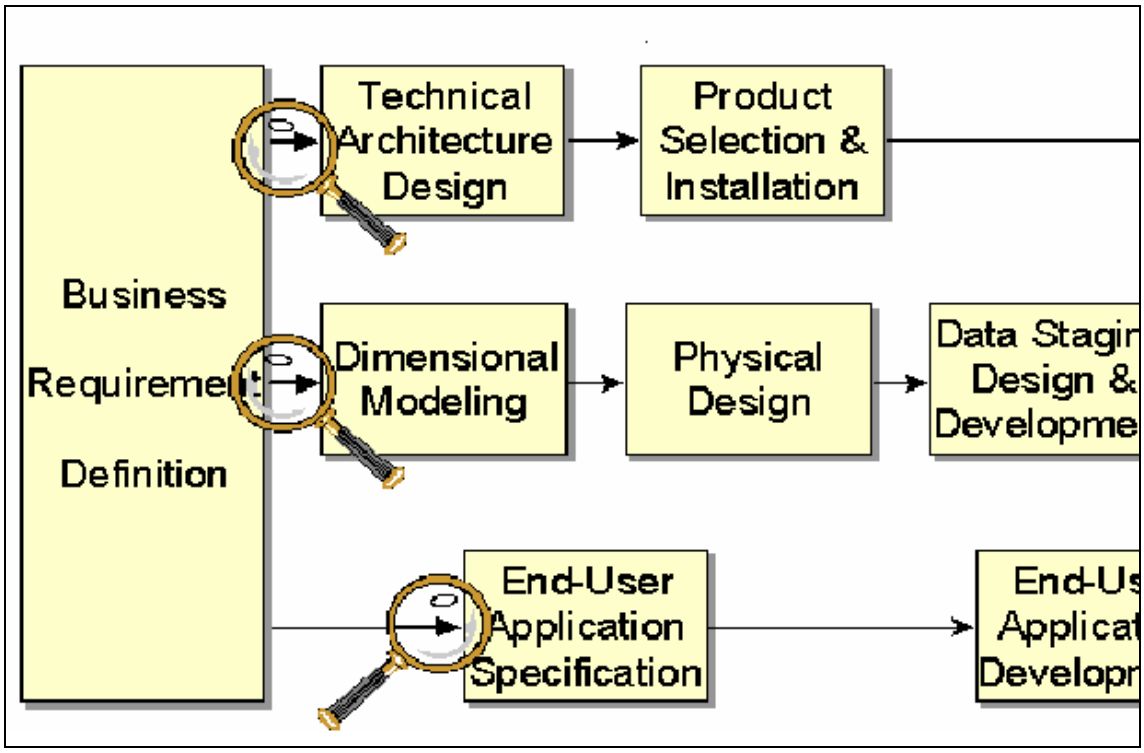


Figura Nº 3.3:"Interacción entre Requerimientos del Negocio y etapas subsiguientes"

La definición de los requerimientos del negocio establece la base para las tres etapas paralelas subsiguientes focalizadas en la tecnología, los datos y las aplicaciones por lo cual es altamente crítica y es el centro de atención del BDL.

Los usuarios finales y sus requerimientos impactan siempre en las implementaciones realizadas de un data warehouse. Según la perspectiva de Kimball [Kim98], los requerimientos del negocio se posicionan en el centro del "universo del data warehouse". Como destaca siempre el autor, los requerimientos del negocio deben determinar el alcance del data warehouse (qué datos debe contener, cómo debe estar organizado, cada cuánto debe actualizarse, quiénes y desde dónde accederán, etc.). Kimball da consejos y técnicas para descubrir eficazmente los requerimientos del negocio. Estas tácticas y estrategias se focalizan sobre las entrevistas de relevamiento (diferentes tipos, preparación de la entrevista, roles a cubrir, búsqueda de información pre-entrevista, selección de entrevistados, desarrollo de los cuestionarios, planificación,

preparación de los entrevistados, conducción de la entrevista, contenido, cierre, revisión de resultados, etc.).

3.3 Modelado Dimensional

La definición de los requerimientos del negocio determina los datos necesarios para cumplir los requerimientos analíticos de los usuarios. Diseñar los modelos de datos para soportar estos análisis requiere un enfoque diferente al usado en los sistemas operacionales. Básicamente se comienza con una matriz donde se determina la dimensionalidad de cada indicador y luego se especifican los diferentes grados de detalle (atributos) dentro de cada concepto del negocio (dimensión), como así también la granularidad de cada indicador (variable o métrica) y las diferentes jerarquías que dan forma al modelo dimensional del negocio (BDM) o mapa dimensional.

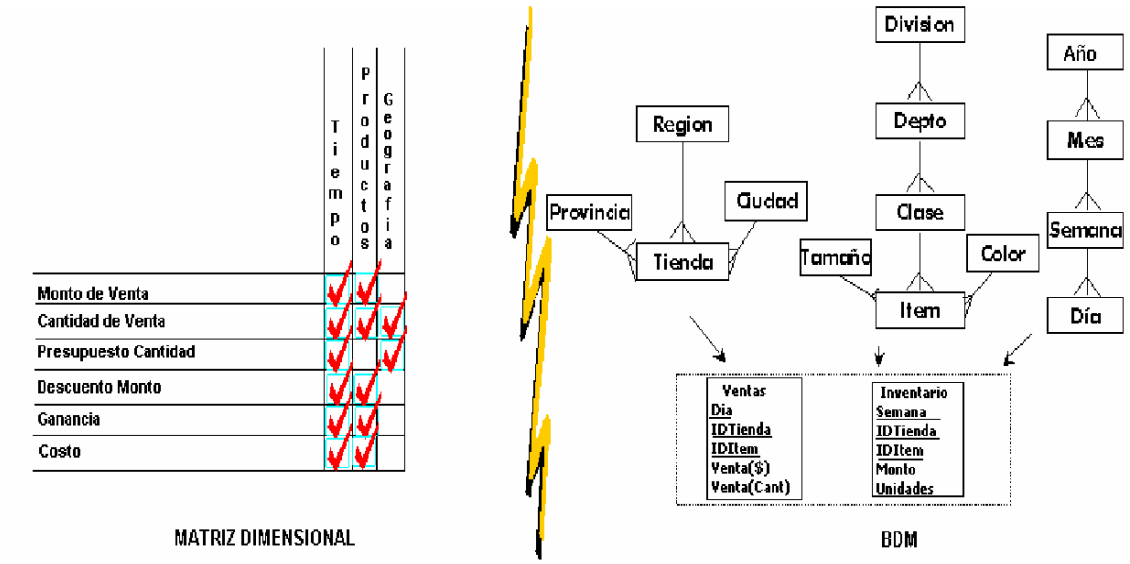


Figura Nº 3.4:"Diagramas para Modelado Dimensional"

Matriz Dimensional que determina la dimensionalidad de cada indicador y modelo dimensional del negocio (BDM) que ilustra las diferentes jerarquías dentro de cada dimensión.

Ralph Kimball es realmente un referente en el tema de modelado dimensional como lo demuestran sus numerosos papers y libros publicados al respecto. Gran parte de las técnicas básicas en este tema (identificación de dimensiones, atributos, star schemas,

snowflake schemas, soluciones verticales, etc.) son tratadas e introducidas en [Kim92] constituyendo la base de la teoría sobre modelado dimensional.

El modelo dimensional utilizado en el modelado de data warehousing, organiza y presenta los datos definiendo dimensiones (líneas o áreas temáticas del negocio). Por ejemplo, en los modelos de retail es muy común encontrar las mediciones estructuradas por las dimensiones sucursal, producto y tiempo.

De esta forma, permiten analizar la información a distintos niveles de agregación dentro de las diferentes dimensiones. Dentro de cada dimensión se puede definir los niveles de agregación o sumarización para cada análisis, a estos niveles de granularidad se los caracteriza con el nombre de atributos.

Dentro de este contexto, un análisis consiste en definir el nivel de agregación que califique el conjunto de datos resultado, independientemente para cada dimensión.

3.3.1 Componentes básicos

A la hora de realizar un modelo dimensional del negocio (BDM) se hace necesaria la presentación de los siguientes componentes que definiremos a lo largo de este capítulo:

- **Dimensiones** (Dimensions)
- **Atributos** (Attributes)
  Elementos (Attributes Elements)
- **Relaciones** (Attribute Relationships)
  Jerarquías (Hierarchies)
- **Variables** o **Indicadores** (Facts o Metrics)

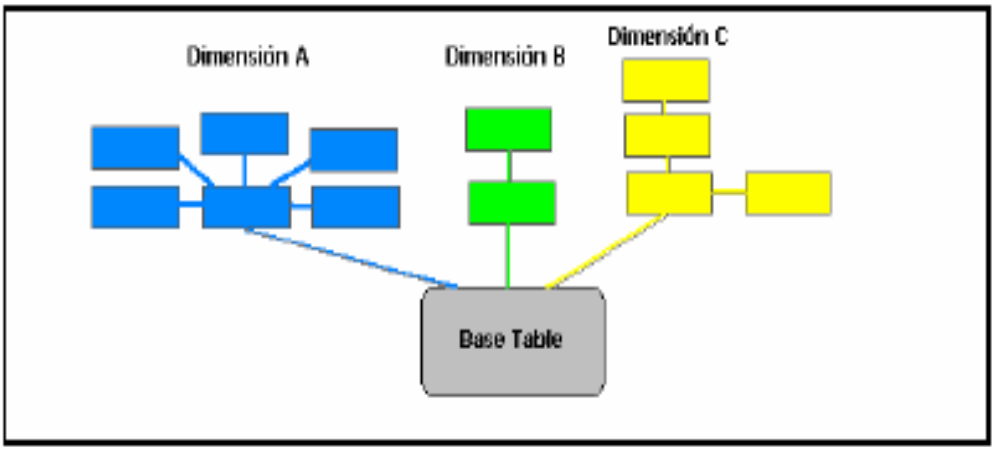


Figura Nº 3.5:"Esquema de un modelo dimensional"

3.3.1.1 Dimensiones

Son las áreas temáticas, líneas del negocio o sujetos del negocio. Las mismas proveen un método general para organizar la información corporativa. Definidas como un grupo de uno o más atributos, separados y distintos uno de otros (es decir, que no se comparten atributos), según el esquema de implementación elegido, puede ser que no se encuentren explícitamente en el data warehouse, sino que se presenten sólo como un recurso conceptual o lógico que ayuda a mostrar múltiples perspectivas de los datos permitiendo realizar análisis por diferentes dimensiones o incluso cruzando información entre distintas dimensiones (Ver más adelante Star Schema y SnowFlaked Schema).

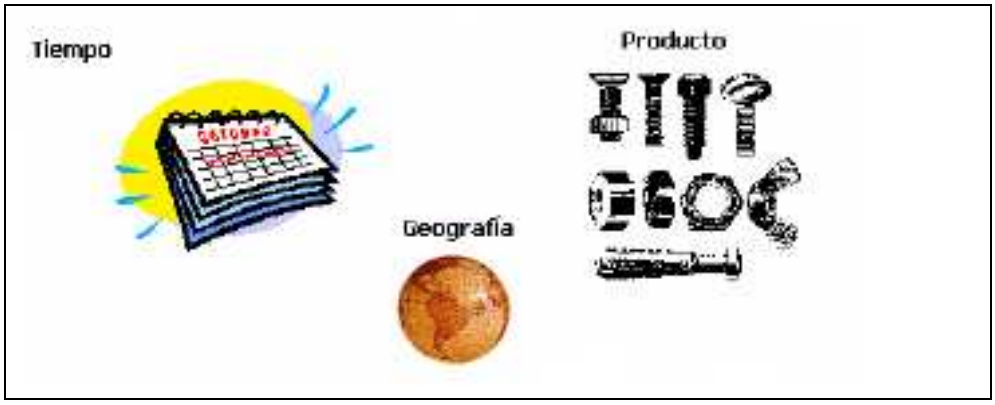


Figura Nº 3.6:"Dimensión"

3.3.1.2 Atributos

Los atributos son una agrupación de elementos o items dentro de una dimensión. Representan categorías o clases de elementos que tienen el mismo nivel lógico dentro de una dimensión donde todos los elementos de un atributo se relacionan con otros atributos de la dimensión de la misma forma. La finalidad de los atributos es ver la información de cada dimensión a diferentes niveles de detalle y agrupar los datos para ser analizados.

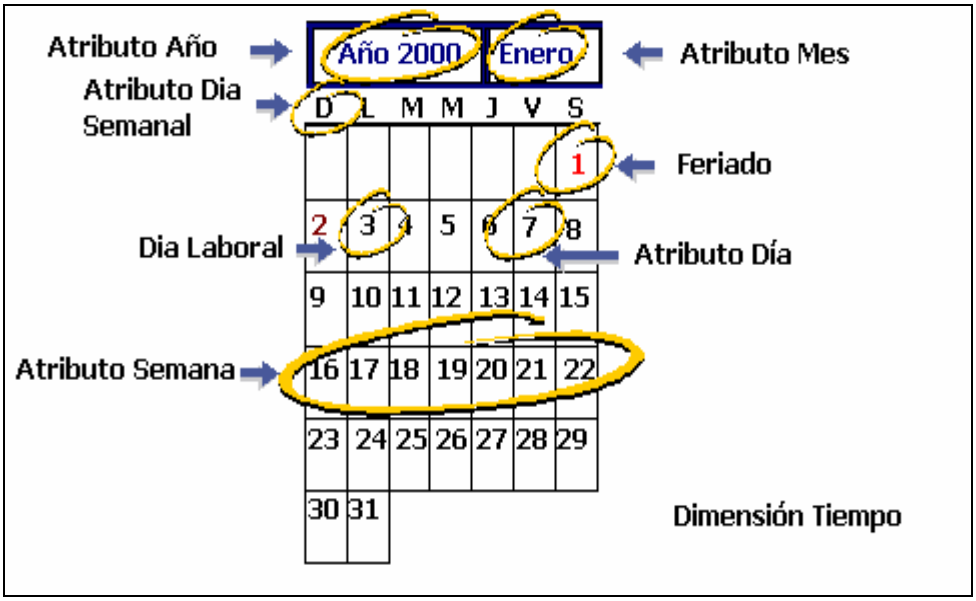


Figura Nº 3.7:"Atributos"

3.3.1.3 Elementos

Son las instancias o valores de los atributos que, como componentes atómicos del modelo, permiten clasificar el rendimiento del negocio. Es importante aclarar que si bien no forman parte del BDM es aconsejable su incorporación para un mayor entendimiento del modelo en etapas tempranas del ciclo de vida (relevamiento)

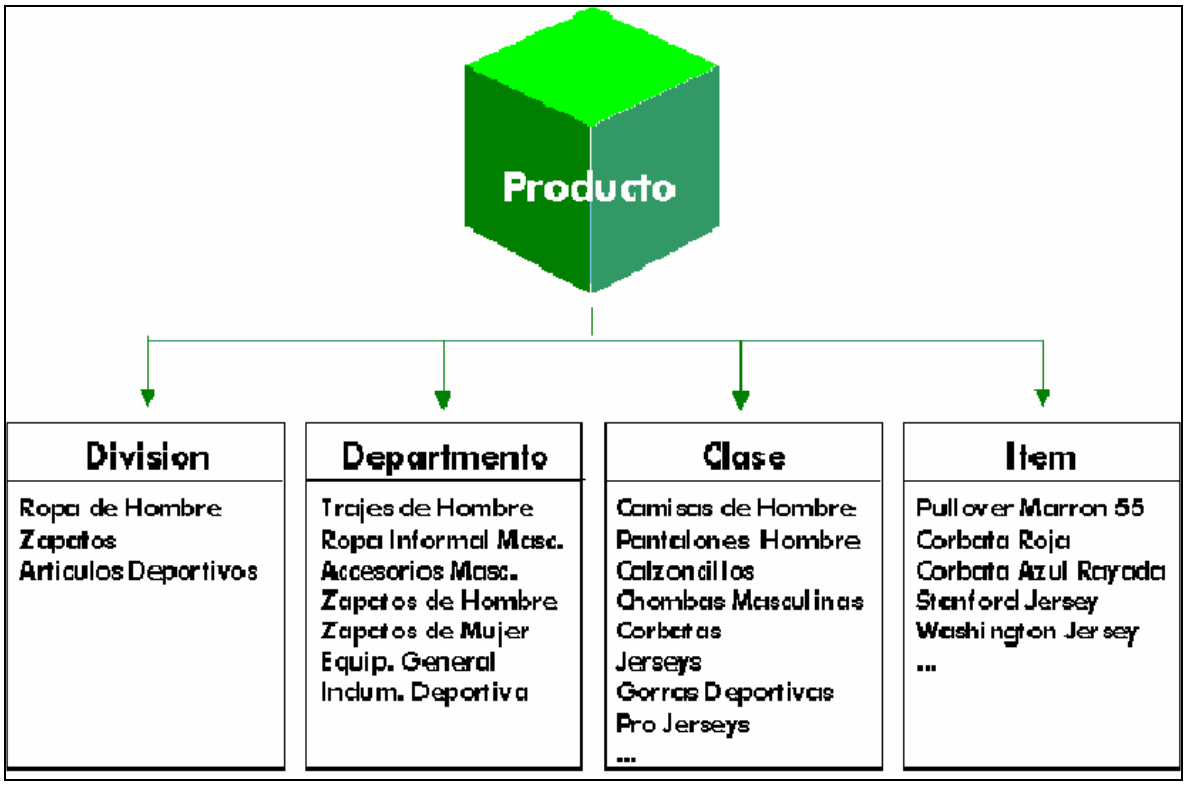


Figura Nº 3.8:"Elementos o instancias de atributos"

3.3.1.4 Relaciones

Son asociaciones lógicas de atributos dentro de una jerarquía definida por las instancias de los atributos y transitivas dentro de una jerarquía. Al igual que en DER existen diferentes tipos de relaciones (clasificadas por su cardinalidad):

- Uno-a-uno (1:1)
- Uno-a-muchos (1:M)
- Muchos-a-uno (M:1)
- Muchos-a-muchos (M:N)

Es importante aclarar que los atributos dentro de una dimensión están directamente relacionados uno con otros a través de los diferentes tipos de relaciones antes definidos. En cambio, los atributos en diferentes dimensiones están relacionados uno con otros a través de los indicadores o variables del negocio definidas como intersección de las dimensiones.

3.3.1.5 Jerarquías

Representadas por un ordenamiento lógico dentro de la dimensión, se encuentran formadas por los diferentes tipos de relaciones entre los atributos de una misma dimensión. Pueden existir múltiples jerarquías dentro de una dimensión pero siempre es posible identificar una jerarquía principal o columna vertebral de la dimensión y jerarquías secundarias o descriptivas compuestas por atributos característicos definidos desde la jerarquía principal.

Dentro del contexto de navegación del modelo dimensional, se puede decir que las diferentes jerarquías definen el mapa de caminos para el "drilling" o la "navegación" de los datos. Haremos drill-up o roll-up cuando nos movamos hacia un atributo superior dentro de la jerarquía (navegación ascendente), drill-down o roll-down cuando analicemos información a mayor nivel de detalle (navegación descendente), drill-within cuando nos movamos entre la jerarquía principal y la característica desde un atributo hacia cualquier otro que no sea ni descendiente ni ancestro dentro de la misma dimensión (navegación intradimensional) y drill-across para analizar información sobre diferentes dimensiones (navegación interdimensional).

Como convenciones del modelado, la jerarquía principal se dibuja verticalmente desde el atributo más agregado (arriba) hasta el más atómico (abajo) y las jerarquías características se adicionan por los costados.

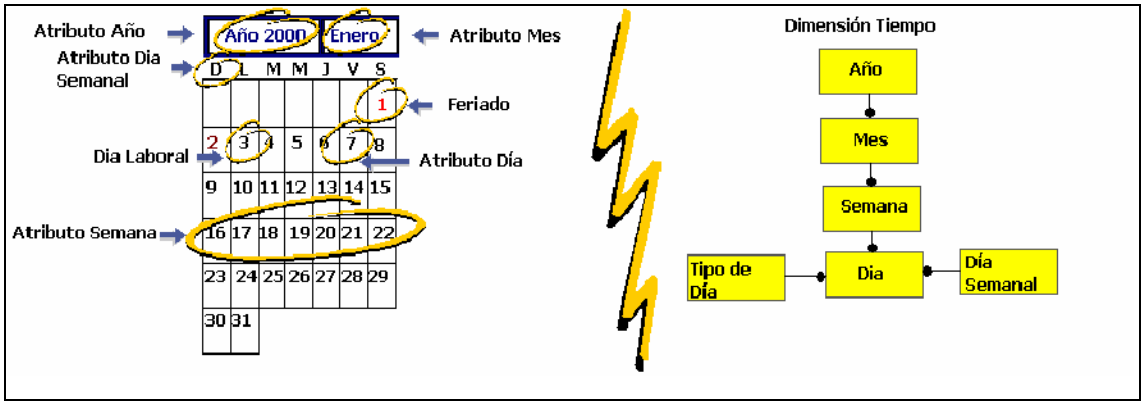


Figura Nº 3.9:"Relaciones y Jerarquías"

3.3.1.6 Indicadores

Son las variables o métricas que ayudarán a medir la performance del negocio. Existen dos tipos de indicadores: básicos y derivados.

Los indicadores básicos, primitivos o crudos existen físicamente en el warehouse junto a los atributos que los caracterizan, pueden venir de diferentes sistemas fuentes y tener distintos niveles de granularidad o agregación. Por ejemplo, la variable Venta($) es llevada diariamente mientras que el indicador Unidades en Stock es seguido semanalmente.

Por otro lado, los indicadores derivados o métricas calculadas se construyen a partir de los indicadores base y pueden o no estar almacenados físicamente en el data warehouse (es típicamente una decisión de tuning). Un ejemplo clásico de métrica derivada es Margen de Ganancia, la cual se define como la resta entre Precio y Costo (ambos indicadores básicos).

3.3.2 Comparación con modelo ER

Al momento de presentar el modelado dimensional, parece inevitable hacer la comparación con los modelos entidad-relación (ERM) dada su gran penetración y aceptación en los ambientes de base de datos relacionales y su incorporación (en la etapa de análisis de requerimientos) dentro del ciclo de vida del software.

Si bien los modelos dimensionales pueden parecer similares a los clásicos DER (Diagrama Entidad-Relación), hay algunas diferencias que enumeraremos a continuación.

- El término entidad no es usado en el modelado dimensional.
- Algunas entidades en un DER serán mapeadas a atributos en un modelo dimensional.
- Una dimensión es una agrupación lógica de atributos que relaciona a cada uno de ellos.
- Indicadores son los valores usados para medir el rendimiento del negocio, por ejemplo, costo, precio, margen, etc.
- Los indicadores se definen como intersección de las diferentes dimensiones.

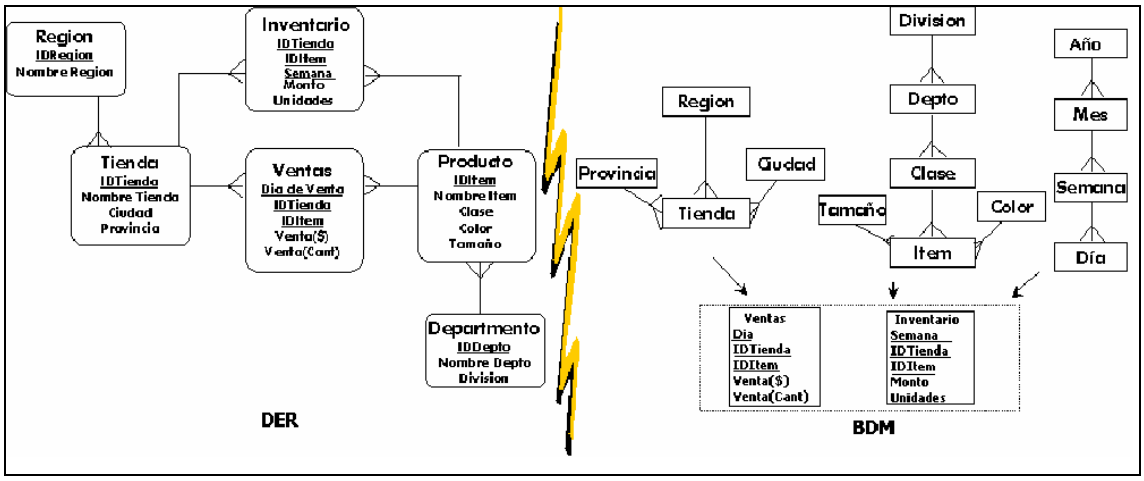


Figura Nº 3.10:"DER vs.BDM"

Según sentencia Kimball [Kim95], no tiene sentido utilizar el DER para realizar el relevamiento de requerimientos y luego aplicar una serie de reglas para transformarlo en un modelo dimensional. El DER puede ser útil para contrastar los requerimientos de los usuarios con la estructura de los sistemas operacionales existentes pero no para ser usado de base en la definición del BDM.

3.4 Diseño Físico

El diseño físico de las base de datos se focaliza sobre la selección de las estructuras necesarias para soportar el diseño lógico. Algunos de los elementos principales de este proceso son la definición de convenciones estándares de nombres y seteos específicos del ambiente de la base de datos. La indexación y las estrategias de particionamiento son también determinadas en esta etapa.

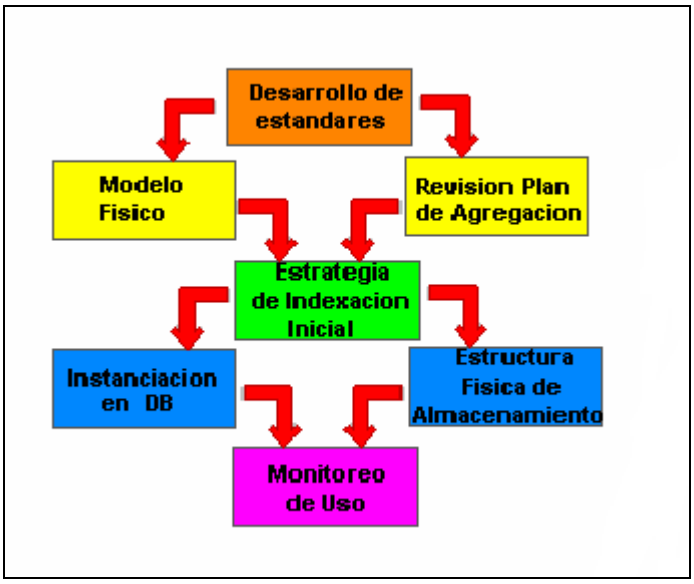


Figura Nº 3.11:"Proceso de diseño físico a alto nivel"

Debemos enfrentarnos entonces al desafío de traducir nuestro modelo dimensional del negocio al *"lenguaje"* que entienden las bases de datos. Este mapeo puede realizarse siguiendo unos pasos básicos que finalizan en un esquema normalizado (el cual se detalla en los siguientes puntos).

Luego de obtener el esquema pueden realizarse diversas técnicas de desnormalización que apunten a mejorar los tiempos de respuestas de las consultas realizadas sobre el esquema definido.

También puede identificarse cuán dinámicas o estáticas son las relaciones entre los atributos con el objetivo de planificar el seguimiento de la historia entre estas relaciones (Versioning).

En este capítulo describiremos entonces las diferentes técnicas y métodos para mapear los conceptos lógicos del modelo dimensional en las correspondientes estructuras del esquema físico. Además presentaremos diferentes esquemas de implementación analizando ventajas y desventajas de cada uno de ellos.

3.4.1 Mapeo desde Modelo Lógico hacia Esquema Físico

Como proceso de mapeo del modelo dimensional a un esquema físico se puede decir que la construcción del esquema involucra la definición con componentes propios

(tablas, filas y columnas), de todos los componentes del modelo dimensional (atributos, relaciones e indicadores).

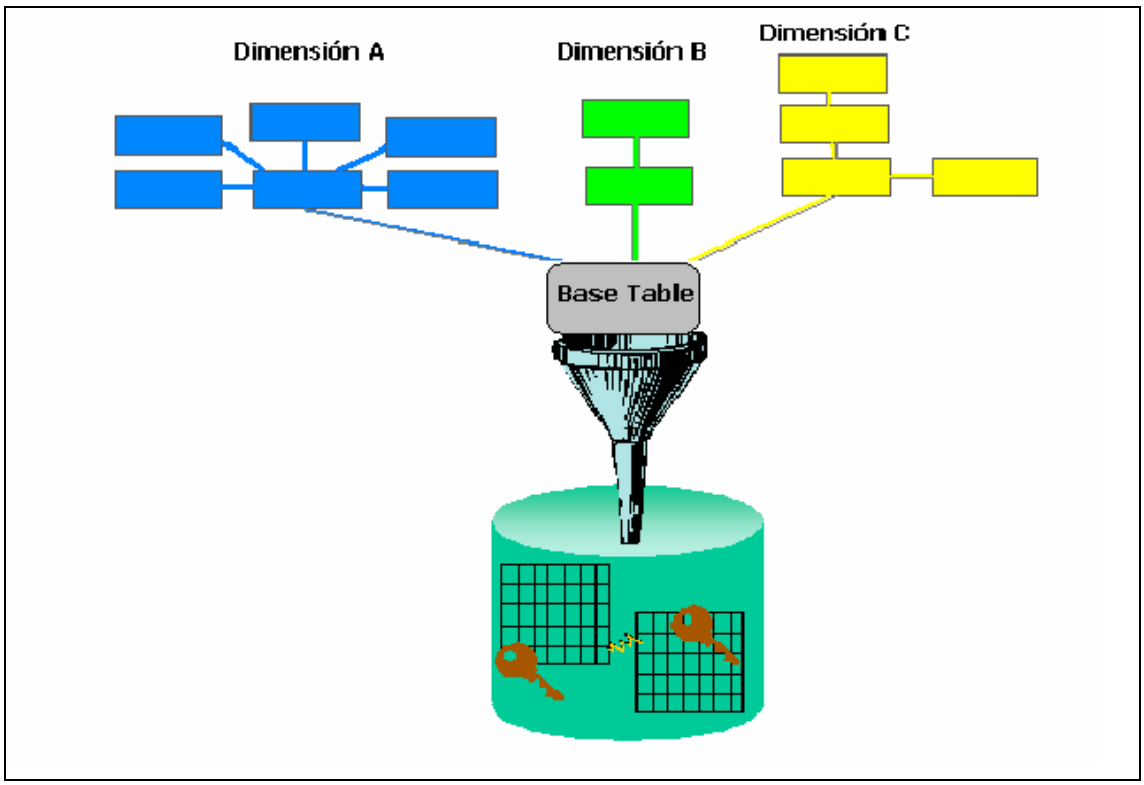


Figura Nº 3.12:"Mapeo del BDM a Esquema Físico Relacional"

Paralelamente o después de la definición de las estructuras iniciales, se debe examinar las estrategias de optimización y tuning del warehouse, como así también definir la frecuencia de refresco y los diferentes perfiles de uso.

A la hora de definir nuestro esquema físico del modelo dimensional se identifican tres tipos de tablas:

- **Tablas lookup**
  Contiene una columna identificatoria (id) y, si existe, una columna de descripción del atributo que representa
- **Tablas relacionales**
  Contiene el id de dos o más atributos definiendo, de esta forma, la asociación entre ellos.
- **Tablas base**
  Contiene columnas que representan los indicadores o variables del negocio y ids de los atributos que indican el nivel de granularidad de las métricas.

Podemos ver entonces que surgen tres tipos de columnas o campos:

- **Identificación**
  Contiene códigos de identificación

Requerido para todos los atributos
Preferiblemente numéricos (procesamiento más rápido)

- **Descripción**
  Contiene descripciones, generalmente, en formato texto.
  Opcional para todos los atributos
- **Variables** (Fact Columns)
  Numéricas

3.4.2 Construcción de tablas Lookup

Se comienza asociando una lookup a cada atributo definido en el BDM de tal forma de ir construyendo un esquema normalizado. Se identifica la clave primaria y la descripción de cada tributo para representarla como los correspondientes campos de la tabla lookup. Dado que el campo descripción es opcional puede dejarse sólo el identificador y utilizarlo con ambos propósitos.

Existen algunos atributos característicos, como por ejemplo el Domicilio de un cliente, que no tienen su propia lookup sino que residen en la lookup de algún otro atributo (por ejemplo, en este caso estaría incluido en la lookup Cliente).

Si las tablas lookups representan una dimensión completa y no un atributo en particular, entonces se las conoce también como tablas dimensionales, las mismas son utilizadas ampliamente en el esquema físico conocido con el nombre de Esquema Estrella (Star Schema).

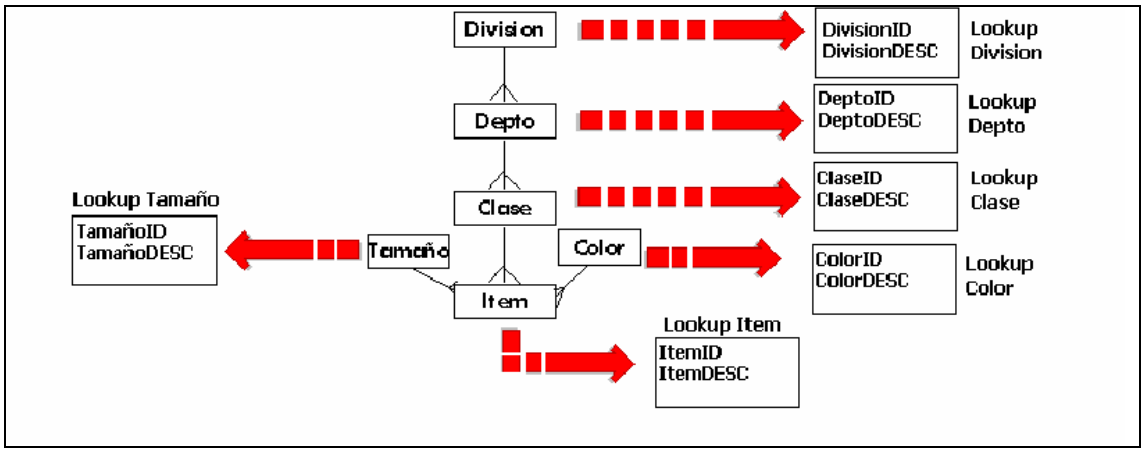


Figura Nº 3.13:"Tablas Lookup"

3.4.3 Relaciones de atributos dentro de una dimensión

Se identifica las relaciones directas entre los distintos atributos (relaciones Padre-Hijo), luego se define una tabla para cada relación que contiene los campos ids de cada lookup que representa los atributos asociados. Como se realiza comúnmente, la definición de la clave primaria estará asociada a la cardinalidad de la relación relevada en el BDM.

En el caso de las relaciones que no sean de muchos-a-muchos se puede evitar la creación de una tabla relacional, incluyendo el id del padre en la lookup del hijo de la relación y definiéndola como clave foránea referenciándola a la lookup padre. De esta forma también se pueden incluir los ids de todos sus ancestros y mejorar los tiempos de respuesta al reducir la cantidad de joins en las consultas, incrementando, por otro lado el mantenimiento de las tablas.

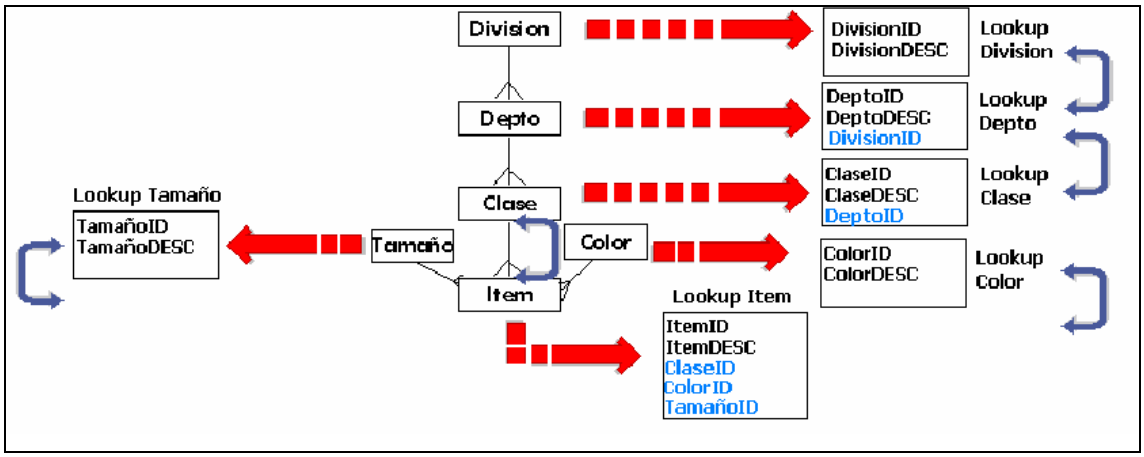


Figura Nº 3.14:"Implementación de relaciones"

3.4.4 Definición de tablas Base

Se debe definir una tabla para cada conjunto de indicadores, agrupados según su granularidad. Como se explicó anteriormente, cada tabla de hechos (fact table o base table) tiene campos numéricos asociados a los indicadores y campos ids para cada atributo que interviene en la *"dimensionalidad"* y *"granularidad"* de la variable a medir. La dimensionalidad de una variable está asociada a los conceptos generales (dimensiones) que participan en la definición de un indicador y la granularidad al nivel

de detalle o agregación de la misma y define los atributos de cada dimensión que caracterizan el valor de la variable. Por ejemplo, si las ventas se cargan diariamente para cada sucursal y cada familia de productos, la dimensionalidad del indicador *Unidades Vendidas* será Tiempo - Geografía-Productos y la granularidad será Día-Sucursal - Familia, en cambio, si el stock de mercadería se sigue semanalmente, la dimensionalidad de *Unidades en Stock* será la misma pero la granularidad será Semana-Sucursal-Familia.

La clave primaria de la tabla base estará compuesta por todos los ids que definan la dimensionalidad de los indicadores y cada id individualmente referenciará, como clave foránea, a la tabla lookup que represente el atributo asociado.

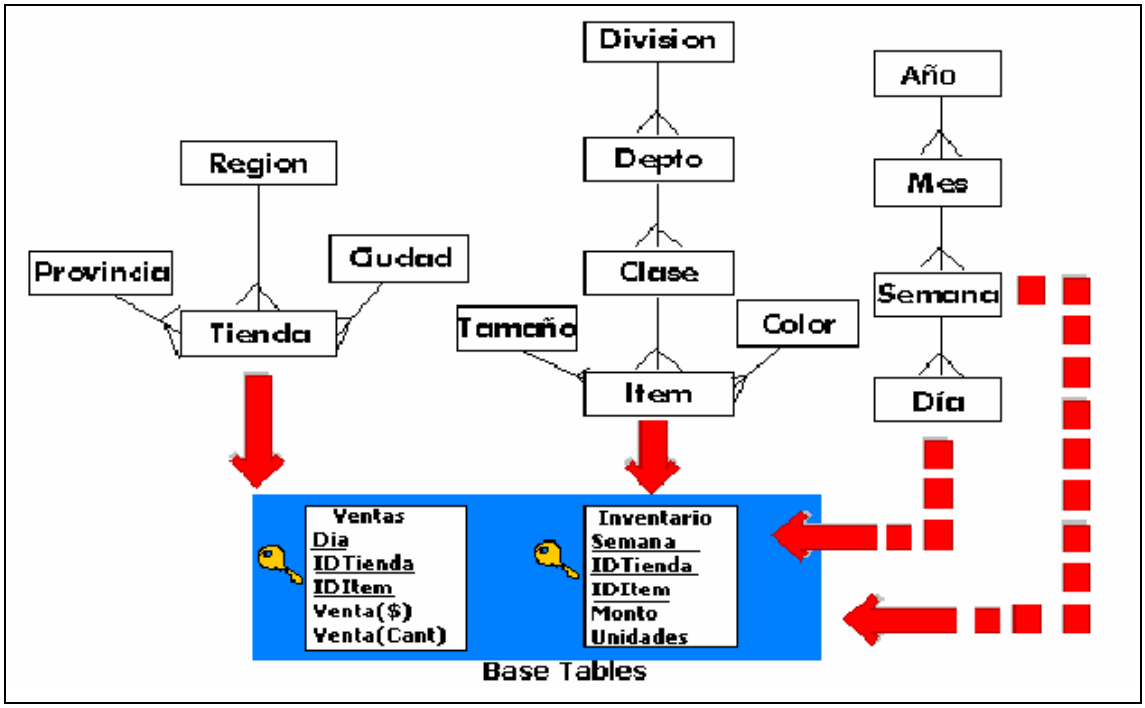


Figura Nº 3.15:"Tabla Base"

3.4.5 Tipos de esquemas

Tomando como base el trabajo de Jeff Bedell [Bed97], en este capítulo analizaremos algunos de los posibles esquemas físicos que se pueden implementar junto a las ventajas y desventajas asociadas a cada uno de ellos.

3.4.5.1 Esquema Estrella Consolidado I

Las tablas lookup representan a una dimensión completa. Contienen una clave genérica, una descripción genérica, un campo de nivel que indica la "altura" del atributo dentro de la jerarquía y campos ids para cada atributo dentro de la dimensión. Por ejemplo, si tenemos tres dimensiones Producto, Geografía y Tiempo, existen tres lookups, cada una con un id genérico del tipo Producto_Key, Geo_Key, Tiempo_Key junto a Producto_DESC, Geo_DESC y Tiempo_DESC mas el campo 'nivel' de tipo numérico y los ids: FamiliaID, LineaID, ItemID en el caso de la dimensión Producto.

Este esquema soporta solamente una tabla base y en la misma caen sólo las claves genéricas de las dimensiones junto a los indicadores del negocio. Al no existir tablas relacionales, pues los atributos y sus relaciones de una dimensión se encuentran todos representados en una única lookup, esta estructura no admite relaciones M:N de forma natural.

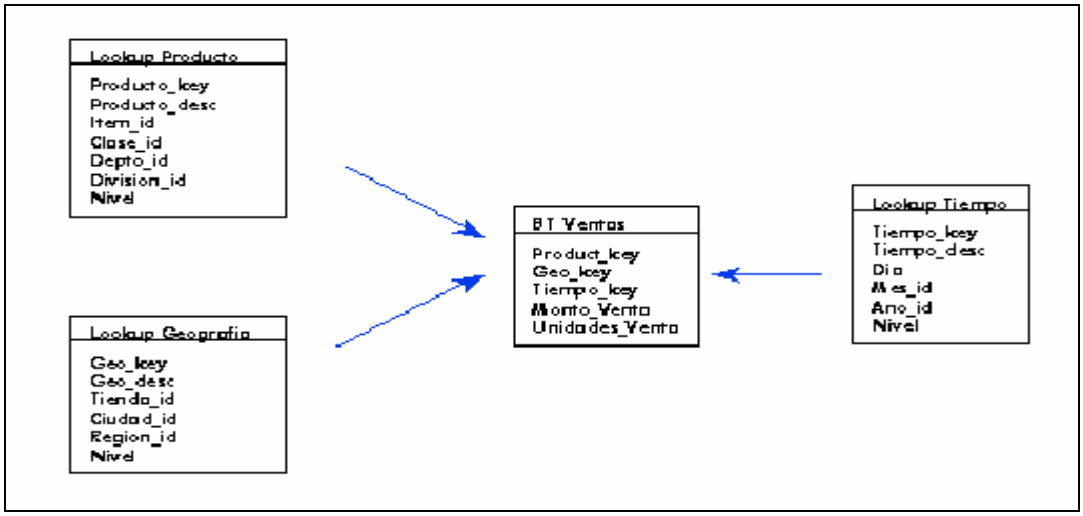


Figura Nº 3.16:"Esquema Estrella Consolidado I"

3.4.5.2 Esquema Estrella Consolidado II

Las tablas lookups tienen la misma estructura básica del Esquema Consolidado I, salvo que el id de cada atributo es reemplazado por la descripción y la descripción genérica es eliminada. La estructura de la tabla base es idéntica al esquema anterior.

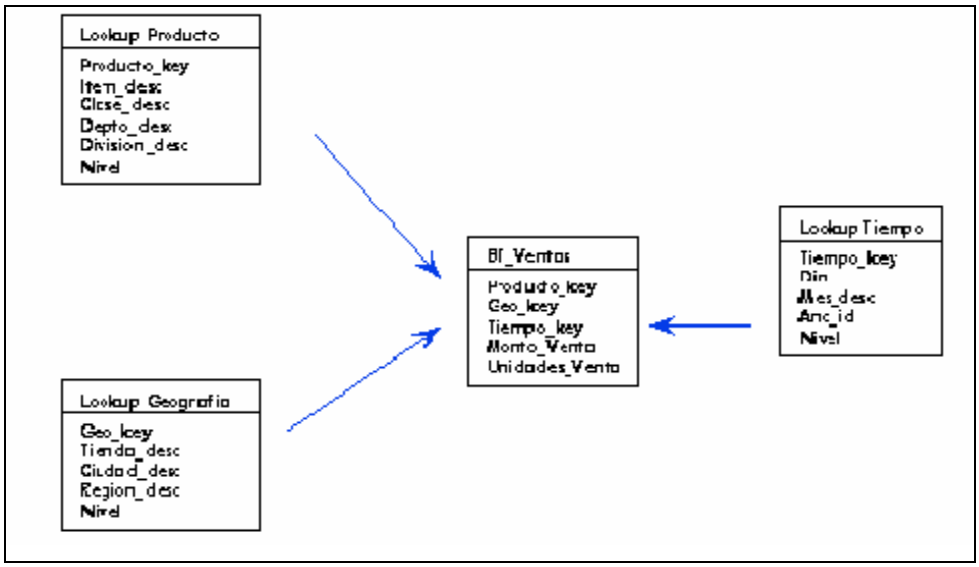


Figura Nº 3.17:"Esquema Estrella Consolidado II"

3.4.5.3 Esquema Estrella Consolidado III

Las tablas lookups son una combinación de los dos esquemas anteriores, contienen la clave genérica de la dimensión, el campo de nivel, los ids y las descripciones de cada atributo de la jerarquía de esa dimensión. La estructura de la tabla de hechos permanece intacta.

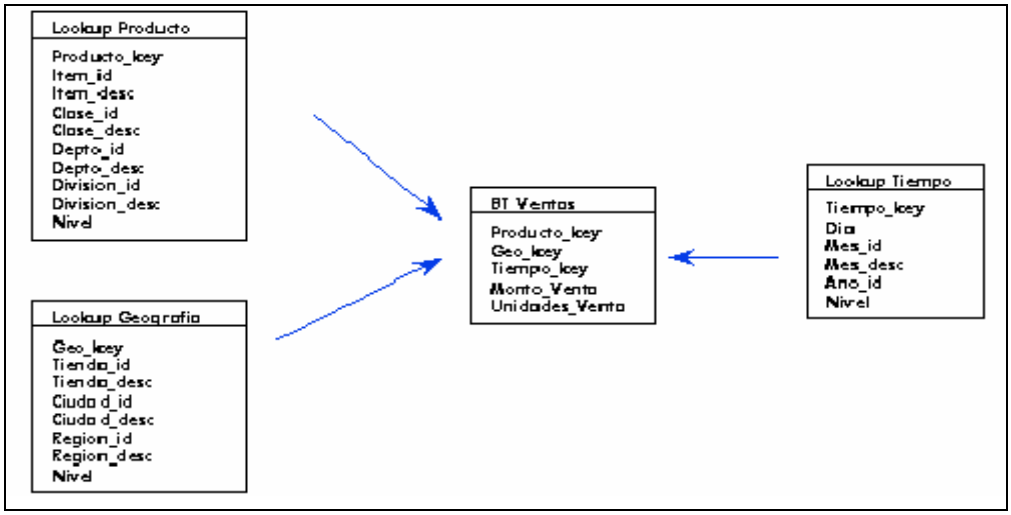


Figura Nº 3.18:"Esquema Estrella Consolidado III

3.4.5.4 Comparación entre Esquemas Estrella Consolidados

Estos esquemas tienen abundancia de campos en NULL, formando una matriz triangular superior en la parte correspondiente a los indicadores de la tabla base. Haciendo un análisis más particular podemos ver que una debilidad que presenta EEC I se manifiesta

a la hora de requerir un análisis que involucre más de un atributo de una misma dimensión. Por ejemplo, ver las ventas discriminadas por Región y Provincia (ambos atributos de la dimensión Geografía). Dado que existe sólo un campo descripción en la lookup Geografía, para producir el efecto deseado se debe realizar un left join para cada atributo de la misma dimensión que se desee mostrar, pudiendo ser esta operación muy costosa.

El EEC II resuelve este último problema pero al tener sólo las descripciones (generalmente texto variable) la búsqueda de los atributos correspondientes para una consulta se hace demasiado lenta en contraposición a la búsqueda por campos numéricos como son generalmente los ids. Por último el EEC III combina las características de los EEC I y II pagando el costo de ser un esquema completamente desnormalizado. En cuanto a la tabla base, la misma es del tipo *"in-table aggregation"* pues todas las consolidaciones (niveles de agregación) se realizan dentro de la misma tabla de hechos, de allí el nombre de Consolidado que reciben estos esquemas.

Entre las ventajas de estos esquemas se encuentra la poca cantidad de tablas y la facilidad de generar SQL automáticamente (para realizar los distintos análisis) ya que la estructura es bastante estática.

Entre las desventajas podemos enumerar la dificultad de relaciones M:N, tabla base extremadamente grande por manejar las agregaciones dentro de sí misma. En general puede decirse que no es un modelo escalable en cuanto a cantidad de atributos dentro de cada dimensión y en cantidad de elementos dentro de cada atributo.

3.4.5.5 Esquema Estrella Normalizado I

Las tablas lookups representan a cada atributo. Contienen el id, la descripción y el id del padre inmediato en la jerarquía de la dimensión a la que pertenecen. En caso de relaciones M:N se define la correspondiente tabla relacional que involucra los ids de los atributos asociados.

La tabla base tiene un único nivel de datos (no tiene agregaciones incluidas). La clave primaria está formada por los ids de los atributos hojas de las dimensiones. Este esquema es comúnmente conocido con el nombre de esquema Copo de Nieve

(SnowFlaked Schema). Como puede verse, este esquema se encuentra en tercera forma normal.

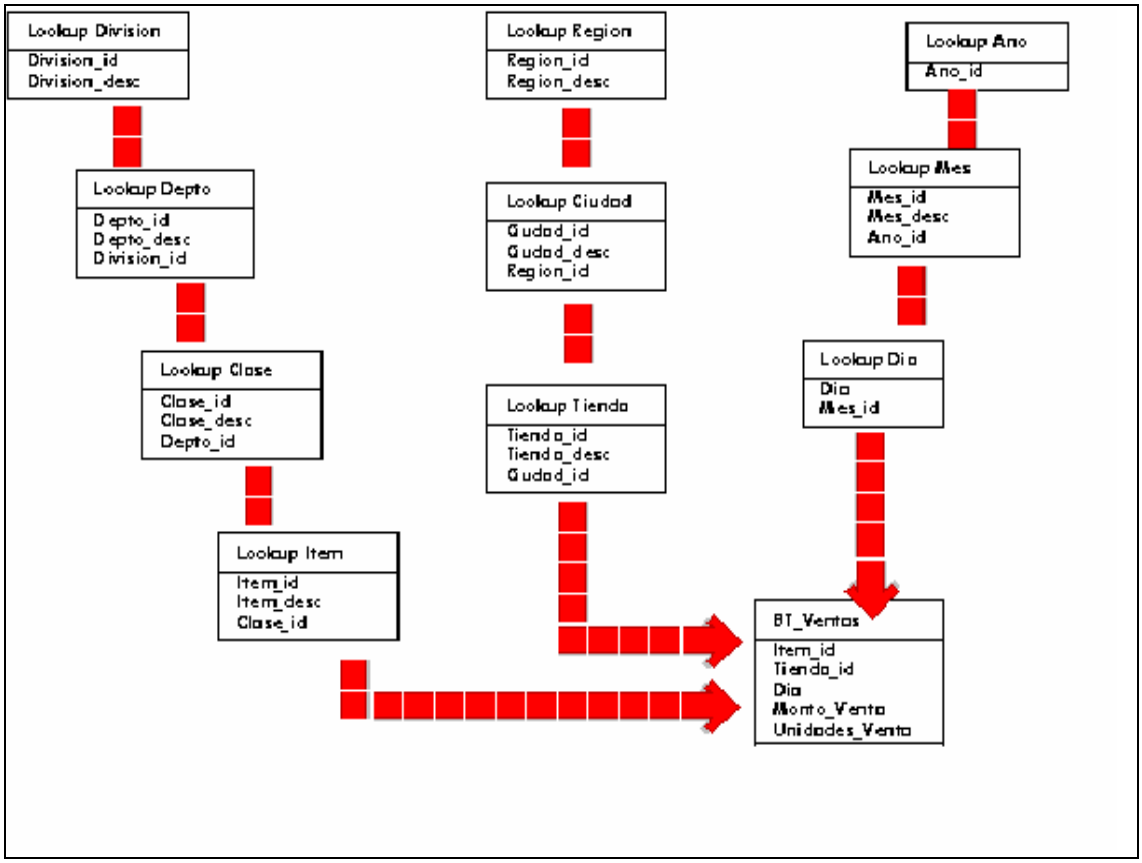


Figura Nº 3.19:"Esquema Estrella Normalizado I"

3.4.5.6 Esquema Estrella Normalizado II

A las tablas lookups se le incorpora los ids de todos sus ancestros (todas los ids de las tablas que están por encima de ella en la jerarquía de la dimensión). La estructura de la tabla base se mantiene inalterable con respecto al esquema anterior.

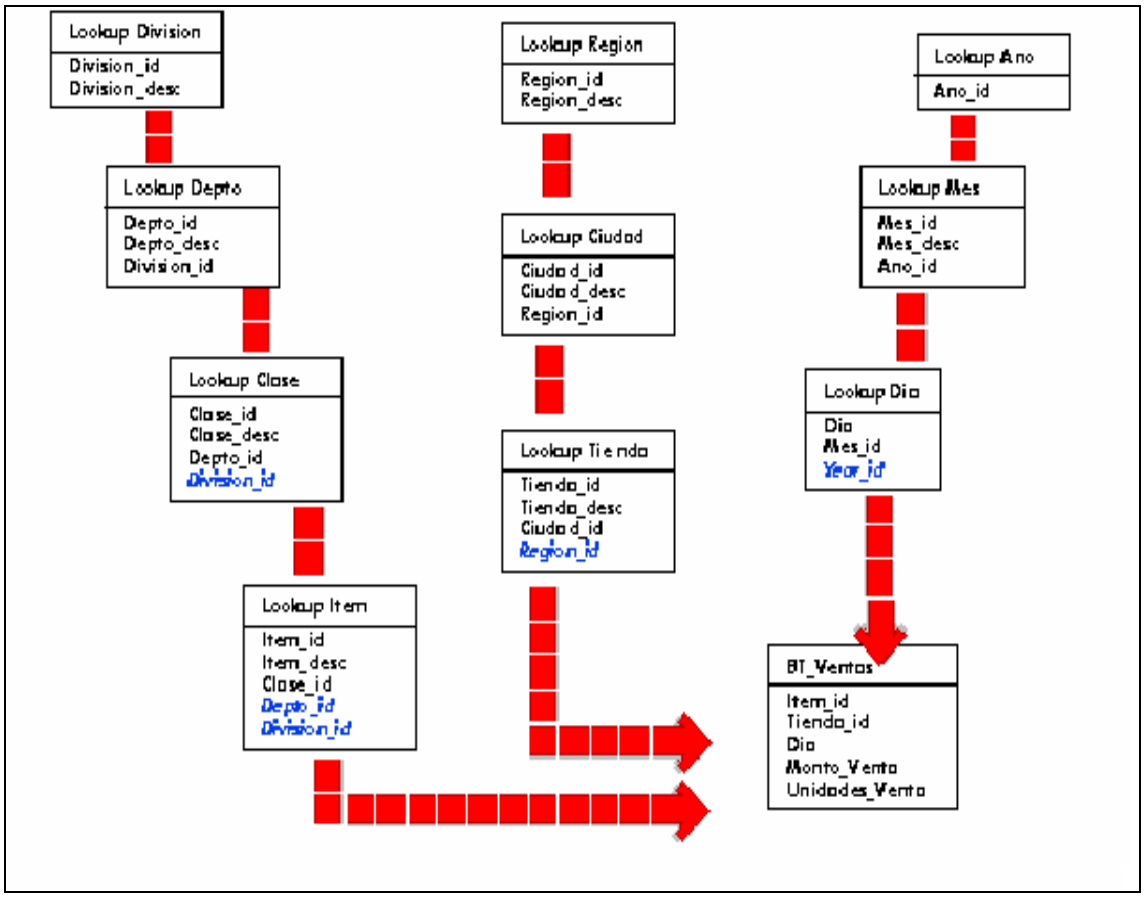


Figura Nº 3.20:"Esquema Estrella Normalizado II"

3.4.5.7 Esquema Estrella Normalizado III

Incluye en las tablas lookups no sólo los ids de sus ancestros sino también las descripciones de los mismos. La estructura de la tabla base no sufre modificaciones.

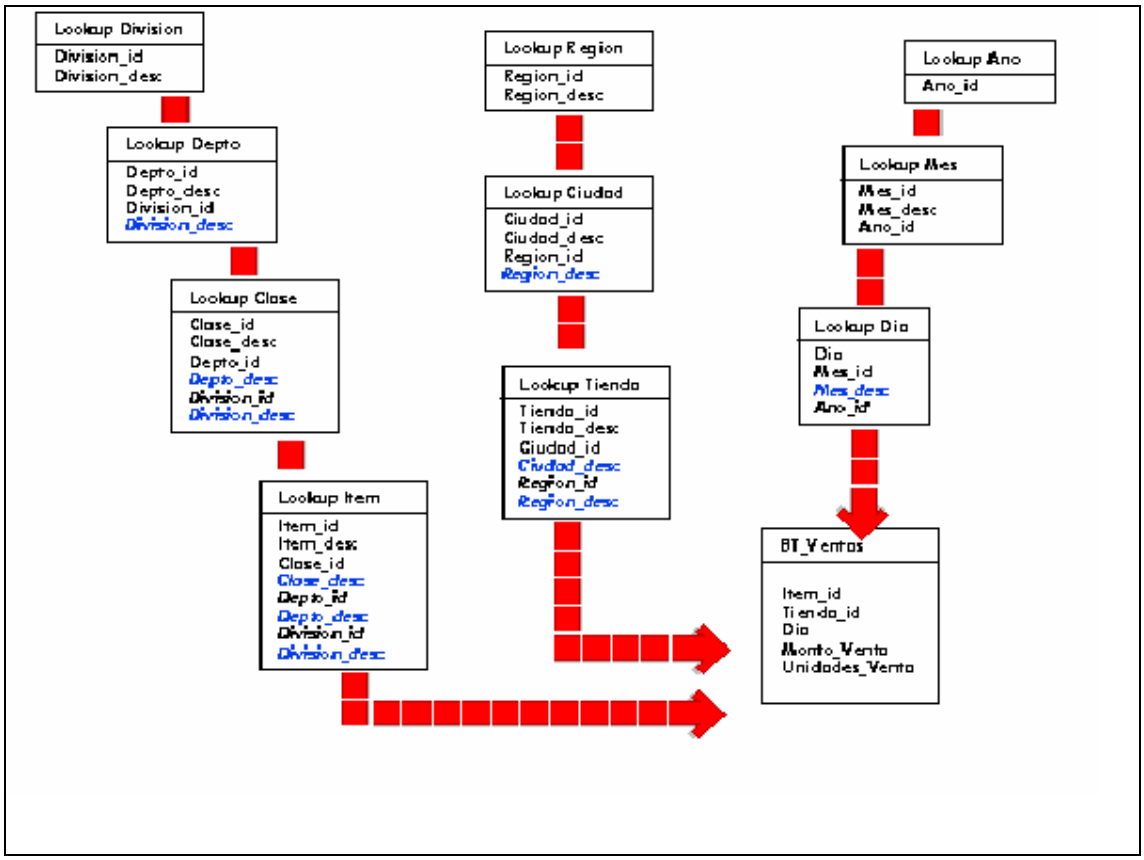


Figura Nº 3.21:"Esquema Estrella Normalizado III"

3.4.5.8 Comparación entre Esquemas Estrella Normalizados

El EEN I requiere mínima cantidad de espacio pero requiere muchos joins a la hora de responder consultas a niveles superiores de las jerarquías o al incluir análisis donde se muestra la descripción de padre e hijo, p.e., nuestro análisis anterior a nivel de Región y Provincia.

El EEN II, que involucra la desnormalización de los ids, requiere espacio adicional de almacenamiento pero reduce el número de joins necesario para resolver las consultas. Sólo realiza joins de tablas para mostrar descripciones. Por último, el EEN III, con desnormalización de ids y descripciones, requiere más espacio de almacenamiento pero elimina la necesidad de consultar más de una tabla por dimensión. Las tablas lookups que no son hojas son utilizadas en consultas particulares que muestran sólo a ese nivel (como puede ser tablas de agregación o listado de elección de elementos de atributos -pick lists-).

Estos esquemas soportan la posibilidad de múltiples fact tables y niveles de agregación de diferente granularidad que hacen a estos esquemas ser esparsos o densos según la cantidad y criterio de las agregaciones.

Entre los beneficios de este tipo de esquemas podemos encontrar entonces la flexibilidad y escalabilidad, relaciones M:N, manejo óptimo de pick lists, agregación en tablas individuales (de fácil acceso y mantenimiento). Como puntos negativos se encuentran el crecimiento en la cantidad de tablas y la dificultad para generar SQL de forma automática.

3.5 Diseño y Desarrollo de Presentación de Datos

Esta etapa es típicamente la más subestimada de las tareas en un proyecto de data warehouse.

Las principales subetapas de esta zona del ciclo de vida son: la extracción, la transformación y la carga (ETL process). Se definen como procesos de extracción a aquellos requeridos para obtener los datos que permitirán efectuar la carga del Modelo Físico acordado. Así mismo, se definen como procesos de transformación los procesos para convertir o recodificar los datos fuente a fin poder efectuar la carga efectiva del Modelo Físico. Por otra parte, los procesos de carga de datos son los procesos requeridos para poblar el Data Warehouse.

Todas estas tareas son altamente críticas pues tienen que ver con la materia prima del data warehouse: los datos. La desconfianza y pérdida de credibilidad del data warehouse serán resultados inmediatos e inevitables si el usuario choca con información inconsistente. Es por ello que la calidad de los datos es un factor determinante en el éxito de un proyecto de data warehousing. Es en esta etapa donde deben sanearse todos los inconvenientes relacionados con la calidad de los datos fuente.

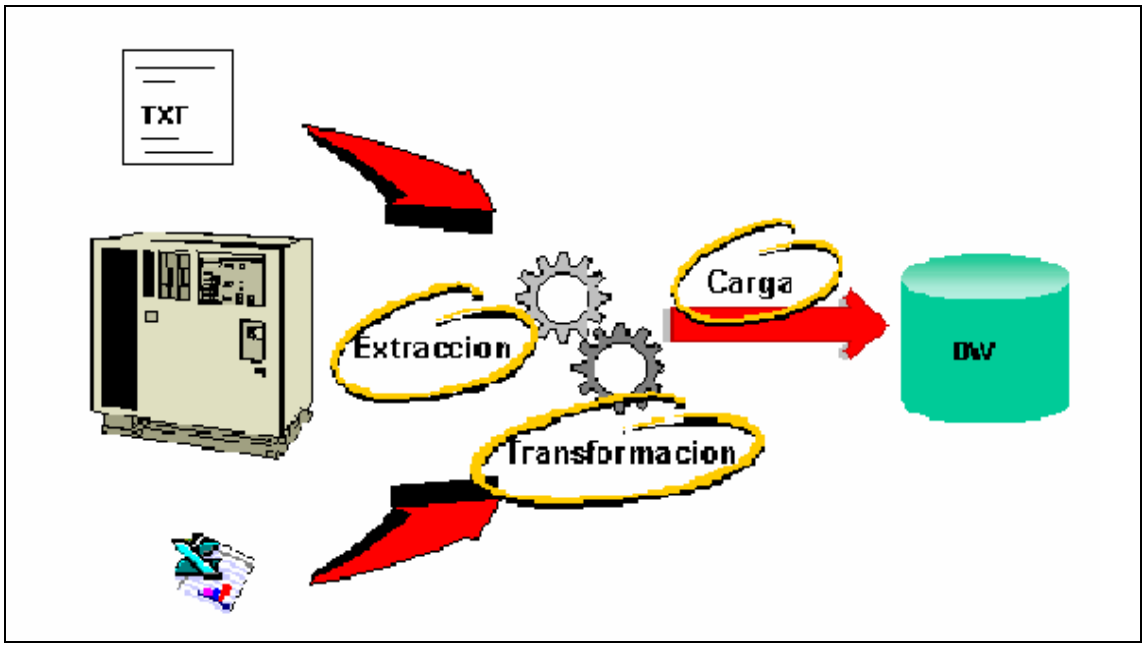


Figura Nº 3.22:"Proceso ETL"

Como advierte Kimball [Kim98], el proceso de Data Staging es el iceberg de un proyecto de datawarehousing. Son muchos los desafíos que deben enfrentarse para lograr datos de alta calidad de los sistemas fuentes. Como comentamos anteriormente, en general es una de las etapas más subestimadas, que siempre termina tomando más tiempo del previsto. Ralph Kimball propone entonces un plan de 10 items que ayudarán a guiar esta etapa del BDL.

Plan:

1. Crear un diagrama de flujo fuente-destino esquemático, de una página y de muy alto nivel.

2. Probar, elegir e implementar una herramienta de data staging.

3. Profundizar en detalle por tabla destino, gráficamente describir las reestructuraciones o transformaciones complejas. Gráficamente ilustrar la generación de las claves surrogadas. Desarrollo preliminar de la secuencialidad de los trabajos.

Carga de dimensiones:

4. Construir y probar la carga de una tabla dimensional estática. La principal meta de este paso es resolver los problemas de infraestructura que pudieran surgir (conectividad, transferencia, seguridad, etc.)

5. Construir y probar los procesos de actualización de una dimensión.

6. Construir y probar las cargas de las restantes dimensiones.

Fact Tables y automatización:

7. Construir y probar la carga histórica de las fact tables (carga masiva de datos). Incluyendo búsqueda y sustitución de claves.

8. Construir y probar los procesos de cargas incrementales.

9. Construir y probar la generación de agregaciones.

10. Diseñar, construir y probar la automatización de los procesos.

3.6 Diseño de la Arquitectura Técnica

Los ambientes de data warehousing requieren la integración de numerosas tecnologías. Se debe tener en cuenta tres factores: los requerimientos del negocio, los actuales ambientes técnicos y las directrices técnicas estratégicas futuras planificadas para de esta forma poder establecer el diseño de la arquitectura técnica del ambiente de data warehousing.

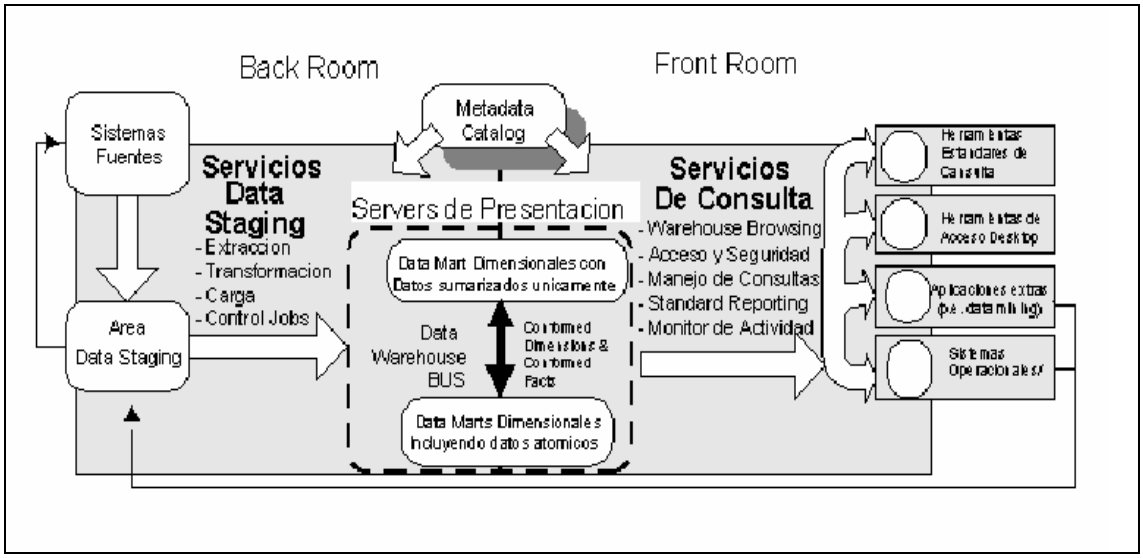


Figura Nº 3.23:"Modelo de Arquitectura Técnica a alto nivel"

Ralph Kimball hace una analogía entre los planos arquitectónicos de una casa y la arquitectura de un warehouse, nadie comenzaría diciendo, *"Hey, tengo algo de madera y algo de concreto, vamos a construir una casa!"*. Hay que tener un plan antes de comenzar, no es simplemente reordenar y explotar la información. Al igual que en una construcción, los planos sirven para comunicar los deseos entre los clientes y el

arquitecto, como así también para medir esfuerzos y materiales necesarios para la obra (comunicación, planificación, flexibilidad y mantenimiento, documentación, productividad y reuso). Finalmente, argumenta Kimball, "un buen conjunto de planos, como cualquier buena documentación, nos ayudará más tarde cuando sea tiempo de remodelar o hacer incorporaciones".

Entre los muchos ensayos y enfoques existentes para la definición y desarrollo de arquitecturas de sistemas, Kimball se basa en el Framework propuesto por John A. Zachman de Zachman International y lo adapta a un ambiente de datawarehousing, simplificándolo considerablemente ya que en este tipo de ambientes, en teoría, no debe ser atacado los problemas de infraestructura de los procesos transaccionales. El Data Warehouse Architecture Framework propuesto por Kimball incluye tres áreas (columnas): el área de arquitectura de datos (qué), el área de arquitectura técnica (cómo) y el área de infraestructura (dónde). A su vez, cada una de estas áreas tienen diferentes niveles de detalles (filas): Nivel de Requerimientos del Negocio, Nivel de Modelo Arquitectónico, Nivel de Modelo Detallado, Nivel de Implementación.

Como podemos inferir de las identificaciones de los diferentes niveles, Kimball recomienda un enfoque top-down, comenzando con una visión global y dividiendo la arquitectura en pequeñas piezas hasta llegar al grado donde las piezas pueden ser realmente implementadas.

Como siempre el énfasis está puesto en los requerimientos del negocio, son nuestra guía primaria para desarrollar nuestra arquitectura y para priorizar. La arquitectura técnica se divide en dos partes, el back room (la parte interna del warehouse) y el front room (la cara pública del warehouse), interactuando constantemente. Mientras los requerimientos del negocio nos dicen qué necesitamos hacer, la arquitectura técnica nos responde el interrogante de cómo lo haremos.

3.7 Selección de Productos e Instalación

Utilizando el diseño de arquitectura técnica como marco, es necesario evaluar y seleccionar componentes específicos de la arquitectura como es la plataforma de hardware, el motor de base de datos, la herramienta de ETL o el desarrollo pertinente, herramientas de acceso, etc. Una vez evaluados y seleccionados los componentes

determinados se procede con la instalación y prueba de los mismos en un ambiente integrado de data warehousing.

Caracteristica	Peso	Producto A	Producto B	Producto C	...
Capacidades Basicas de ETL					
xxxxxxxxxxxxxxxxxxxxxx	85				
xxxxxxxxxxxxxxxxxxxxxx	70				
Job Control & Scheduling					
xxxxxxxxxxxxxxxxxxxxxx	90				
xxxxxxxxxxxxxxxxxxxxxx	85				
Metada & Estandares					
xxxxxxxxxxxxxxxxxxxxxx					
xxxxxxxxxxxxxxxxxxxxxx					
Info de Vendedor	70				
xxxxxxxxxxxxxxxxxxxxxx	85				
xxxxxxxxxxxxxxxxxxxxxx					
Puntaje Total					
Ranking					

Figura Nº 3.24:"Ejemplo de matriz de evaluación de productos"

3.8 Especificación de Aplicaciones para Usuarios Finales

No todos los usuarios del warehouse necesitan el mismo nivel de análisis. Es por ello que en esta etapa se identifican los diferentes roles o perfiles de usuarios para determinar los diferentes tipos de aplicaciones necesarias en base al alcance de los diferentes perfiles (gerencial, analista del negocio, vendedor, etc.)

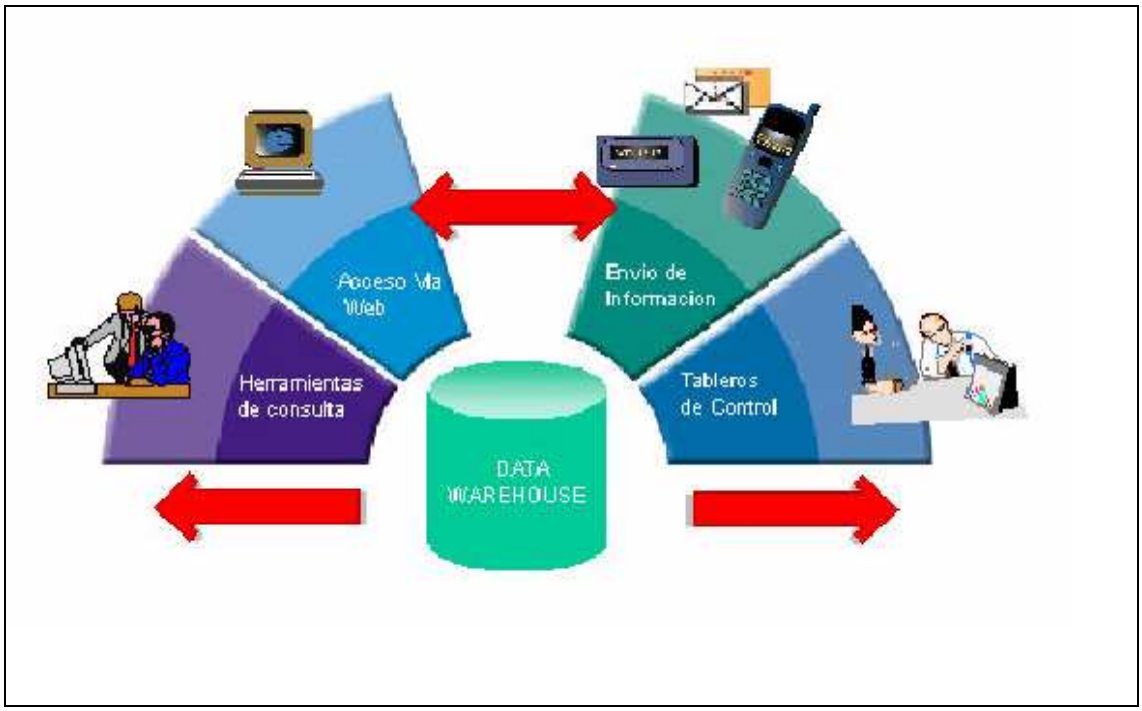


Figura Nº 3.25:"Variedad de interfaces por perfil"

Los diferentes roles o perfiles de usuarios determinan la interface o ventana al warehouse. Herramientas de diseño de reportes y consultas avanzadas para analistas, tableros de control para gerentes, acceso mediante inter/entra net para usuarios internos/externos remotos, envío de información por dispositivos no estándares para usuarios internos/externos, etc.

Kimball se concentra sobre el proceso de creación de aplicaciones *"templetes"*. Comienza definiendo el concepto de la aplicación para usuario final y su rol en el acceso a la información del negocio. Brinda un marco metodológico bastante estándar en lo que ha desarrollo de aplicaciones (como piezas de software) se refiere. Divide el proceso de creación de las aplicaciones para usuarios finales en dos grandes fases: especificación y desarrollo. Clasifica a los usuarios según su perfil de consulta, desde usuarios con un perfil más estratégico y menos predecibles (power users) hasta usuarios netamente operacionales que consumen una serie de reportes estándares (final users) pasando por los usuarios gerenciales con uso de interfaces push-button (EIS users).

Kimball destaca, como uno de los requisitos a cumplir por las aplicaciones, la posibilidad de hacer análisis AD HOC. Análisis ad hoc es simplemente la habilidad para los usuarios de cambiar los parámetros sobre un reporte para crear sus propias versiones personalizadas de ese reporte. De esta forma se dará respuesta a necesidades como "Quiero ver este reporte, pero por mes en lugar de trimestre", "Puedo ver este reporte a nivel provincia en lugar de región?", "Puedo ver esta evolución de ventas mensual pero sólo de mi equipo de ventas?". Más importante aún es que las respuestas a estos interrogantes lo resuelven los propios usuarios sin la necesidad de la intervención del departamento sistemas y maximizando el tiempo de análisis por sobre el tiempo de construcción e integración de la información.

Kimball destaca cuatro pasos principales (siempre enfatizando el hecho de involucrar a los usuarios en cada uno de estos pasos) [Kim98]:

- Determinación del conjunto de templates iniciales (identificar reportes candidatos, clasificarlos y priorizarlos)
- Diseño de la estrategia de navegación dentro de la aplicación (esquema de pantallas, esquema de carpetas –directorios-, criterios de agrupamiento -por datos, por dueño, por regla del negocio, etc.-)

- Determinación de estándares (nombre de objetos, ubicación de objetos, formato de las salidas)
- Detalle de las especificaciones (definición: nombre, descripción o propósito, frecuencia, parámetros, restricciones, layout, etc.)

3.9 Desarrollo de Aplicaciones para Usuarios Finales

Siguiendo a la especificación de las aplicaciones para usuarios finales, el desarrollo de las aplicaciones de los usuarios finales involucra configuraciones del metadata y construcción de reportes específicos.

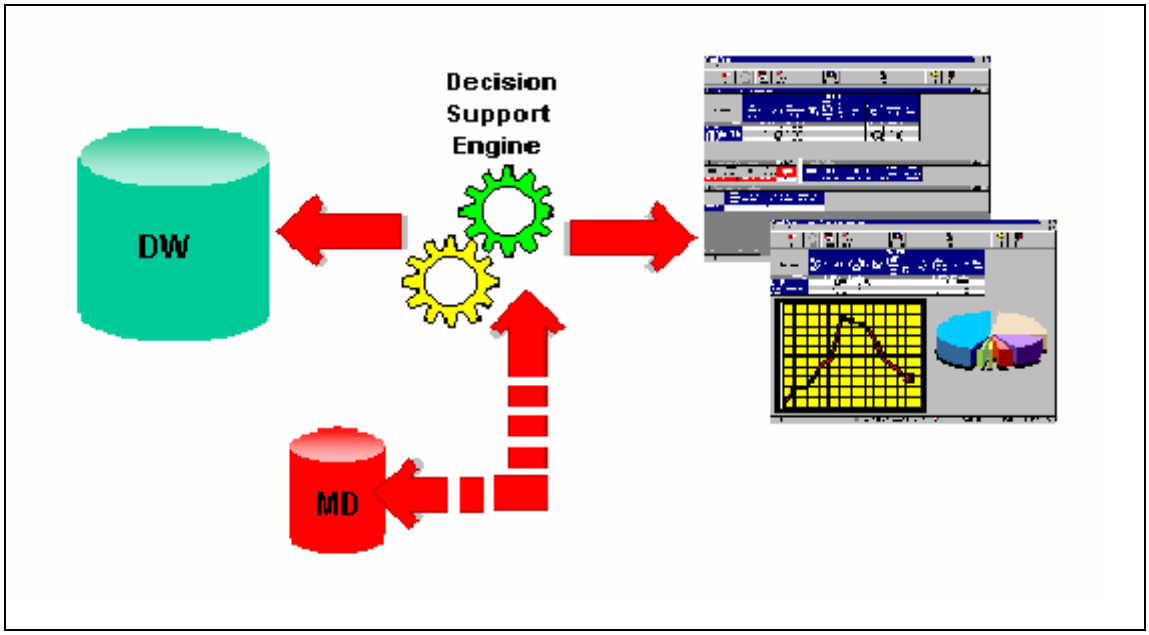


Figura Nº 3.26:"Configuración de Metadata y construcción de reportes"

Una vez que se ha cumplido con todos los pasos de la especificación y se tiene la posibilidad de trabajar con algunos datos de prueba, comienza el desarrollo de la aplicación [Kim98]:

- Selección de un enfoque de implementación
  - Basado en Web
  - Inter/Intra net
  - Usuarios altamente distribuidos
  - Manejo centralizado de nuevas versiones

- Herramienta propietaria
    - Mayor complejidad de uso
    - Para usuarios más capacitados
    - Instalación local
- EIS
    - Acceso estructurado
    - Secuencialidad de pantallas
    - Push-Button
- Interface personalizada
    - API (Application Programming Interface)
    - Desarrollos propios sobre la base de un conjunto de funcionalidades
- Desarrollo de la aplicación
    - Definición de herramienta de acceso al MetaData
    - Desarrollo de templates y esquema de navegación de la aplicación
    - Selección de reportes para pre-ejecución
- Prueba y verificación de datos
    - Descripciones
    - Información duplicada
    - Relaciones entre atributos
    - Consistencia e integridad de datos con sistemas fuentes
- Documentación y Roll Out
    - Retroalimentación con los resultados de la puesta en producción
- Mantenimiento
    - Nuevos templates
    - Incorporación de nuevos sistemas fuentes
    - Monitoreo de performance
    - Eliminación de templates en desuso

3.10 Implementación

La implementación representa la convergencia de la tecnología, los datos y las aplicaciones de usuarios finales accesible desde el escritorio del usuario del negocio. Hay varios factores extras que aseguran el correcto funcionamiento de todas estas

piezas, entre ellos se encuentran la capacitación, el soporte técnico, la comunicación, las estrategias de feedback. Todas estas tareas deben ser tenidas en cuenta antes de que cualquier usuario pueda tener acceso al data warehouse.

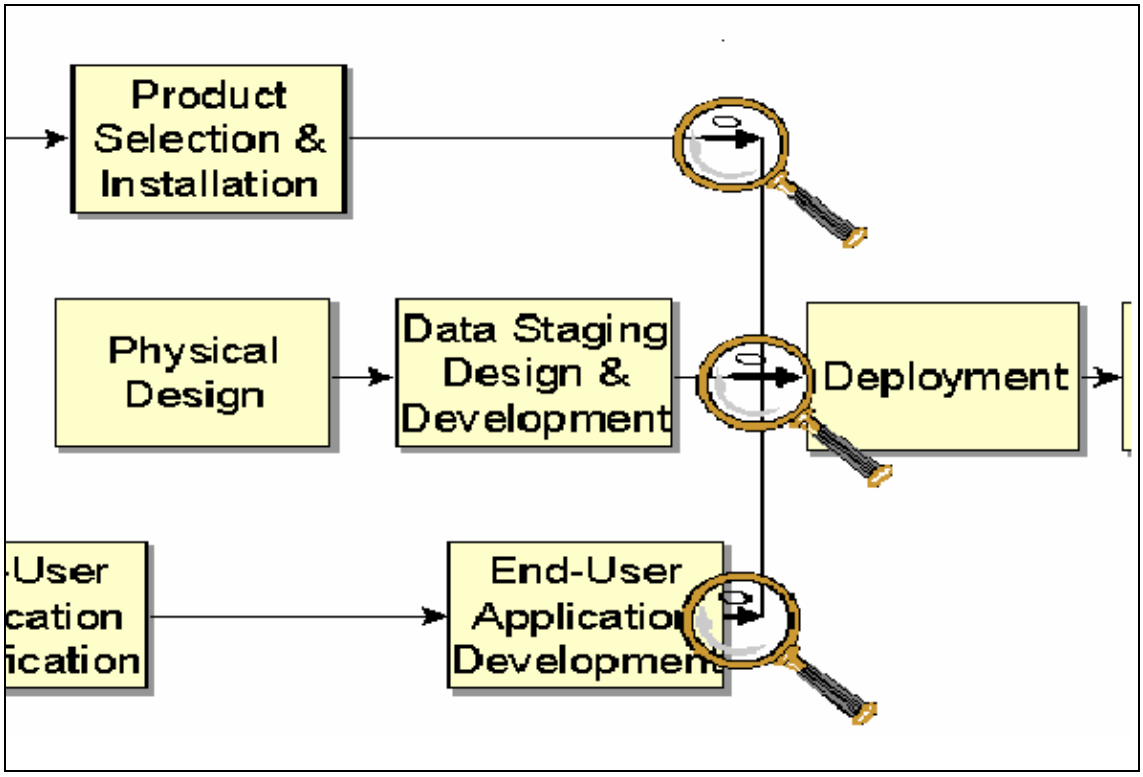


Figura Nº 3.27:"Implementación como convergencia de la tecnología, los datos y las aplicaciones"

La tecnología que reside en el escritorio del usuario es la última pieza que debe ser ubicada antes de la salida a producción (roll out o deployment). Desafortunadamente, afirma Kimball, las organizaciones frecuentemente subestiman el esfuerzo y el tiempo requerido para esta etapa. Kimball propone entonces una checklist sobre actividades que deberían ocurrir antes de la implantación para asegurar que la infraestructura correspondiente al ambiente del usuario este correcta. Esta checklist incluye: configuración de hardware, conexión a las bases, acceso a intranet o internet, direcciones LAN (si no son dinámicamente asignadas), auditorías de tecnología sobre las configuraciones en las que se encontraban las PCs, preveer actualizaciones de hardware y software (determinando responsables, proyecto o área de usuario), verificaciones de seguridad (logon de red y base de datos), prueba de procedimientos de instalación en una variedad de máquinas, planificación de instalación con la correspondiente educación a los usuarios, etc.

Ralph Kimball dedica gran parte de este capítulo a la educación de los usuarios finales afirmando que "una robusta estrategia de educación para usuarios de negocios es un prerrequisito para el éxito del proyecto". Entre sus recomendaciones para cumplir exitosamente con esta tarea se encuentran dos aspectos, por un lado el alcance de la educación, y por otro, la metodología de presentación. Para lo primero, debe instruirse al usuario en tres aspectos claves: contenido del warehouse, aplicación y herramientas de acceso. En cuanto a lo segundo, para los usuarios finales no es fácil entender los límites entre las aplicaciones, los contenidos y las herramientas del warehouse. Para los usuarios el data warehouse es un todo, no la suma de componentes discretos, por lo tanto la educación también debe reflejar la misma perspectiva. **No** debe interpretarse educación de los usuarios finales como educación sólo sobre las herramientas de acceso. La educación sobre herramientas es inútil al menos que se complemente con educación sobre el contenido del warehouse (cuáles datos están disponibles, qué significan, cómo se usan y para qué usarlos). Otro factor clave es la correcta entrega de niveles de educación según el perfil del usuario (usuario final, power user, etc.)

El armado de un ambiente de training (como subconjunto de datos de producción) es siempre recomendable, dado que los tiempos de respuesta en producción, si bien pueden ser apropiados para el análisis de la información, tal vez no sean los mejores para un curso de un día de duración. Además, el contenido del warehouse puede variar entre curso y curso haciendo más complicada la administración y mantenimiento del material educativo. Kimball establece dos premisas para regular las relaciones entre la educación y el warehouse.

Realizar educación de usuarios finales sólo si el warehouse está listo (en tiempo y forma) y establecer la política de "Sin Educación, entonces Sin Acceso". Cumpliendo estas dos premisas se puede garantizar el éxito del proyecto en esta etapa final del ciclo de vida.

El soporte es la otra columna que ayudará a implementar con éxito el proyecto. Se deberá definir claramente los niveles de soporte (a nivel aplicación, modelo, calidad de datos) de forma transparente para el usuario. La documentación juega un papel importantísimo en la ayuda a usuarios finales, dado que el datawarehouse evolucionará rápidamente, la documentación escrita puede quedar obsoleta también de forma

prematura. La documentación en línea empieza entonces a tener vida propia, ya sea dentro del warehouse en sí mismo o como herramientas anexas, como el data warehouse web site propuesto por Kimball.

Finalmente, la propuesta de Kimball se orienta a la metodología de implementación. Para ello propone un esquema de versionado (release framework). Primero se pasa por la versión Alpha, primera oportunidad para el grupo de trabajo de conducir una prueba del sistema de principio a fin.

Todos los componentes del sistema deben ser testeados (infraestructura técnica, extracción, transformación, carga, procedimientos de calidad, performance, templates, etc.). Típicamente es de naturaleza iterativa y sólo puede decirse que se termina esta versión cuando el grupo de trabajo confía plenamente en la calidad del warehouse como un todo (por lo tanto a la hora de estimar los tiempos debe tenerse en cuenta que este será el objetivo final de la versión Alpha por lo cual debe asignarse un tiempo prudencial). Durante la etapa Alpha un conjunto limitado de usuarios finales son provistos con acceso al warehouse (sólo aquellos que pertenecen al grupo de trabajo). Luego viene la versión Beta.

El objetivo de esta versión es conducir una prueba a nivel usuario de principio a fin. El grupo Beta está formado por los power users, la determinación de la cantidad de este grupo es también importante, pues si son pocos la aplicación no se probará adecuadamente, y si son muchos se estará salteado la etapa de testing cayendo directamente en una salida a producción. Una vez superadas estas dos versiones, al igual que en un proceso de diseño e implementación de software, llegamos a un estado GA (general availability). Lo que sugiere Kimball es que, todo cambio y/o modificación que se realice posteriormente al warehouse, pase internamente por un estado Alpha y Beta aunque externamente sea una nueva versión, release o pack desde el último GA.

3.11 Mantenimiento y crecimiento

Como se remarca siempre, Data Warehousing es un proceso (de etapas bien definidas, con comienzo y fin, pero de naturaleza espiral) pues acompaña a la evolución de la organización durante toda su historia. Se necesita continuar con los relevamientos de forma constante para poder seguir la evolución de las metas por conseguir. Según

afirma Kimball [Kim98], "*si se ha utilizado el BDL el data warehouse está preparado para evolucionar y crecer*". Al contrario de los sistemas tradicionales, los cambios en el desarrollo deben ser vistos como signos de éxito y no de falla. Es importante establecer las prioridades para poder manejar los nuevos requerimientos de los usuarios y de esa forma poder evolucionar y crecer.

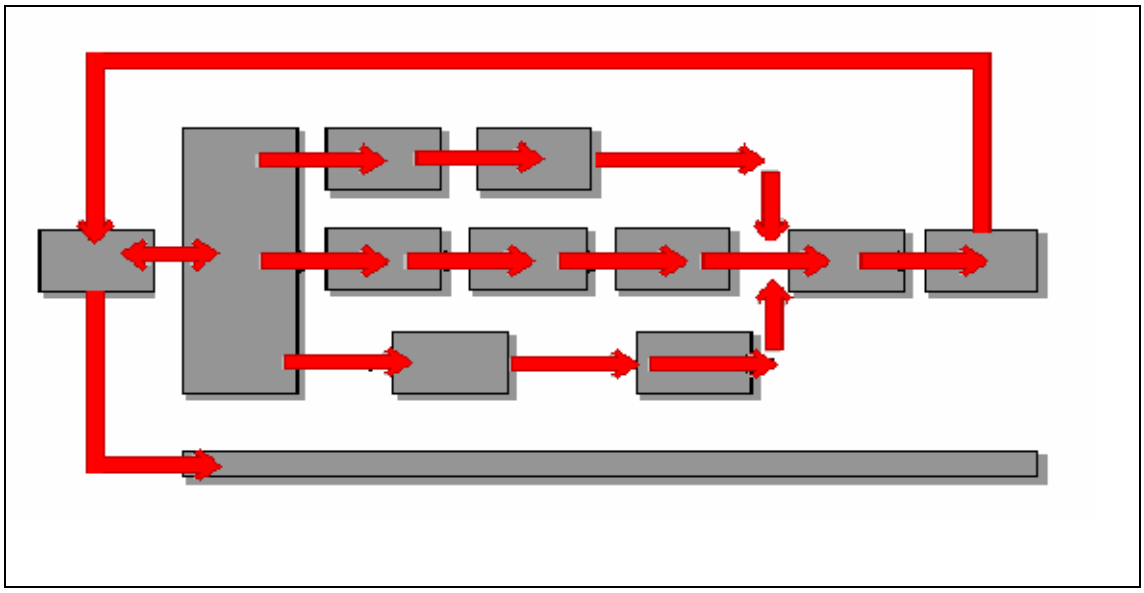


Figura Nº 3.28:"Data Warehousing como proceso de naturaleza espiral"

El sentido de las flechas fija claramente el concepto evolutivo y espiral del BDL.

Una vez que se ha construido e implantado el data warehouse no hay tiempo para el descanso, rápidamente debemos estar preparados para administrar el mantenimiento y crecimiento del mismo. Si bien las tareas pueden llegar a parecer similares a las tratadas en otras etapas del BDL, existe una diferencia clave: los usuarios están ahora accediendo al warehouse.

Kimball [Kim98] brinda entonces una serie de puntos a tener en cuenta para mantener exitosamente el warehouse. Entre ellos se destacan: el continuo soporte y la constante capacitación a usuarios de negocios, el manejo de la infraestructura (monitoreo de base de datos, tráfico, etc.), tuning de rendimiento sobre las consultas, mantenimiento del metadata y procesos ETLs. Otros aspectos involucran el monitoreo regular del cumplimiento de las expectativas sobre el warehouse (variables de medición del éxito del proyecto que se habían fijado en la etapa de relevamiento), relevamiento de casos de estudio (situaciones reales donde una decisión basada en información del warehouse tuvo impacto sobre el negocio –ROI-), constante publicidad interna del uso del warehouse (permitiendo acceso siempre y cuando se tenga la capacitación

correspondiente) y constante comunicación con los sectores de negocios y sistemas para asegurar la buena salud del datawarehouse.

En cuanto a cómo prepararse para el crecimiento del warehouse, Kimball recomienda la creación de un comité conformado por analistas del negocio y sistemas que establezca prioridades y procedimientos. El manejo de prioridades lo discrimina según sean mejoras menores o mayores (clasificadas según el impacto en el warehouse y excluyendo los errores, los cuales deben ser resueltos inmediatamente). Como mejoras menores incluye incorporación de datos sencillos, nuevas agregaciones, cambios en niveles superiores del modelo (atributos de alto nivel, lejanos a las bases tables), tareas que en definitiva no lleven más de días o semanas. Advierte sobre el razonamiento simplista de decidir no aplicar una mejora pues el cambio es muy menor, *"1000 pequeñas mejoras es una ventaja estratégica"*. Dentro de las mejoras de mayor impacto se encuentran las que modifican y/o crean bases tables o atributos asociados a las mismas, tareas que pueden disparar nuevos proyectos de meses de duración.

Finalmente, Kimball concluye su libro garantizando que el seguimiento y cumplimiento de todas las etapas BDL ayudará a obtener un warehouse mantenible y de controlar crecimiento. Como es ilustrado en el diagrama del BDL, se deberá volver a iterar sobre el mismo pasando nuevamente por cada una de las etapas para de esta forma administrar correctamente los cambios a warehouses existentes, como así también para respetar y aprovechar definiciones realizadas para otros proyectos.

3.12 Gerenciamiento del Proyecto

El gerenciamiento del proyecto asegura que las actividades del BDL se lleven en forma y sincronizadas. Como lo indica el diagrama, el gerenciamiento acompaña todo el ciclo de vida. Entre sus actividades principales se encuentra el monitoreo del estado del proyecto y la comunicación entre los requerimientos del negocio y las restricciones de información para poder manejar correctamente las expectativas en ambos sentidos.

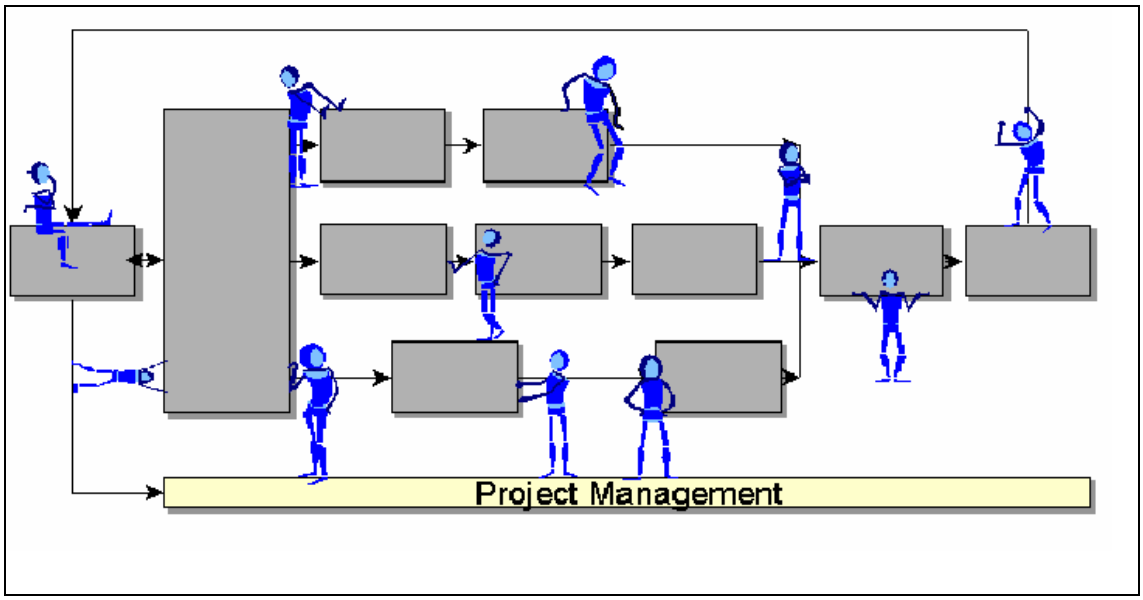


Figura Nº 3.29:"Gerenciamiento del proyecto"
El gerenciamiento del proyecto debe asegurar que las actividades del BDL se lleven en forma y sincronizadas.

3.13 Puntos problemáticos de un DW

Según identifica Jeff Bedell [Bed97], los principales puntos de atención que pueden llegar a complicar un proyecto de data warehousing se discriminan según las siguientes tres áreas:

- **Rutinas de Carga**. Incluye programas de extracción y limpieza de datos. Surgen problemas en este punto dada la falta de integración y estructura consistente (alineada) entre los sistemas fuentes.
- **Mantenimiento**. Dados los diferentes períodos de almacenamiento para OLTP y OLAP y el hecho de que los DW son sistemas secundarios de información, otro problema surge para sincronizar los datos entre los sistemas operacionales fuentes y los warehouses.
- **Tuning**. Dado los patrones de uso y los métodos de acceso típicos de los sistemas OLAP, diseñadores y administradores deben realizar cambios significativos a los implementados en el tuning de sistemas OLTPs.

Por nuestra experiencia en este campo, podemos decir que, además de estos puntos problemáticos propios de un ambiente OLAP, un proyecto de data warehousing no está ajeno a otros inconvenientes que son comunes a todos los proyectos tales como, falta de compromiso, relevamientos contradictorios, requerimientos incompletos, etc.

Capítulo 4: Metodología de Desarrollo para la Construcción del Datamart Establecimientos Caso Diners Club

Diners Club SA es una empresa cuyo objeto principal es la emisión y administración de tarjetas de crédito con las cuales los usuarios de éstas pueden realizar compras de bienes y/ o servicios en establecimientos comerciales afiliados al sistema que se establezca, sin necesidad de pagar en efectivo, para luego hacer el pago de tales operaciones a través de la empresa.

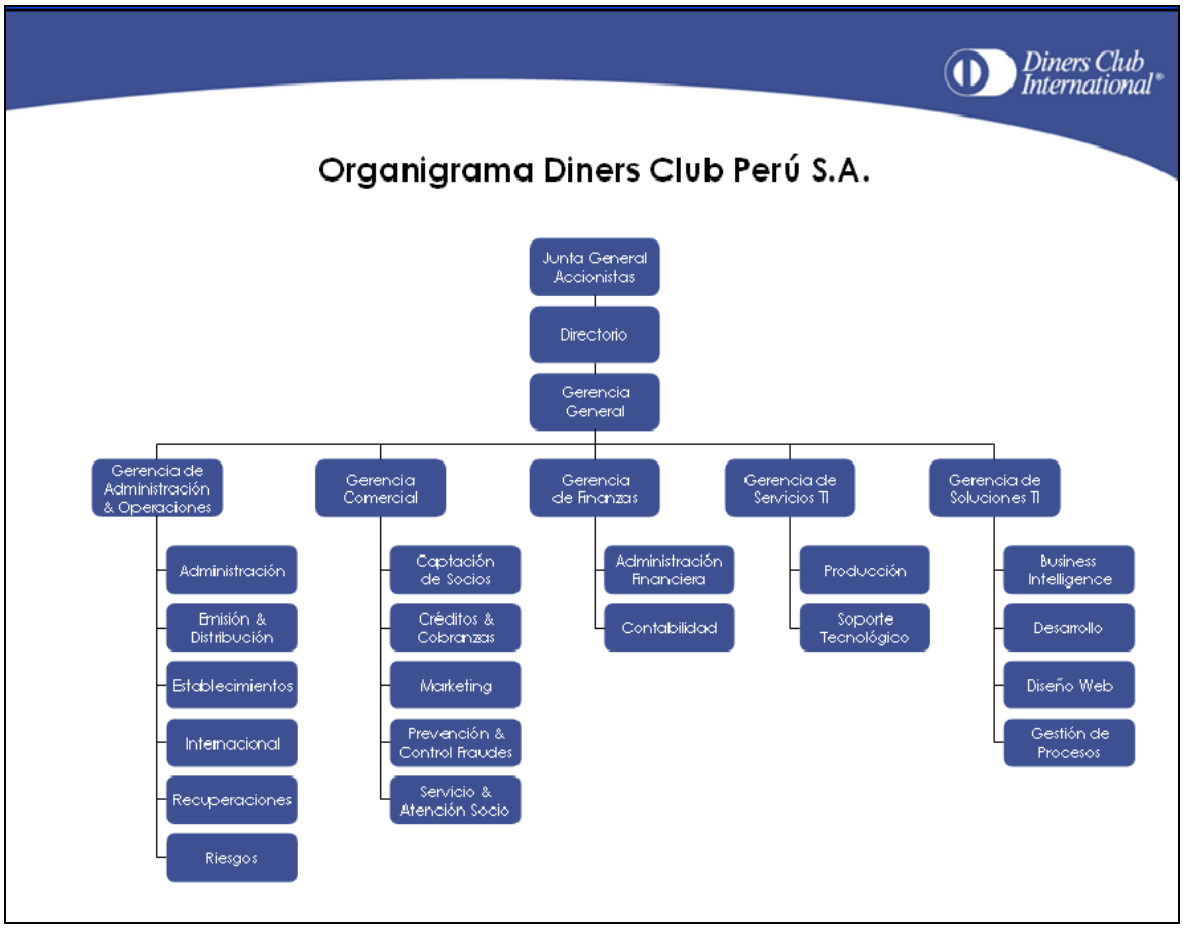


Figura Nº 4.1:"Organigrama Diners Club SA"

Para Diners Club sus principales socios externos son los comercios, pues los clientes utilizan las tarjetas de crédito en los establecimientos afiliados a nivel mundial, en este sentido el número de establecimientos afiliados se ha incrementado año a año, haciendo difícil y casi imposible poder analizar la información que compete al área de

establecimientos, lo cual llevó a proponer el desarrollo de una solución de inteligencia de negocios pues la cantidad de datos operativos producidos, desbordaban a líderes y gerentes, no alcanzando distinguir cuál es la situación real de la organización.

Para ello se tiene como referencia el marco metodológico propuesto por Ralph Kimball para el diseño e implementación de un Datawarehouse o un Data Mart, adaptando esta metodología a la situación real de la empresa se ha considerado las siguientes etapas para el desarrollo del proyecto. Ver figura 4.2

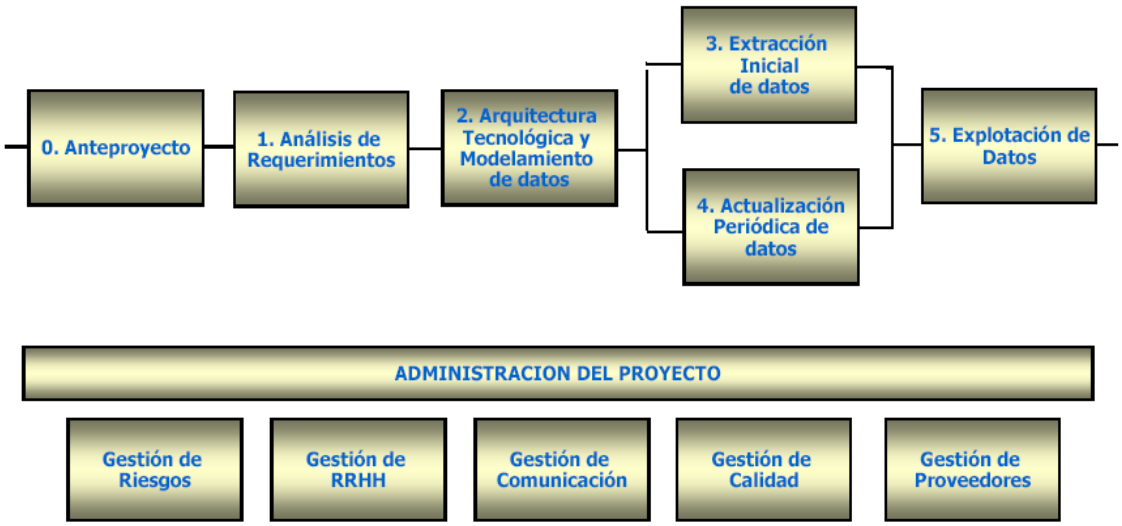


Figura Nº 4.2:"Metodología Implementación"

A continuación presentaremos los objetivos, tareas generales y entregables que deben llevarse a cabo en cada una de las etapas del presente proyecto.

4.1 Organización del Proyecto - Anteproyecto

En esta etapa se organiza e inicia el proyecto de implementación del Data Mart. Se lleva a cabo la reunión de inicio del Proyecto en la cual se define la organización y sus responsabilidades.

Objetivos

- Definir las metas y objetivos del proyecto
- Crear la organización y el ambiente a través del cual se va administrar el proyecto

Tareas

- Definir el Comité de Proyecto
- Definir el equipo de trabajo y responsabilidades

- Realizar la reunión de inicio del proyecto
- Identificar usuarios claves a ser entrevistados
- Desarrollar Plan de Trabajo detallado – Cronograma

Durante esta etapa se realizaron las reuniones con los responsables y usuarios del área de establecimientos, definiéndose el Comité de Proyecto, equipo de trabajo, responsabilidades y el plan de trabajo detallado.

4.2 Análisis de Requerimientos de Información

En esta etapa se definen los usuarios responsables, se establece el plan de entrevistas, se llevan a cabo las entrevistas a los usuarios claves y con la gerencia del área de establecimientos de la empresa Diners Club SA. A través de ellas los usuarios validan la información, la cual permitirá soportar sus requerimientos o estrategias de análisis. Se analizan los reportes actuales y se verifica la existencia de los requerimientos de información en las fuentes de datos, finalmente se formaliza el documento "Alcance de Requerimientos". Ver figura 4.3

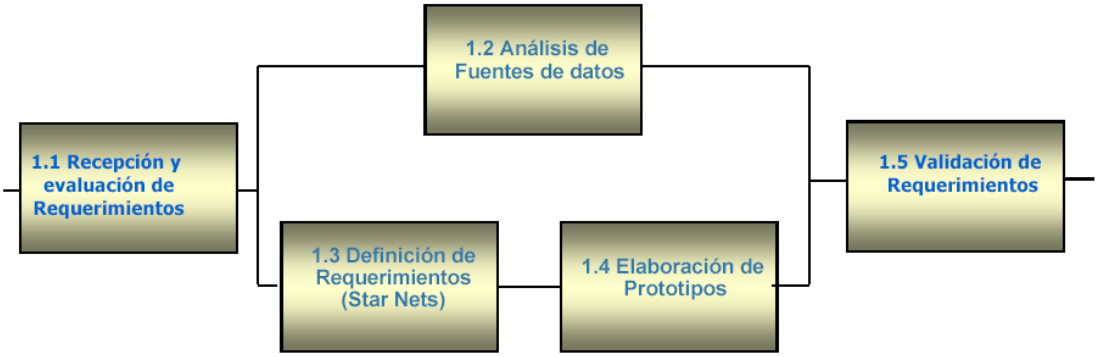


Figura Nº 4.3:"Análisis de Requerimientos"

Objetivos

- Comprender e identificar la información relevante que será almacenada en el Data Mart de Establecimientos
- Conocer los requerimientos y las vistas de análisis de los usuarios finales identificando las medidas y su relación con las dimensiones
- Identificar las fuentes de información para verificar su existencia, determinar su creación o posible proceso de transformación

Tareas

- Analizar y validar reportes actuales
- Identificar y validar los requerimientos de información
- Definir el alcance histórico y periodicidad de los datos
- Validar técnicamente la existencia de la información requerida en las fuentes de datos
- Verificar preliminarmente la calidad de los datos

Durante esta etapa se lleva a cabo las entrevistas con los usuarios claves con el fin de recoger los requerimientos de información de cada uno de ellos, lográndose identificar las necesidades de análisis, se identificaron los sistemas fuentes, se listaron y entendieron los requerimientos a través de la compresión del negocio.

Aquí se presenta un extracto de las necesidades encontradas luego de realizadas las entrevistas:

Establecimiento

- Se requiere complementar la información del establecimiento que existe actualmente en el Datamart Comercial.
- Se requiere contar con un reporte de bajas de establecimiento en un periodo de tiempo, indicando el motivo de bajas. El reporte debe permitir agrupar la información por otros atributo como por ejemplo giros, cadenas, zonas, distrito.
- El usuario requiere el TOP de establecimientos por facturación.
- El usuario requiere un Pareto de establecimientos frente a las colocaciones.
- Se requiere elaborar un reporte que muestre el detalle de establecimientos afiliados en el mes, indicando adicionalmente, el número de establecimientos activos hasta la emisión del reporte.
- Se acordó identificar aquellos datos que podrían añadirse al análisis de los establecimientos CAD

Visitas

- El Datamart debe permitir obtener información de los establecimientos a los cuales no se les ha realizado visitas en un periodo de tiempo. Este reporte debe permitir agrupar los datos tanto por ejecutivo, zona, giros, distrito, etc.
- Se requiere conocer el número de visitas por establecimientos en un rango de fechas, además obtener la fecha de la última visita
- Se requiere un reporte que contemple los establecimientos pendientes de visita por Procesos por el tema de capacitación.
- El Datamart debe permitir analizar a los centros nocturnos, con su dirección, última visita, giro y facturación.
- Se requiere un reporte del número de visitas realizadas a los establecimientos por ejecutivo. Incluyendo los siguientes datos: establecimiento, anotaciones, fecha de visita, contacto, motivo de visita.

Afiliaciones

- Se tiene como requerimiento elaborar un análisis según la procedencia de afiliación (DC o MC Procesos). Dicho análisis deberá realizarse por giros, sub giros, zonas, periodos, cadenas, zonas, representantes y ejecutivos.
- Se requiere conocer cuantos establecimientos de provincias han sido afiliados por Procesos en un determinado periodo de tiempo y de estos cuales fueron afiliados por un ejecutivo de Procesos Provincia y cuales por ejecutivos Procesos Lima.
- Se tiene como requerimiento elaborar un reporte para analizar el número de afiliaciones en un periodo de tiempo por giros y zonas. Identificando cuales fueron realizadas por el personal de Procesos del Perú y cuales por los ejecutivos Diners. Además realizar comparativos de afiliaciones por año, mes.

Pagos

- Se requiere analizar los pagos realizados a los establecimientos en un periodo de tiempo.

- Se requiere un reporte de los pagos no ejecutados a los establecimientos, motivo de no pago, comparaciones y estadísticas de establecimientos a los cuales no se ha efectuado pagos y la comisiones de estos

Ajustes

- Se desea contar con un reporte que permita analizar los ajustes que se ejecutan a los establecimientos versus la facturación, en un periodo de tiempo y por los motivos del ajuste.
- Se acordó considerar los ajustes por devolución de establecimientos

Costo de Materiales

- El Datamart deberá almacenar los costos de los materiales entregados a los establecimientos para determinar un ratio que se obtendrá de comparar el costo de materiales versus las comisiones. Dicho análisis deberá realizarse por giros, sub giros, zonas, periodos, cadenas, zonas
- Se requiere un reporte de los materiales entregados a los establecimientos para analizar la distribución de los mismos

Metas

- Se requiere registrar en el Datamart las metas de las visitas de los ejecutivos a los establecimientos. Actualmente el análisis del cumplimiento de las metas se realiza manualmente mediante la utilización del formato Conteo de Visitas.
- Se tiene como requerimiento elaborar reportes comparativos de la facturación y comisiones del año actual versus el año anterior por meses y por giros TOP.
- El usuario comento que cuenta con las metas de visitas por ejecutivo del año 2008
- Se acordó incluir las metas por número de afiliaciones al Datamart como parte del alcance del proyecto

Al final de esta etapa se presento el documento de “Alcance Funcional del Proyecto” el cual en una reunión fue validado por los involucrados en el proyecto.

4.3 Arquitectura, Tecnología y Modelamiento Dimensional

En esta etapa se realiza el diseño del modelo de base de datos del Data Mart utilizando técnicas de modelamiento dimensional especializadas y orientadas a responder las consultas de los usuarios finales en términos del negocio y no en términos de transacciones. Es fundamental que el modelo de datos soporte este tipo de consultas para obtener un tiempo de respuesta adecuado.

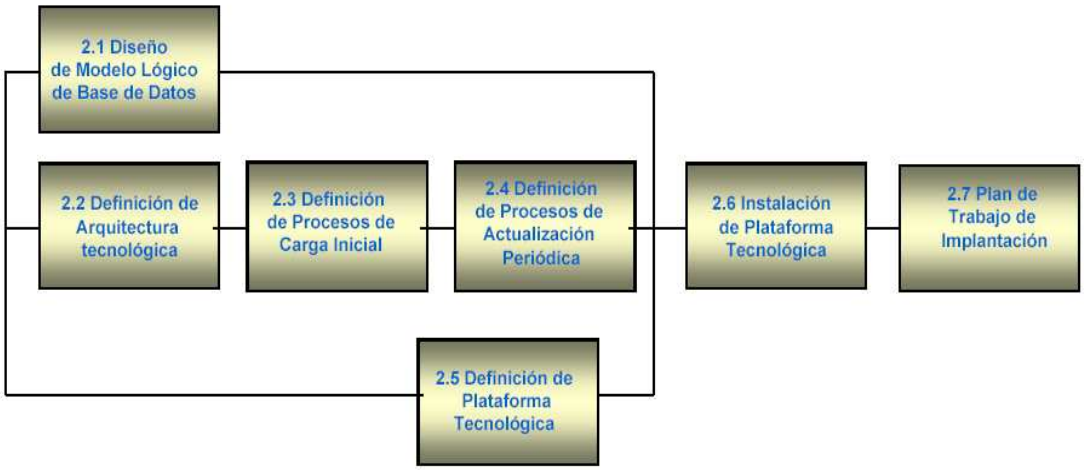


Figura Nº 4.4:"Arquitectura, tecnología y modelo dimensional"

Objetivos

- Diseñar el modelo lógico y físico de la base de datos para el Data Mart

Tareas

- Definir las relaciones entre entidades y atributos
- Identificar niveles de agregación de datos
- Identificar requerimientos de performance
- Diseñar el modelo lógico y físico de la base de datos
- Identificar las métricas
- Identificar las dimensiones, indicadores y jerarquías de análisis

Durante esta etapa se definieron las siguientes dimensiones, tablas de hechos y matriz de indicadores según los requerimientos de los usuarios:

4.3.1 Dimensiones y Tabla de Hechos

Dimensión Establecimiento

Campo	Descripción
GIRO	Giro del establecimiento
SUB-GIRO	Sub-giro del establecimiento
DESC_CADENA	Descripción de la cadena del establecimiento
AGRUPACION_COMERCIAL	Agrupación comercial a la que pertenece el establecimiento
CENTRO_COMERCIAL	Centro comercial a la que pertenece el establecimiento
DESC_FRANQUICIA	Descripción de la franquicia del establecimiento
RAZON_SOCIAL	Razón social del establecimiento
COD_ESTABLECIMIENTO	Código único de establecimiento
NOMBRE_ESTABLECIMIENTO	Nombre comercial del establecimiento
RUC	RUC del establecimiento
DESCR_ZONA	Zona del establecimiento
EJECUTIVO_ZONA	Ejecutivo de la zona del establecimiento
COD_ZONA	Código de la zona del establecimiento
ESTADO	Estado de la afiliación del establecimiento a Diners
FECHA_AFILIACION	Fecha de afiliación del establecimiento a Diners
FECHA_DESAFILIACION	Fecha en que el establecimiento se desafilió
AFILIADO_CONTADO_SOLES	Indicador si establecimiento se encuentra afiliado al sistema de trabajo CONTADO SOLES (si/no)
AFILIADO_CONTADO_DOLARES	Indicador si establecimiento se encuentra afiliado al sistema de trabajo PLAZOS SOLES (si/no)
AFILIADO_PLAZOS_DOLARES	Indicador si establecimiento se encuentra afiliado al sistema de trabajo CONTADO DOLARES (si/no)

Campo	Descripción
AFILIADO_CONTADO_DOLARES	Indicador si establecimiento se encuentra afiliado al sistema de trabajo PLAZOS DOLARES (si/no)
FLAG_AFILIADO_CAD	Indica si el establecimiento está afiliado a CAD
FLAG_AFILIADO_DELIVERY	Indica si el establecimiento está afiliado a DELIVERY
FLAG_AFILIADO_INTERNET	Indica si el establecimiento está afiliado a INTERNET
FLAG_AFILIADO_MOTO	Indica si el establecimiento está afiliado a MOTO
ESTEBLC_DINERS	Indica si el establecimiento es interno (si/no)
ESTABLECIMIENTO_TOP	Indica si el establecimiento es TOP (si/no)
PORC_COMISION_CONTADO_SOL	Porcentaje de comisión aplicable a consumos contado soles en el establecimiento
PORC_COMISION_CONTADO_DOL	Porcentaje de comisión aplicable a consumos contado dólares en el establecimiento
PORC_COMISION_FINANCIADO_SOL	Porcentaje de comisión aplicable a consumos financiados soles en el establecimiento
PORC_COMISION_FINANCIADO_DOL	Porcentaje de comisión aplicable a consumos financiados dólares en el establecimiento
DIRECCION_OFICINA	Dirección de oficina del establecimiento
COD_INTES	Código de cadena internacional
DESC_INTES	Descripción de cadena internacional
DEPARTAMENTO_ESTABLECIMIENTO	Departamento de dirección Comercial del establecimiento
PROVINCIA_ESTABLECIMIENTO	Provincia de dirección Comercial del establecimiento
DISTRITO_ESTABLECIMIENTO	Distrito de dirección Comercial del establecimiento
CODIGO_POSTAL	Código Postal de dirección Comercial del establecimiento
DIRECCION_COMERCIAL	Dirección Comercial del establecimiento
DEPARTAMENTO_ESTABLECIMIENTO_OFICINA	Departamento de dirección oficina del establecimiento
PROVINCIA_ESTABLECIMIENTO_OFICINA	Provincia de dirección oficina del establecimiento
DISTRITO_ESTABLECIMIENTO_OFICINA	Distrito de dirección oficina del establecimiento
CODIGO_POSTAL_OFICINA	Código Postal de dirección oficina del establecimiento

Campo	Descripción
DIRECCION_OFICINA	Dirección de oficina del establecimiento
REPRESENTANTE_LEGAL	Representante Legal
EJECUTIVO_AFILIO	Ejecutivo que Afilio
FORMA_PAGO_CONTADO_SOLES	Forma Pago Contado Soles
BANCO_PAGO_CONTADO_SOLES	Banco Pago Contado Soles
TIPO_CTA_CONTADO_SOLES	Tipo de Cuenta Contado Soles
MONEDA_CTA_CONTADO_SOLES	Moneda de Cuenta Contado Soles
NRO_CTA_CONTADO_SOLES	Número de Cuenta Contado Soles
FORMA_PAGO_PLAZOS_SOLES	Forma Pago Plazos Soles
BANCO_PAGO_PLAZOS_SOLES	Banco Pago Plazos Soles
TIPO_CTA_PLAZOS_SOLES	Tipo de Cuenta Plazos Soles
MONEDA_CTA_PLAZOS_SOLES	Moneda de Cuenta Plazos Soles
NRO_CTA_PLAZOS_SOLES	Número de Cuenta Plazos Soles
FORMA_PAGO_CONTADO_DOLARES	Forma Pago Contado Dólares
BANCO_PAGO_CONTADO_DOLARES	Banco Pago Contado Dólares
TIPO_CTA_CONTADO_DOLARES	Tipo de Cuenta Contado Dólares
MONEDA_CTA_CONTADO_DOLARES	Moneda de Cuenta Contado Dólares
NRO_CTA_CONTADO_DOLARES	Número de Cuenta Contado Dólares
FORMA_PAGO_PLAZOS_DOLARES	Forma Pago Plazos Dólares
BANCO_PAGO_PLAZOS_DOLARES	Banco Pago Plazos Dólares
TIPO_CTA_PLAZOS_DOLARES	Tipo de Cuenta Plazos Dólares
MONEDA_CTA_PLAZOS_DOLARES	Moneda de Cuenta Plazos Dólares
NRO_CTA_PLAZOS_DOLARES	Número de Cuenta Plazos Dólares

Campo	Descripción
RETENCION_LEGAL	Indicador Retención Legal (SI, NO)
MOTIVO_DESAFILIACION	Motivo Desafiliación
CEP_SOLES	Código de establecimiento de MC Proceso Soles (Código marcado como principal)
CEP_DOLARES	Código de establecimiento de MC Proceso Dólares (Código marcado como principal)
POS	Indicador si establecimiento cuenta con POS (SI/NO)
COD_IATA	Código IATA
TELEFONO_PRINCIPAL	Teléfono Principal
COD_COMISION	Código de Comisión

Tabla Nº 4.1:"Dimensión Establecimiento"

Dimensión Artículo

Campo	Descripción
COD_ARTICULO	Código de artículo
NOMBRE_ARTICULO	Nombre de artículo
UNIDAD_MEDIDA	Unidad de medida (UND, CJA, FCO, etc.)

Tabla Nº 4.2:"Dimensión Artículo"

Dimensión TipoVisita

Campo	Descripción
COD_TIPOVISITA	Tipo de visita
DESC_TIPOVISITA	Descripción tipo de visita (capacitación, mantenimiento, afiliación, etc.)

Tabla Nº 4.3:"Dimensión Tipo Visita"

Dimensión Representante

Campo	Descripción
COD_REPRESENTANTE	Código de representante

Campo	Descripción
NOMBRE_REPRESENTANTE	Nombre del representante

Tabla Nº 4.4:"Dimensión Representante"

Dimensión Ejecutivo

Campo	Descripción
COD_TIPO_EMPL	Código del tipo de empleado
DESCR_TIPO_EMPL	Descripción del tipo de empleado (Ej.: gestor, sectorista, usuario call center, etc)
COD_EMPLEADO	Código del empleado
DESCR_EMPLEADO	Descripción del empleado

Tabla Nº 4.5:"Dimensión Ejecutivo"

Dimensión Tiempo

Campo	Descripción
PERIODODIA	Fecha calendario (YYYYMMDD)
ANIO	Año
SEMESTRE	Semestre (1er Semestre / 2do Semestre)
TRIMESTRE	Trimestre (1er Trimestre, 2do….)
DESCR_MES	Nombre de mes (Enero, Febrero, …., Diciembre)
NUM_MES	Número de mes en el año (1,2,3….,12)
SEMANA	Semana en el año (Semana 1 …. Semana 52)
DESCR_DIA	Nombre de día de la semana (Lunes, Martes, …. , Domingo)
NUM_DIA	Número de día en el mes (1,…., 31)
FECHA	Fecha calendario (YYYY/MM/DD)

Tabla Nº 4.6:"Dimensión Tiempo"

Dimensión Forma Pago Establecimiento

Campo	Descripción
COD_FORMAPAGO	Código forma de pago a establecimiento
DESC_FORMAPAGO	Descripción forma de pago a establecimiento

Tabla Nº 4.7:"Dimensión Forma Pago Establecimiento"

Dimensión Geografía

Campo	Descripción
DEPARTAMENTO	Departamento
PROVINCIA	Provincia
DISTRITO	Distrito
COD_POSTAL	Código postal

Tabla Nº 4.8:"Dimensión Geografía"

Tabla de Hechos Visita

Campo	Descripción
NUMERO_VISITA	Identificador de la visita
COD_ESTABLECIMIENTO	Establecimiento visitado
COD_EJECUTIVO	Ejecutivo
COD_REPRESENTANTE	Representante
CONTACTO	Nombre del Contacto
COD_TIPOVISITA	Tipo de visita
ACCIONES_REALIZADAS	Acción realizada
SUGERENCIAS	Sugerencia

Campo	Descripción
FECHA_VISITA	Fecha de la visita al establecimiento
ESTADO	Estado de la visita (ACTIVO, ELIMINADO)
ESTADO_ATENDIDO	Estado de atendido (PENDIENTE, ATENDIDA)
FECHA_GENERACION	Fecha de generación de visita
FECHA_DIGITACION	Fecha de digitación de la visita
MOTIVO_AFILIA_PENDIENTE	Motivo de la Afiliación Pendiente(POR DOCUMENTACION,POR COMISIÓN, POR PLAZOS)
PRIMERA_VISITA	Indicador de primera visita al establecimiento (Si/No)
VISITA_NOCTURNA	Indicador de visita nocturna (SI/NO)
EJECUTIVO_SOLICITO	Sólo para caso de solicitar materiales o entrega de documentos

Indicador

NRO_VISITAS	Cantidad de visitas realizadas

Tabla Nº 4.9:"Tabla de Hechos Visita"

Tabla de Hechos Entrega de Materiales

Campo	Descripción
NUMERO_VISITA	Identificador de la visita
COD_ARTICULO	Artículo
DESCARGA	Ejecutivo a quien se le descontó su stock (quien solicita o quien entrega)

Indicadores

CANTIDAD_ENTREGADA	Número de artículos entregados al establecimiento
COSTO_ARTICULO_SOLES	Costo de articulo expresado en soles
COSTO_ARTICULO_DOLARES	Costo de articulo expresado en dólares

Tabla Nº 4.10:"Tabla de Hechos Entrega de Materiales"

Tabla de Hechos Ajustes

Campo	Descripción
COD_ESTABLECIMIENTO	Establecimiento
ORIGEN	Origen del memo (POR SOCIO, POR ESTABLECIMIENTO)
TIPO_MEMO	Tipo de Memorando(CARGOEST,ABONOSOC/CARGOEST)
RECAP	Nro de Recap de consumo
ITEM	Nro de item del recap
NRO_TICKET	Nro de Ticket
NRO_TARJETA	Nro de Tarjeta
FECHA_INGRESO	Fecha Ingreso del ajuste
FECHA_ACTIVACION	Fecha de Activación
MONEDA	Moneda (S/. O US$)
TIPO_AJUSTE	Tipo del Ajuste (CARGO,ABONO)
COMPROBANTE_PAGO	Comprobante de pago
ESTADO	Estado(PROCESADO, PAGADO, SALDO,xASOCIAR, ASOCIADO)
ID_MEMO	Id Memorando
MOTIVO	Motivo de Ajuste (ANULACION, DEVOLUCION PARCIAL,DUPLICIDAD…)

Indicadores

IMPORTE_AJUSTE_SOLES	Importe del Ajuste expresado en Soles
IMPORTE_AJUSTE_DOLARES	Importe del Ajuste expresado en Dólares
SALDO_SOLES	Saldo del Ajuste expresado Soles
SALDO_DOLARES	Saldo del Ajuste expresado Dólares

Tabla Nº 4.11:"Tabla de Hechos Ajustes"

Tabla de Hechos Pago

Campo	Descripción
COD_ESTABLECIMIENTO	Establecimiento
BANCO	Banco donde se realiza el pago
COD_FORMAPAGO	Forma de pago
TIPO_CUENTA	Tipo de cuenta Cuenta Corriente / Ahorro
NRO_CUENTA_CHEQUE	Número de cuenta / cheque
FECHA_PROGRAMADA	Fecha programada
FECHA_PAGO	Fecha de pago
MONEDA_PAGO	Moneda S/.o US$ que se realizará el pago
COMPROBANTE	Número de comprobante
FECHA_ENTREGA_CHEQUE	Fecha de entrega del cheque
ESTADO	Estado de pago
MONEDA_ORIGEN	Monde S/. o US$ en que ingresaron los consumos

Indicadores

CARGO_SOLES	Cargo sobre el pago expresado en soles
CARGO_DOLARES	Cargo sobre el pago expresado en dólares
ABONO_SOLES	Abono sobre el pago expresado en soles
ABONO_DOLARES	Abono sobre el pago expresado en dólares
IMPORTE_PAGO_SOLES	Importe original del pago expresado en soles
IMPORTE_PAGO_DOLARES	Importe original del pago expresado en dólares
PAGO_NETO_SOLES	Pago neto expresado en soles (incluyendo ajustes)

PAGO_NETO_DOLARES	Pago neto expresado en dólares (incluyendo ajustes)

Tabla Nº 4.12:"Tabla de Hechos Pagos"

Tabla de Hechos Proyección de Afiliaciones

Campo	Descripción
ZONA_PROYECCION	Zona de Ejecutivo
FECHA_PROYECCION	Fecha de proyección

Indicador

NRO_AFILIACION_PROYECTADA	Proyección

Tabla Nº 4.13:"Tabla de Hechos Proyección de Afiliaciones"

Tabla de Hechos Proyección de Visitas

Campo	Descripción
COD_EJECUTIVO	Ejecutivo
FECHA_PROYECCION	Fecha de proyección
COD_TIPOVISITA	Tipo de visita

Indicador

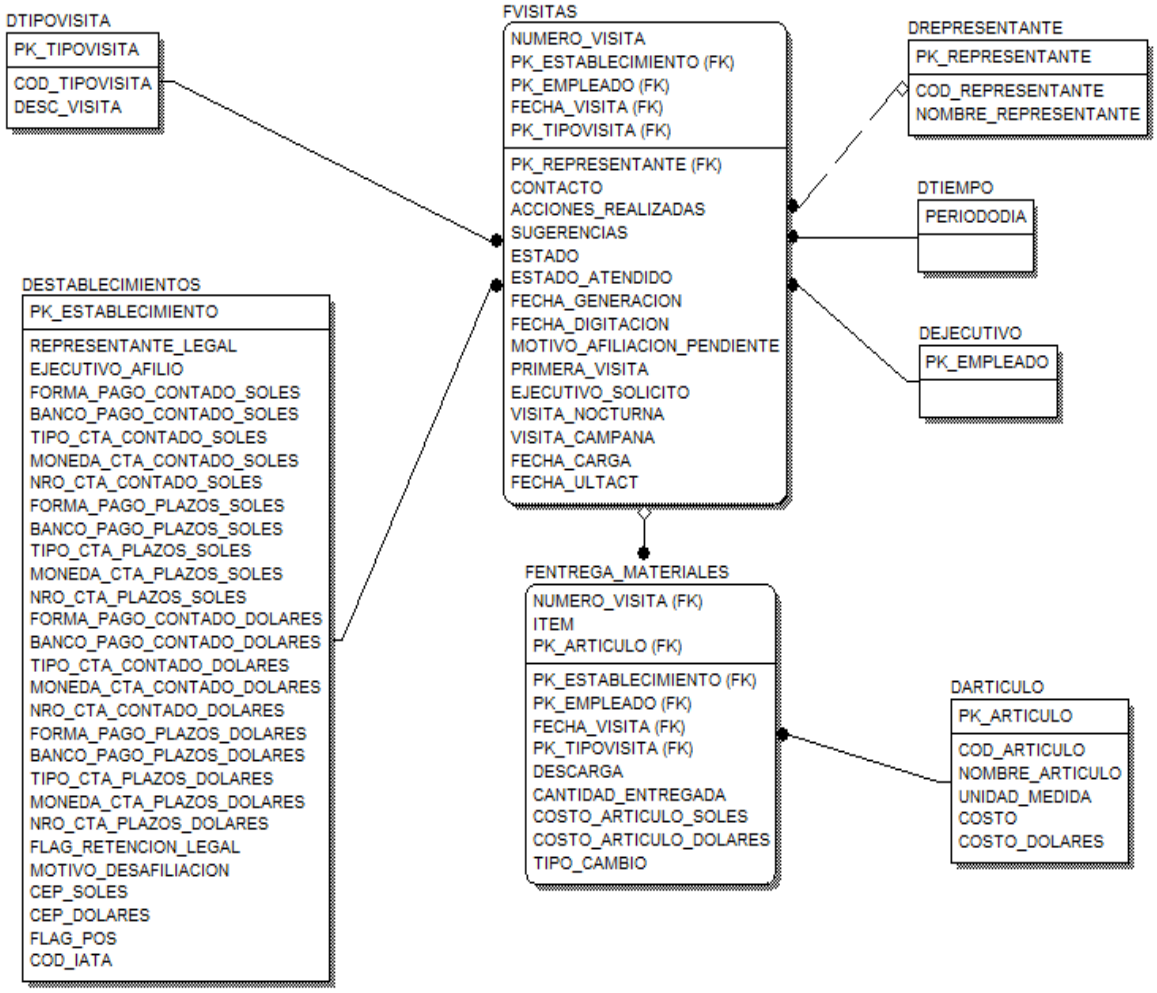
NRO_VISITAS_PROYECTADA	Proyección

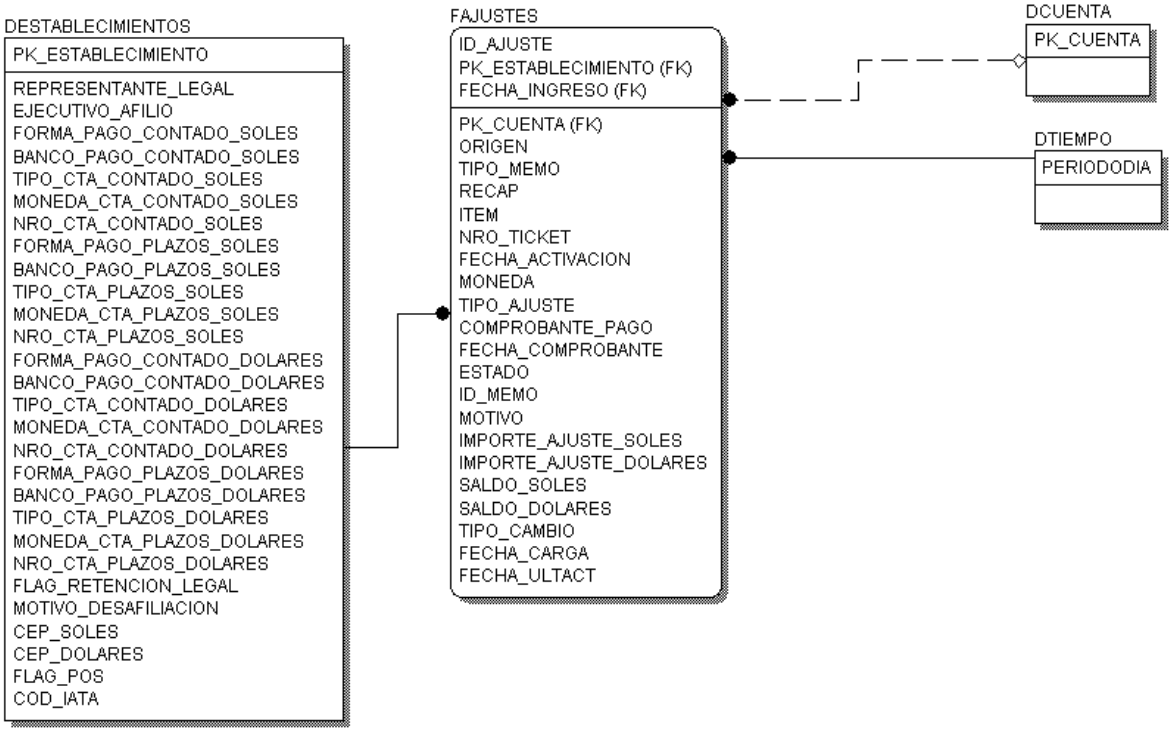
Tabla Nº 4.14:"Tabla de Hechos Proyección de Visitas"

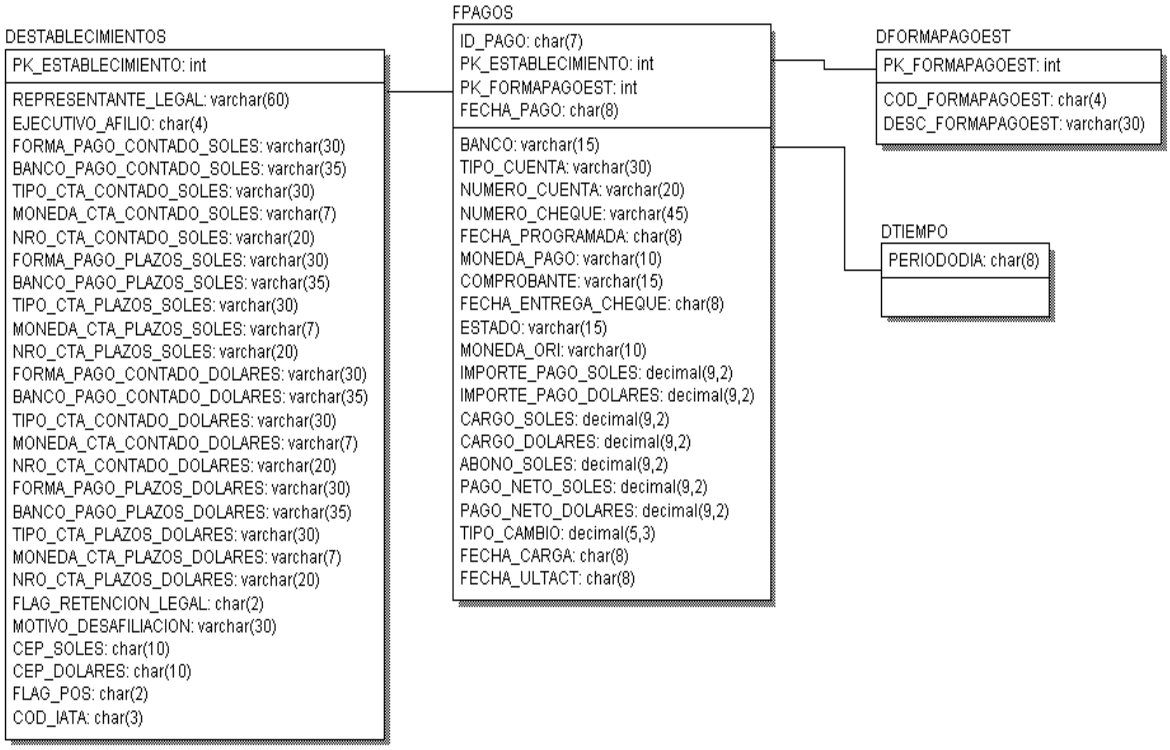
4.3.2 **Matriz de Indicadores** Tabla Nº 4.15

	INDICADORES																		
	VISITAS	**AFILIACIONES**	**AJUSTES**				**PAGOS**								**ENTREGA MATERIALES**			**PROYECCION DE VISITAS**	**PROYECCION DE AFILIACION**
DIMENSIONES	NRO_VISITAS	NRO_AFILIACIONES	IMPORTE_AJUSTE_SOLES	IMPORTE_AJUSTE_DOLARES	SALDO_SOLES	SALDO_DOLARES	CARGO_SOLES	CARGO_DOLARES	ABONO_SOLES	ABONO_DOLARES	IMPORTE_PAGO_SOLES	IMPORTE_PAGO_DOLARES	PAGO_NETO_SOLES	PAGO_NETO_DOLARES	CANTIDAD_ENTREGADA	COSTO_ARTICULO_SOLES	COSTO_ARTICULO_DOLARES	NRO_VISITAS_PROYECTADA	NRO_AFILIACIONES_PROYECTADA
ESTABLECIMIENTO	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
VISITA	✓														✓	✓	✓	✓	
ARTICULO	✓														✓	✓	✓		
FORMAPAGO							✓	✓	✓	✓	✓	✓	✓	✓					
REPRESENTANTE	✓	✓													✓	✓	✓		
EJECUTIVOS	✓	✓													✓	✓	✓	✓	✓
TIEMPO	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

4.3.3 Modelo Lógico




DESTABLECIMIENTOS
PK_ESTABLECIMIENTO
REPRESENTANTE_LEGAL
EJECUTIVO_AFILIO
FORMA_PAGO_CONTADO_SOLES
BANCO_PAGO_CONTADO_SOLES
TIPO_CTA_CONTADO_SOLES
MONEDA_CTA_CONTADO_SOLES
NRO_CTA_CONTADO_SOLES
FORMA_PAGO_PLAZOS_SOLES
BANCO_PAGO_PLAZOS_SOLES
TIPO_CTA_PLAZOS_SOLES
MONEDA_CTA_PLAZOS_SOLES
NRO_CTA_PLAZOS_SOLES
FORMA_PAGO_CONTADO_DOLARES
BANCO_PAGO_CONTADO_DOLARES
TIPO_CTA_CONTADO_DOLARES
MONEDA_CTA_CONTADO_DOLARES
NRO_CTA_CONTADO_DOLARES
FORMA_PAGO_PLAZOS_DOLARES
BANCO_PAGO_PLAZOS_DOLARES
TIPO_CTA_PLAZOS_DOLARES
MONEDA_CTA_PLAZOS_DOLARES
NRO_CTA_PLAZOS_DOLARES
FLAG_RETENCION_LEGAL
MOTIVO_DESAFILIACION
CEP_SOLES
CEP_DOLARES
FLAG_POS
COD_IATA
FAJUSTES
ID_AJUSTE
PK_ESTABLECIMIENTO (FK)
FECHA_INGRESO (FK)
PK_CUENTA (FK)
ORIGEN
TIPO_MEMO
RECAP
ITEM
NRO_TICKET
FECHA_ACTIVACION
MONEDA
TIPO_AJUSTE
COMPROBANTE_PAGO
FECHA_COMPROBANTE
ESTADO
ID_MEMO
MOTIVO
IMPORTE_AJUSTE_SOLES
IMPORTE_AJUSTE_DOLARES
SALDO_SOLES
SALDO_DOLARES
TIPO_CAMBIO
FECHA_CARGA
FECHA_ULTACT
DCUENTA
PK_CUENTA
DTIEMPO
PERIODODIA


DESTABLECIMIENTOS
PK_ESTABLECIMIENTO: int
REPRESENTANTE_LEGAL: varchar(60)
EJECUTIVO_AFILIO: char(4)
FORMA_PAGO_CONTADO_SOLES: varchar(30)
BANCO_PAGO_CONTADO_SOLES: varchar(35)
TIPO_CTA_CONTADO_SOLES: varchar(30)
MONEDA_CTA_CONTADO_SOLES: varchar(7)
NRO_CTA_CONTADO_SOLES: varchar(20)
FORMA_PAGO_PLAZOS_SOLES: varchar(30)
BANCO_PAGO_PLAZOS_SOLES: varchar(35)
TIPO_CTA_PLAZOS_SOLES: varchar(30)
MONEDA_CTA_PLAZOS_SOLES: varchar(7)
NRO_CTA_PLAZOS_SOLES: varchar(20)
FORMA_PAGO_CONTADO_DOLARES: varchar(30)
BANCO_PAGO_CONTADO_DOLARES: varchar(35)
TIPO_CTA_CONTADO_DOLARES: varchar(30)
MONEDA_CTA_CONTADO_DOLARES: varchar(7)
NRO_CTA_CONTADO_DOLARES: varchar(20)
FORMA_PAGO_PLAZOS_DOLARES: varchar(30)
BANCO_PAGO_PLAZOS_DOLARES: varchar(35)
TIPO_CTA_PLAZOS_DOLARES: varchar(30)
MONEDA_CTA_PLAZOS_DOLARES: varchar(7)
NRO_CTA_PLAZOS_DOLARES: varchar(20)
FLAG_RETENCION_LEGAL: char(2)
MOTIVO_DESAFILIACION: varchar(30)
CEP_SOLES: char(10)
CEP_DOLARES: char(10)
FLAG_POS: char(2)
COD_IATA: char(3)
FPAGOS
ID_PAGO: char(7)
PK_ESTABLECIMIENTO: int
PK_FORMAPAGOEST: int
FECHA_PAGO: char(8)
BANCO: varchar(15)
TIPO_CUENTA: varchar(30)
NUMERO_CUENTA: varchar(20)
NUMERO_CHEQUE: varchar(45)
FECHA_PROGRAMADA: char(8)
MONEDA_PAGO: varchar(10)
COMPROBANTE: varchar(15)
FECHA_ENTREGA_CHEQUE: char(8)
ESTADO: varchar(15)
MONEDA_ORI: varchar(10)
IMPORTE_PAGO_SOLES: decimal(9,2)
IMPORTE_PAGO_DOLARES: decimal(9,2)
CARGO_SOLES: decimal(9,2)
CARGO_DOLARES: decimal(9,2)
ABONO_SOLES: decimal(9,2)
PAGO_NETO_SOLES: decimal(9,2)
PAGO_NETO_DOLARES: decimal(9,2)
TIPO_CAMBIO: decimal(5,3)
FECHA_CARGA: char(8)
FECHA_ULTACT: char(8)
DFORMAPAGOEST
PK_FORMAPAGOEST: int
COD_FORMAPAGOEST: char(4)
DESC_FORMAPAGOEST: varchar(30)
DTIEMPO
PERIODODIA: char(8)

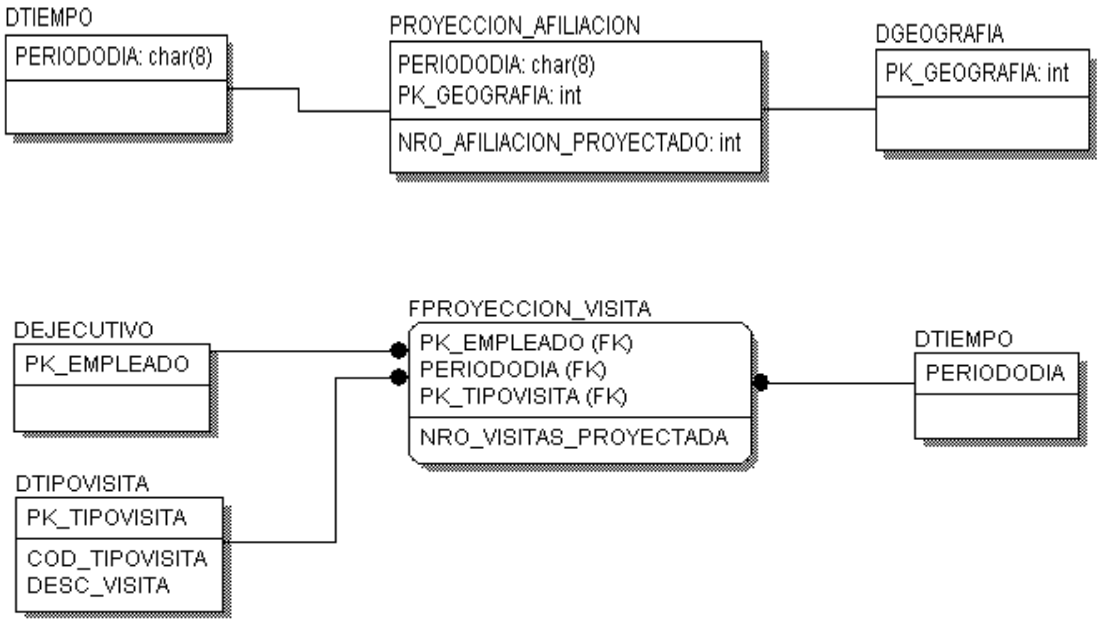
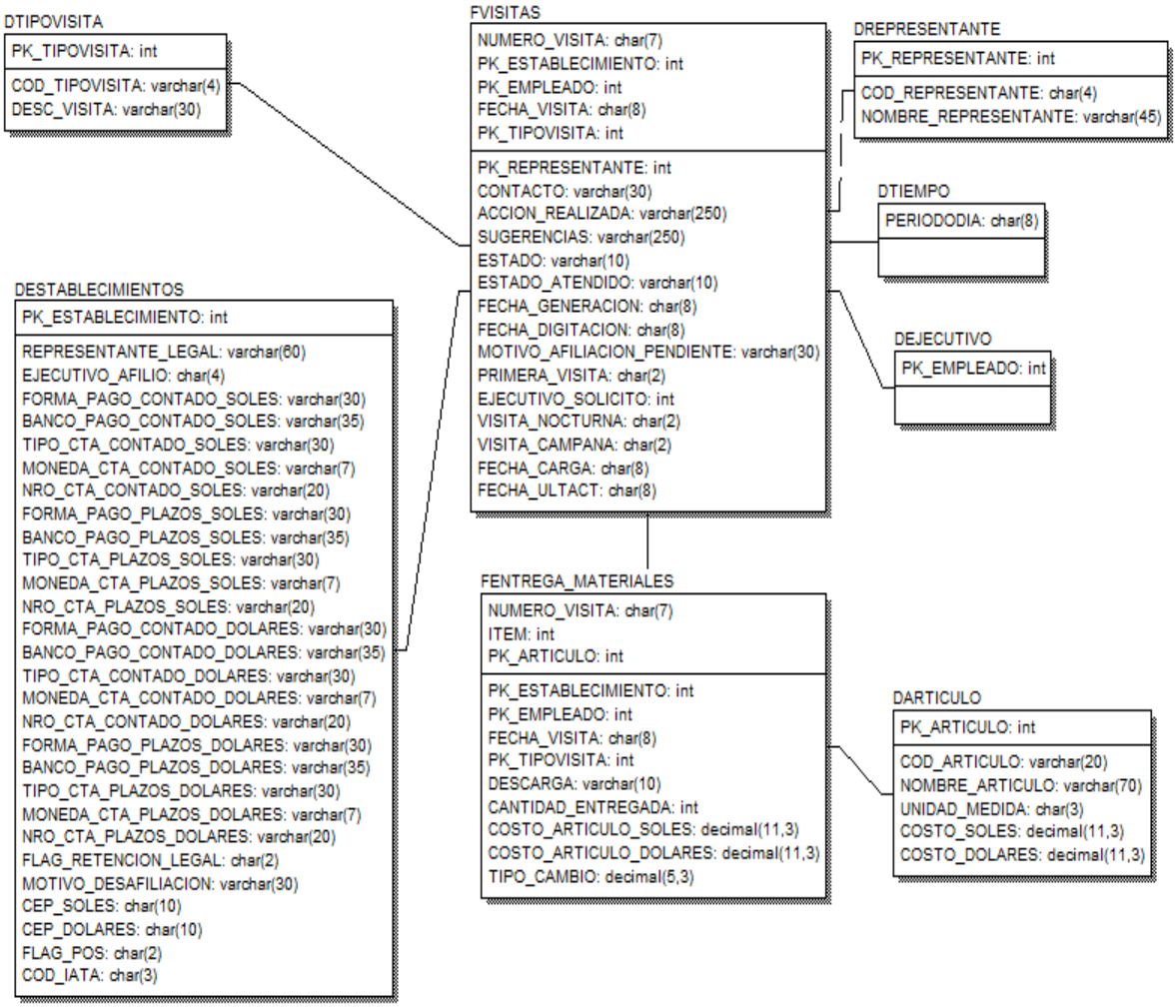
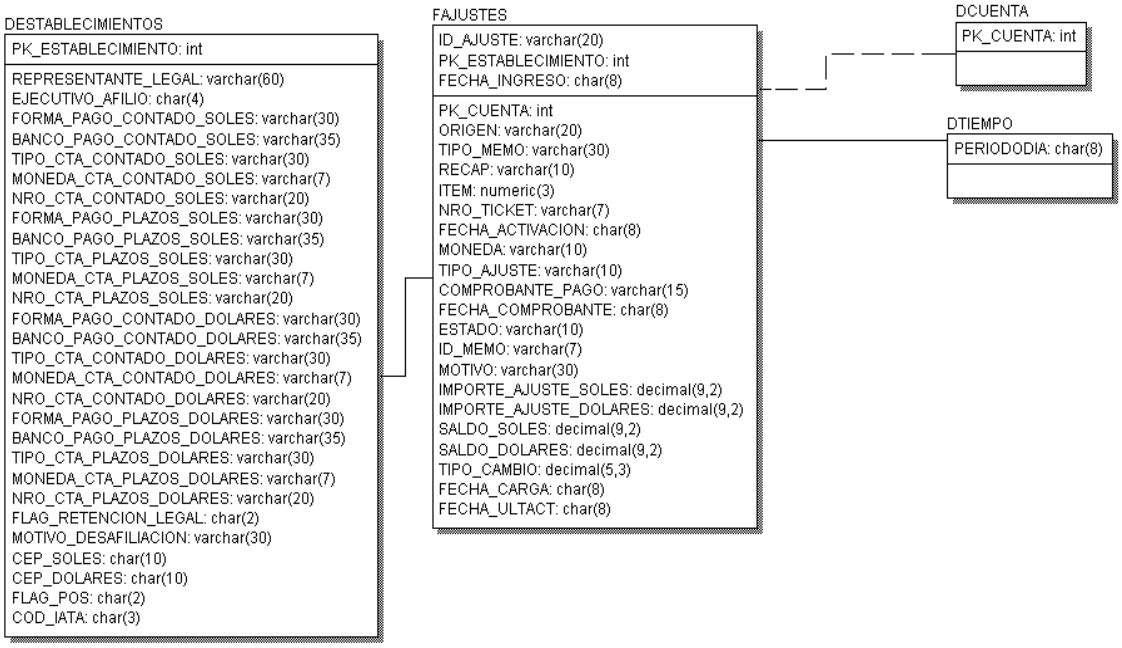
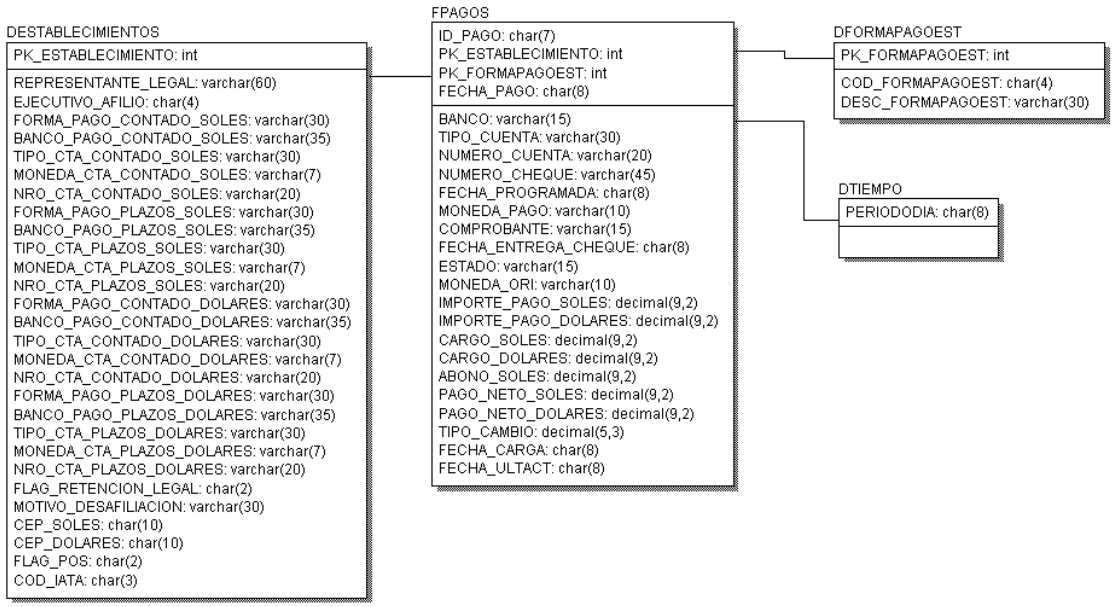


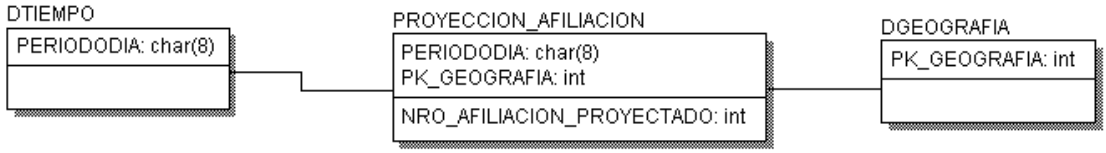
Figura Nº 4.5:"Modelo Lógico"

4.3.4 Modelo Dimensional




DESTABLECIMIENTOS
PK_ESTABLECIMIENTO: int
REPRESENTANTE_LEGAL: varchar(60)
EJECUTIVO_AFILIO: char(4)
FORMA_PAGO_CONTADO_SOLES: varchar(30)
BANCO_PAGO_CONTADO_SOLES: varchar(35)
TIPO_CTA_CONTADO_SOLES: varchar(30)
MONEDA_CTA_CONTADO_SOLES: varchar(7)
NRO_CTA_CONTADO_SOLES: varchar(20)
FORMA_PAGO_PLAZOS_SOLES: varchar(30)
BANCO_PAGO_PLAZOS_SOLES: varchar(35)
TIPO_CTA_PLAZOS_SOLES: varchar(30)
MONEDA_CTA_PLAZOS_SOLES: varchar(7)
NRO_CTA_PLAZOS_SOLES: varchar(20)
FORMA_PAGO_CONTADO_DOLARES: varchar(30)
BANCO_PAGO_CONTADO_DOLARES: varchar(35)
TIPO_CTA_CONTADO_DOLARES: varchar(30)
MONEDA_CTA_CONTADO_DOLARES: varchar(7)
NRO_CTA_CONTADO_DOLARES: varchar(20)
FORMA_PAGO_PLAZOS_DOLARES: varchar(30)
BANCO_PAGO_PLAZOS_DOLARES: varchar(35)
TIPO_CTA_PLAZOS_DOLARES: varchar(30)
MONEDA_CTA_PLAZOS_DOLARES: varchar(7)
NRO_CTA_PLAZOS_DOLARES: varchar(20)
FLAG_RETENCION_LEGAL: char(2)
MOTIVO_DESAFILIACION: varchar(30)
CEP_SOLES: char(10)
CEP_DOLARES: char(10)
FLAG_POS: char(2)
COD_IATA: char(3)
FAJUSTES
ID_AJUSTE: varchar(20)
PK_ESTABLECIMIENTO: int
FECHA_INGRESO: char(8)
PK_CUENTA: int
ORIGEN: varchar(20)
TIPO_MEMO: varchar(30)
RECAP: varchar(10)
ITEM: numeric(3)
NRO_TICKET: varchar(7)
FECHA_ACTIVACION: char(8)
MONEDA: varchar(10)
TIPO_AJUSTE: varchar(10)
COMPROBANTE_PAGO: varchar(15)
FECHA_COMPROBANTE: char(8)
ESTADO: varchar(10)
ID_MEMO: varchar(7)
MOTIVO: varchar(30)
IMPORTE_AJUSTE_SOLES: decimal(9,2)
IMPORTE_AJUSTE_DOLARES: decimal(9,2)
SALDO_SOLES: decimal(9,2)
SALDO_DOLARES: decimal(9,2)
TIPO_CAMBIO: decimal(5,3)
FECHA_CARGA: char(8)
FECHA_ULTACT: char(8)
DCUENTA
PK_CUENTA: int
DTIEMPO
PERIODODIA: char(8)


DESTABLECIMIENTOS
PK_ESTABLECIMIENTO: int
REPRESENTANTE_LEGAL: varchar(60)
EJECUTIVO_AFILIO: char(4)
FORMA_PAGO_CONTADO_SOLES: varchar(30)
BANCO_PAGO_CONTADO_SOLES: varchar(35)
TIPO_CTA_CONTADO_SOLES: varchar(30)
MONEDA_CTA_CONTADO_SOLES: varchar(7)
NRO_CTA_CONTADO_SOLES: varchar(20)
FORMA_PAGO_PLAZOS_SOLES: varchar(30)
BANCO_PAGO_PLAZOS_SOLES: varchar(35)
TIPO_CTA_PLAZOS_SOLES: varchar(30)
MONEDA_CTA_PLAZOS_SOLES: varchar(7)
NRO_CTA_PLAZOS_SOLES: varchar(20)
FORMA_PAGO_CONTADO_DOLARES: varchar(30)
BANCO_PAGO_CONTADO_DOLARES: varchar(35)
TIPO_CTA_CONTADO_DOLARES: varchar(30)
MONEDA_CTA_CONTADO_DOLARES: varchar(7)
NRO_CTA_CONTADO_DOLARES: varchar(20)
FORMA_PAGO_PLAZOS_DOLARES: varchar(30)
BANCO_PAGO_PLAZOS_DOLARES: varchar(35)
TIPO_CTA_PLAZOS_DOLARES: varchar(30)
MONEDA_CTA_PLAZOS_DOLARES: varchar(7)
NRO_CTA_PLAZOS_DOLARES: varchar(20)
FLAG_RETENCION_LEGAL: char(2)
MOTIVO_DESAFILIACION: varchar(30)
CEP_SOLES: char(10)
CEP_DOLARES: char(10)
FLAG_POS: char(2)
COD_IATA: char(3)
FPAGOS
ID_PAGO: char(7)
PK_ESTABLECIMIENTO: int
PK_FORMAPAGOEST: int
FECHA_PAGO: char(8)
BANCO: varchar(15)
TIPO_CUENTA: varchar(30)
NUMERO_CUENTA: varchar(20)
NUMERO_CHEQUE: varchar(45)
FECHA_PROGRAMADA: char(8)
MONEDA_PAGO: varchar(10)
COMPROBANTE: varchar(15)
FECHA_ENTREGA_CHEQUE: char(8)
ESTADO: varchar(15)
MONEDA_ORI: varchar(10)
IMPORTE_PAGO_SOLES: decimal(9,2)
IMPORTE_PAGO_DOLARES: decimal(9,2)
CARGO_SOLES: decimal(9,2)
CARGO_DOLARES: decimal(9,2)
ABONO_SOLES: decimal(9,2)
PAGO_NETO_SOLES: decimal(9,2)
PAGO_NETO_DOLARES: decimal(9,2)
TIPO_CAMBIO: decimal(5,3)
FECHA_CARGA: char(8)
FECHA_ULTACT: char(8)
DFORMAPAGOEST
PK_FORMAPAGOEST: int
COD_FORMAPAGOEST: char(4)
DESC_FORMAPAGOEST: varchar(30)
DTIEMPO
PERIODODIA: char(8)


DTIEMPO
PERIODODIA: char(8)
PROYECCION_AFILIACION
PERIODODIA: char(8)
PK_GEOGRAFIA: int
NRO_AFILIACION_PROYECTADO: int
DGEOGRAFIA
PK_GEOGRAFIA: int

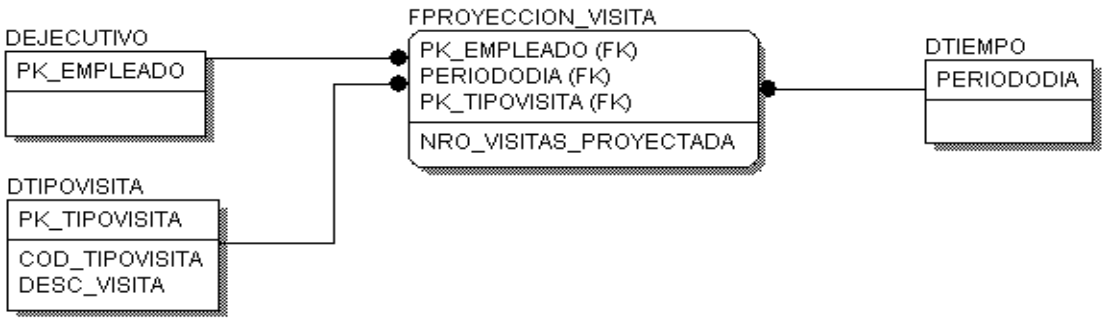


Figura Nº 4.6:"Modelo Dimensional"

Al finalizar esta etapa se entrega el modelo dimensional así como la documentación del mismo.

4.4 Diseño e Implementación de Procesos de Extracción, Transformación y Carga (ETL)

En esta fase se realiza el diseño y construcción de los procesos de extracción, transformación y carga de datos, tanto para la carga inicial como para la carga periódica de datos. Para realizar esta labor, se utilizará la herramienta ETL que se considere necesaria. Para esta etapa se consideran las siguientes acciones:

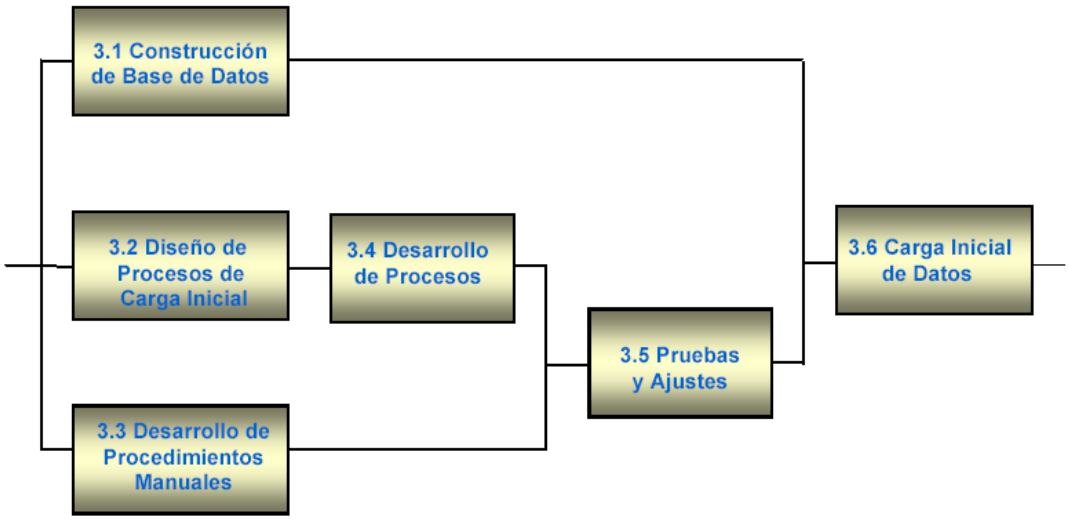


Figura Nº 4.6:"Etapas Procesos ETL I"

a. Extraer: Capturar la data de las Fuentes de Datos apropiadas

b. Transformar, Integrar y Limpiar: Consolidar la data de múltiples Fuentes de Datos dentro de una visión de área o negocio, adoptar una única representación de los datos y validar la data antes que ingrese al Data Mart tomando acciones apropiadas para asegurar su calidad.

c. Cargar: Mover físicamente la data apropiada de las Fuentes de Datos hacia el Data Mart

Objetivos

- Diseñar, programar y ejecutar los procesos de extracción, transformación y carga de datos
- Realizar la prueba integral de los procesos

Tareas

- Definición detallada de las fuentes de datos
- Definición de la lógica y las condiciones de carga
- Diseño e implementación de procesos de extracción, transformación y carga
- Prueba integral del proceso
- Desarrollo de procedimientos de contingencia
- Definir periodicidad de carga de datos

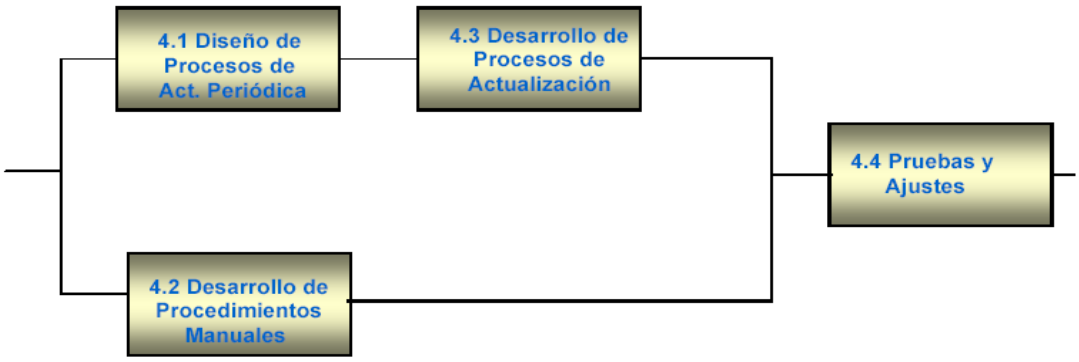


Figura Nº 4.8:"Etapas Procesos ETL II"

Los procesos de extracción y carga se efectuaron en la plataforma AS400 utilizando programas hechos en lenguaje RPGLE, SQLRPGLE y CL.

Se construye la base Datawarehouse_DC, para ello se utiliza Microsoft SQL Server 2005, esta base será la fuente de datos para realizar Datawarehousing.

Extracción y Transformación de Datos

El proceso de extracción, transformación y carga se divide en 2 partes bien marcadas, la primera desde la plataforma AS400 y la otra utilizando SQL SERVER 2005.

4.4.1 Proceso Central AS400

El proceso tiene por finalidad la carga diaria de las tablas temporales en el sistema AS400 alimentándose básicamente de los sistemas transaccionales, para ello el sistema ejecuta diferentes procesos desarrollados utilizando programas CL, RPG RPGLE y SQLRPGLE. Esta etapa forma parte de las fases de extracción y transformación de datos de la solución BI.

4.4.1.1 ETL Dimensión ESTABLECIMIENTOS

El proceso tiene por finalidad cargar la tabla temporal Establecimientos (F76407).

Programa: IR761003

Tablas: PAR06, F41101, F41513, F41514, F41506, F41508, F41264, F41502, L4120601, L4121402, F71501, L4121805, F41267, F71506, F41144, F41292, L4120702, F20531, F41217, F76301, F76511.

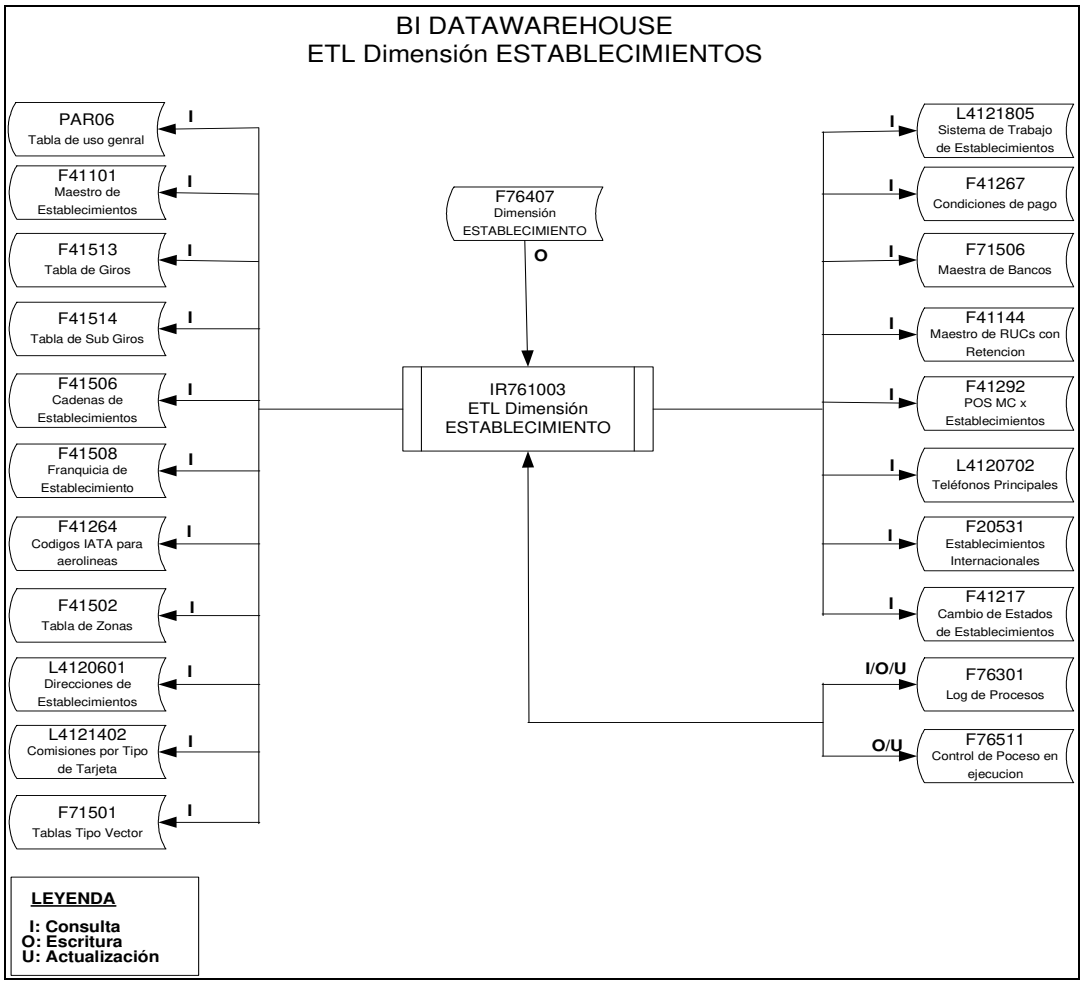


Figura Nº 4.8:"AS400 ETL Dimensión ESTABLECIMIENTOS"

4.4.1.2 ETL Dimensiones Varias Establecimiento

El proceso tiene por finalidad cargar las tablas temporales Artículo (F76425), Visita (F76426), FormaPago Estab (F76427) y Representante (F76428).

Programa: IR761017

Tablas: PAR06, TARTI, F71501, F41102, F76301, F76511.

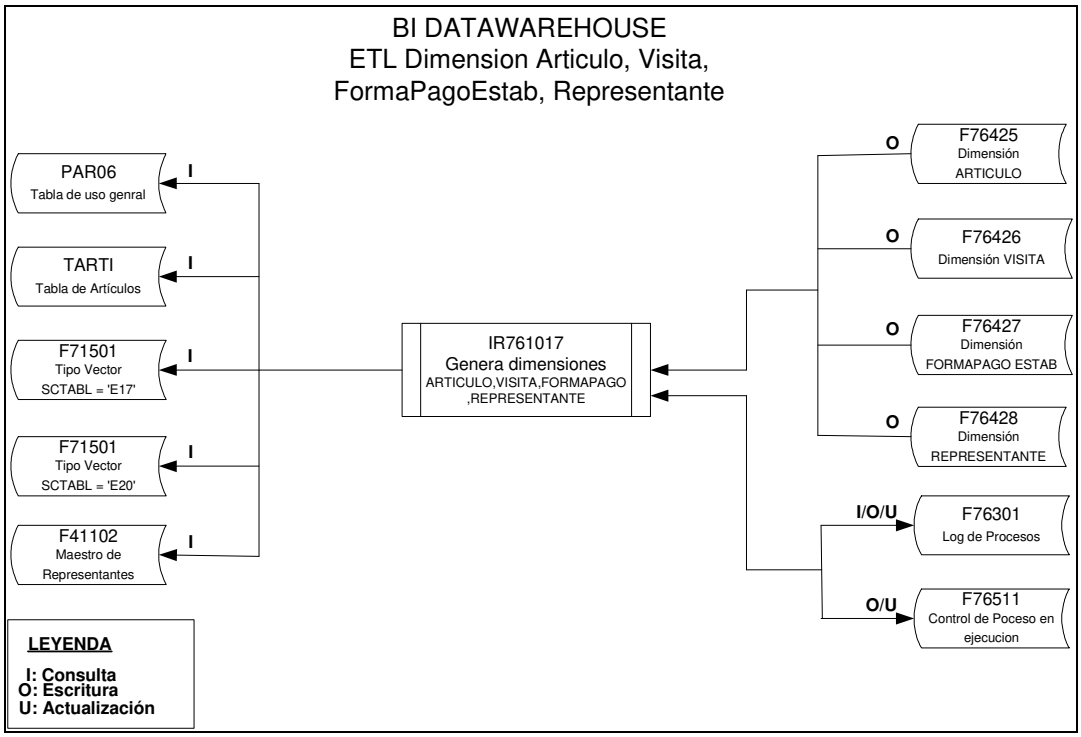


Figura Nº 4.10:"AS400 ETL Dimensiones VARIAS"

4.4.1.3 ETL Fact Table VISITAS Y MATERIALES

El proceso tiene por finalidad cargar las tablas temporales Visitas (F76429) y Materiales (F76430).

Programa: IR761018

Tablas: PAR06, F41105, F41331, TARTI, F76301, F76511.

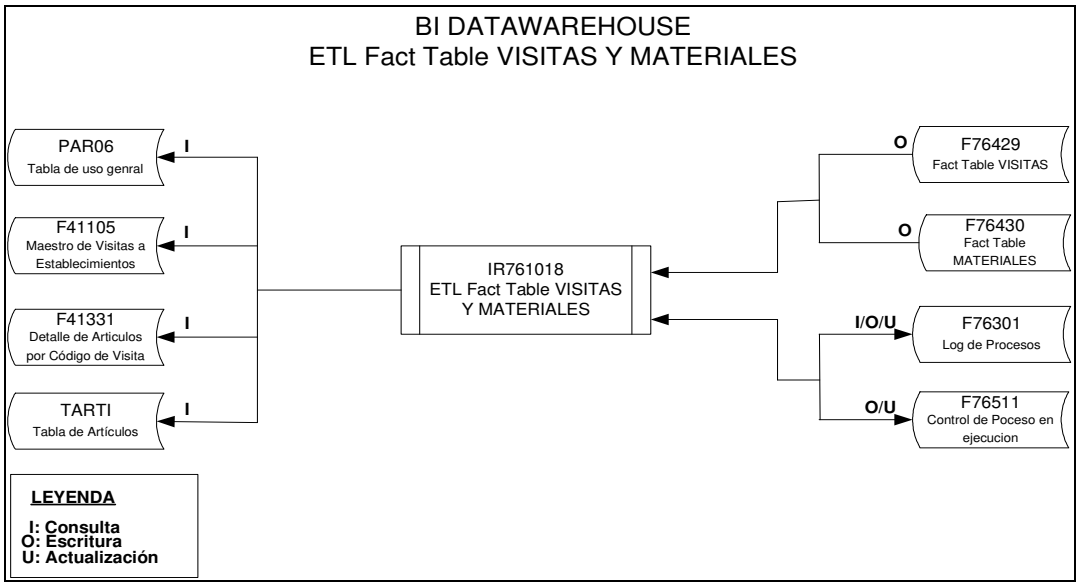


Figura Nº 4.11:"AS400 Fact Tables VISITAS Y ENTREGA DE MATERIALES"

4.4.1.4 ETL Fact Table AJUSTES

El proceso tiene por finalidad cargar la tabla temporal de Ajustes (F76431)

Programa: IR761019

Tablas: PAR06, F41130, L4112903, L4112904, F71501, CS01L32, F76301, F76511.

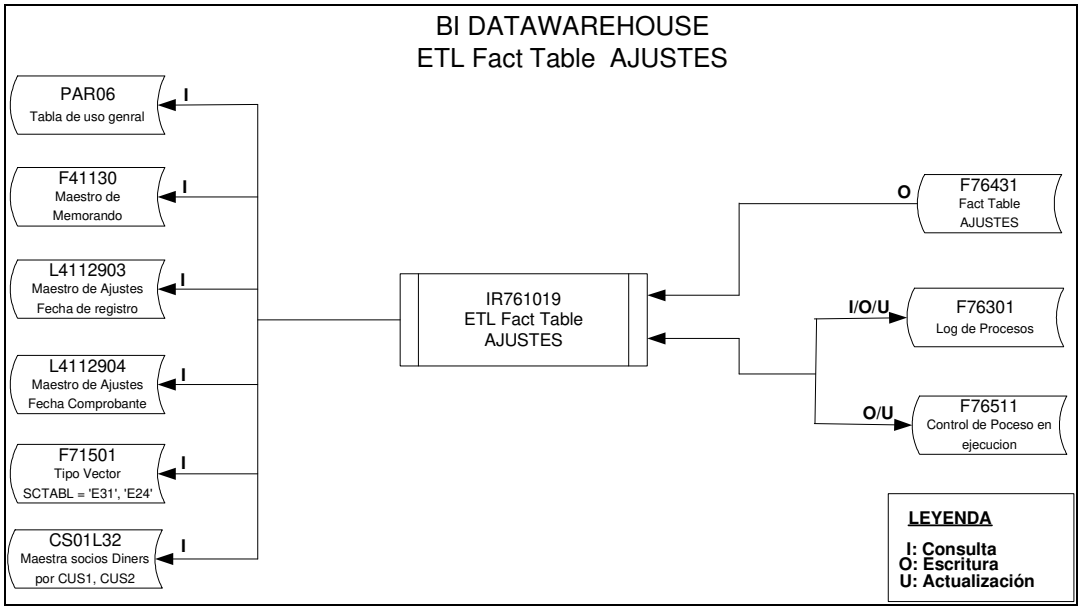


Figura Nº 4.12:"AS400 Fact Tables AJUSTES"

4.4.1.5 ETL Fact Table PAGOS

El proceso tiene por finalidad cargar la tabla temporal Pagos (F76432).

Programa: IR761020

Tablas: PAR06, L4125006, F71506, F71501, F41131, F41249, L4124501, F41129, F76310, F76511.

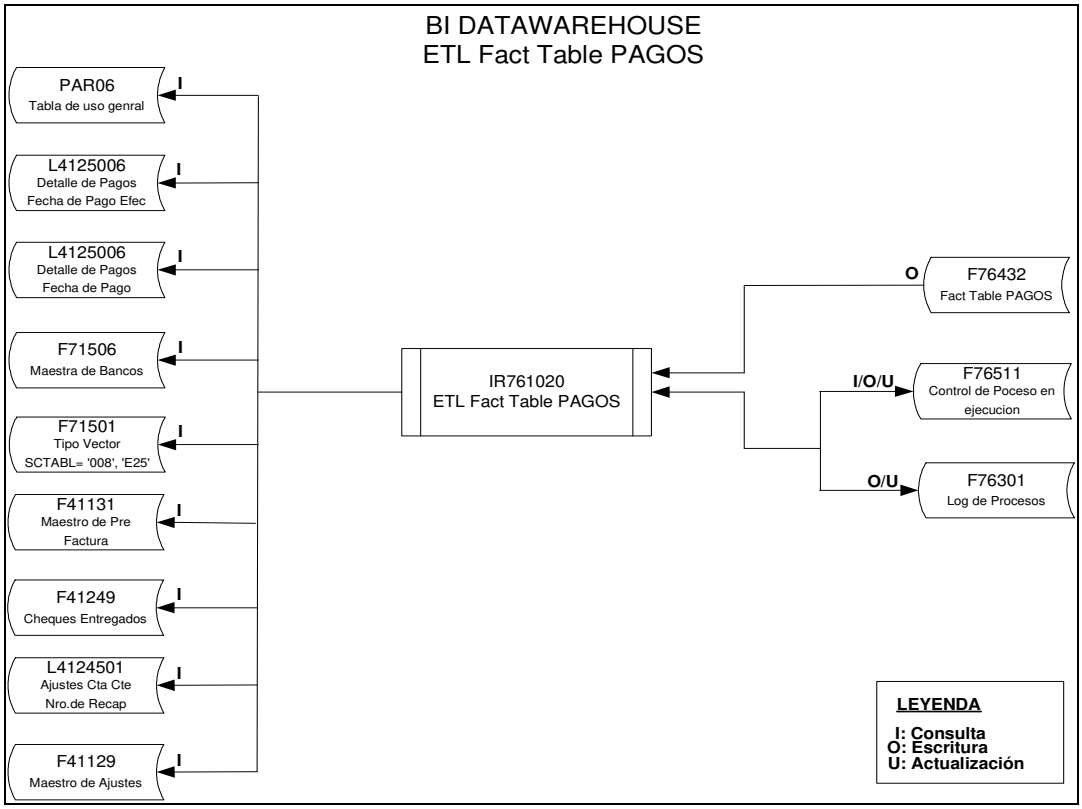


Figura Nº 4.13:"AS400 Fact Tables PAGOS"

4.4.2 Proceso Central SQL

El proceso tiene por finalidad la carga de información al Data Warehouse. Esta etapa forma parte de las fases de transformación y carga final de la solución.

4.4.2.1 Dimensiones Generales Dimensión Articulo

El proceso tiene por finalidad cargar la dimensión Artículo.

Paquete: "*DMC-Dimensiones Generales*"

Secuencia: "Dimensión Artículo"

Pasos:

1.- "Borrar Temporal" Tablas: F76425 (SQL)

2.- "Transfiere Dimensión Artículo" Tablas F76425 (SQL), F76425 (DACLB)

3.- "Actualiza Dimensión Artículo" Procedure: sp_DME_Actualiza_DArticulo

Tablas: F76425 (SQL), DARTICULO

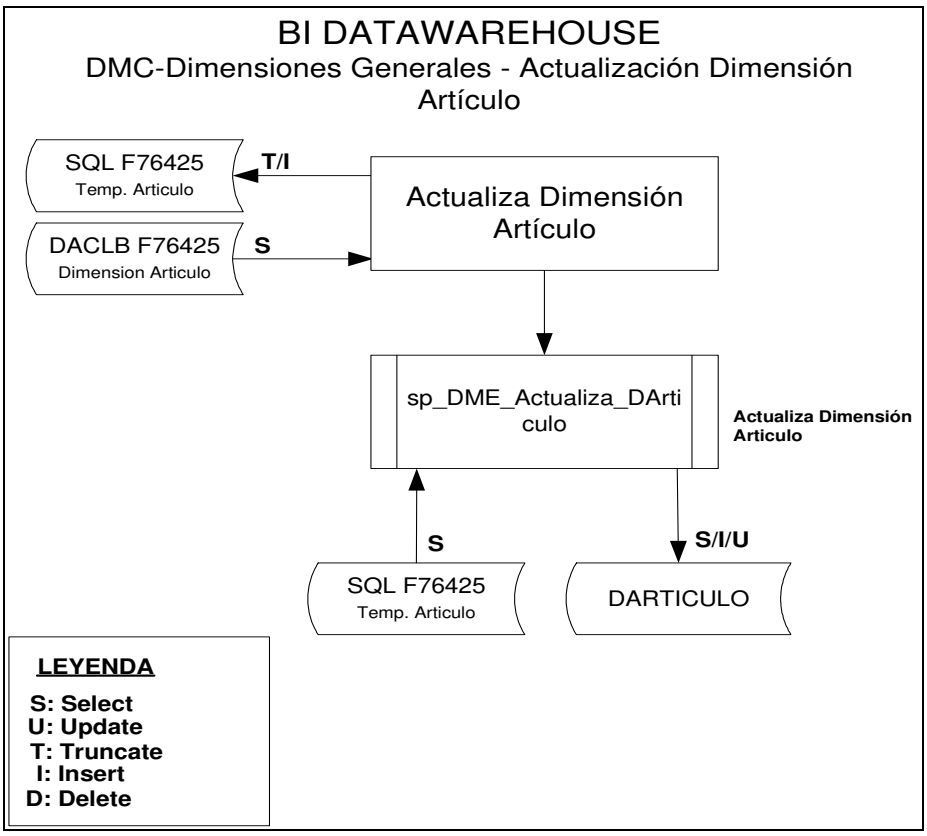


Figura Nº 4.14:"SQL Dimensión ARTÍCULO"

4.4.2.2 Dimensiones Generales Dimensión Representante

El proceso tiene por finalidad cargar la dimensión Representante.

Paquete: "*DMC-Dimensiones Generales*"

Secuencia: "Dimensión Representante"

Pasos:

1.- "Borrar Temporal" Tablas: F76428 (SQL)

2.- "Transfiere Dimensión Representante" Tablas F76428 (SQL), F76428 (DACLB)

3.- "Actualiza Dimensión Representante" Procedure: sp_DME_Actualiza_DRepresentante

Tablas: F76428 (SQL), DREPRESENTANTE

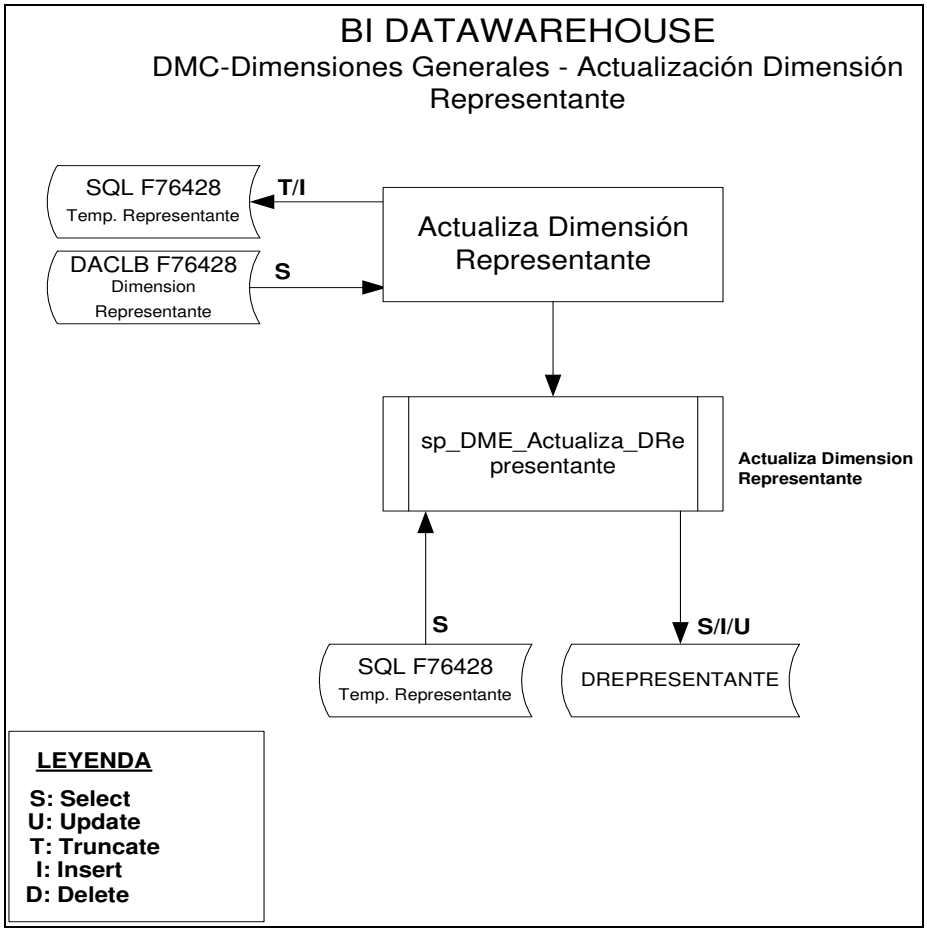


Figura Nº 4.15:"SQL Dimensión REPRESENTANTE"

4.4.2.3 Dimensiones Generales Dimensión Forma Pago Establecimiento

El proceso tiene por finalidad cargar la dimensión Forma Pago Establecimiento.

Paquete: "*DMC-Dimensiones Generales*"

Secuencia: "Dimensión Forma Pago Establecimiento"

Pasos:

1.- "Borrar Temporal" Tablas: F76427 (SQL)

2.- "Transfiere Dimensión Forma Pago" Tablas F76427 (SQL), F76427 (DACLB)

3.- "Actualiza Dimensión Forma Pago Establecimiento"

Procedure: sp_DME_Actualiza_DFormaPagoEst

Tablas: F76427 (SQL), DFORMAPAGOEST

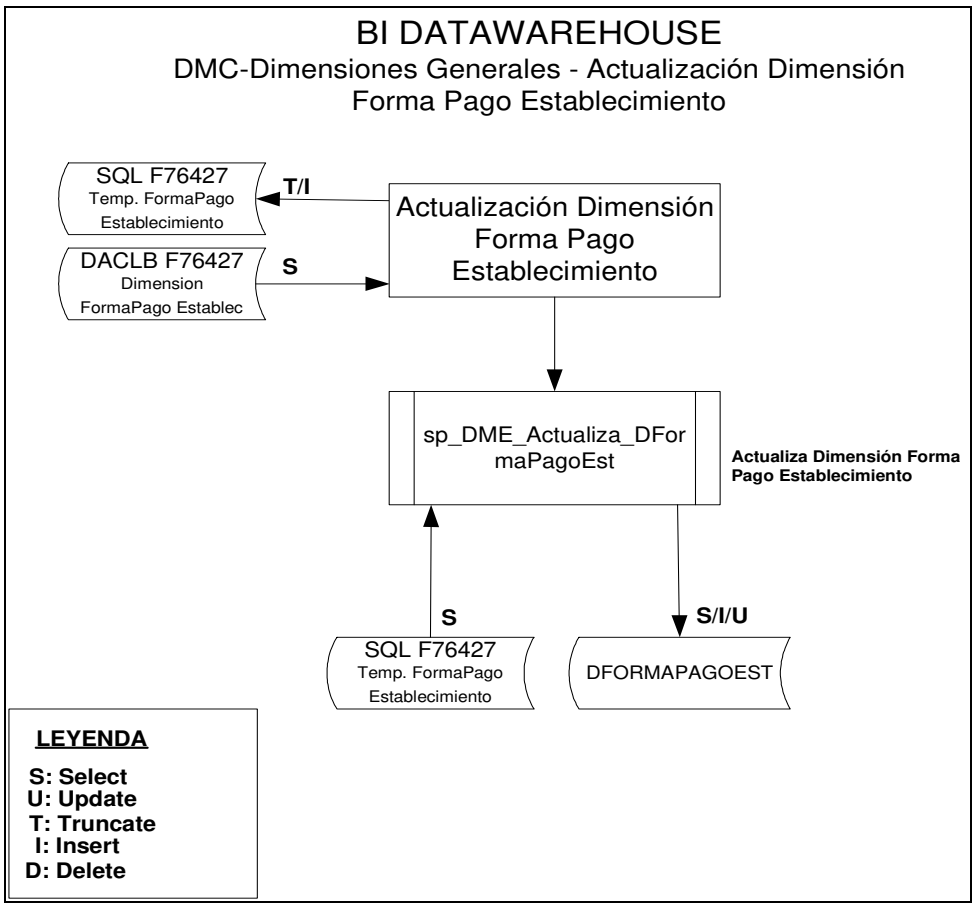


Figura Nº 4.16:"SQL Dimensión FORMA PAGO ESTABLECIMIENTO"

4.4.2.4 Dimensiones Generales Dimensión Tipo Visita

El proceso tiene por finalidad cargar la dimensión Tipo Visita.

Paquete: "*DMC-Dimensiones Generales*"

Secuencia: "Dimensión Tipo Visita"

Pasos:

1.- "Borrar Temporal" Tablas: F76426 (SQL)

2.- "Transfiere Dimensión Tipo Visita" Tablas F76426 (SQL), F76426 (DACLB)

3.- "Actualiza Dimensión Tipo Visita" Procedure: sp_DME_Actualiza_DTipoVisita

Tablas: F76426 (SQL), DTIPOVISITA

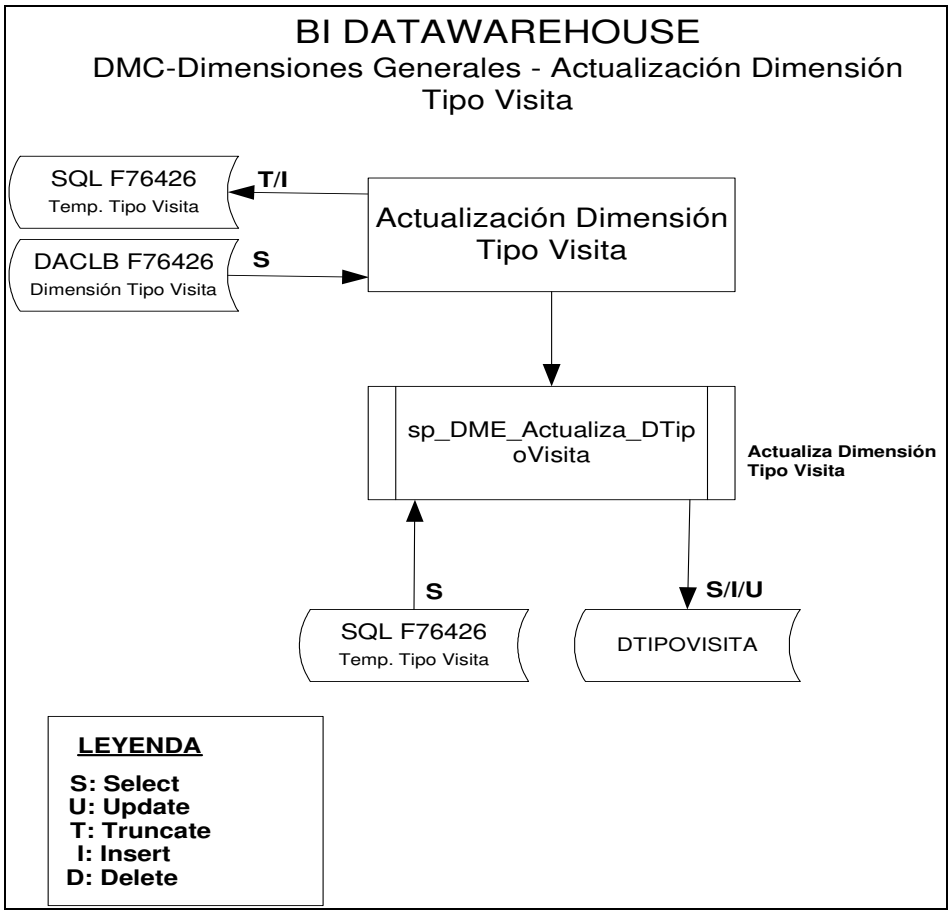


Figura Nº 4.17:"SQL Dimensión TIPO VISITA"

4.4.2.5 Dimensión Establecimiento

El proceso tiene por finalidad cargar la dimensión Establecimiento y Afiliación CAD.

Paquete: "*DMC-Dimensión Establecimiento*"

Secuencia: "Actualización Dimensión Establecimientos y Afiliados a CAD"

Pasos:

1.- "Tarea Ejecutar SQL" Tablas: F76407 (SQL), F76409 (SQL)

2.- "Transfiere tabla temporal Establecimientos" Tablas F76407 (SQL), F76409 (SQL), F76407 (DACLB), F76409 (DACLB)

3.- "Actualiza Dimensión Establecimiento" Procedure: sp_DMC_Actualiza_DEstablecimiento Tablas: F76407 (SQL), DGEOGRAFIA, DEMPLEADO, DESTABLECIMIENTO

4.- "Actualiza Dimensión Afiliaciones CAD" Procedure: sp_DMC_Actualiza_DAfiliadosCAD Tablas: F76409 (SQL), DESTABLECIMIENTO, DAFIL_CAD

5.- "Actualiza Histórico de Establecimiento" Procedure: sp_DME_Actualiza_DEstablecimientoHist Tablas: F76511 (SQL), DESTABLECIMIENTO, DESTABLECIMIENTOHIST

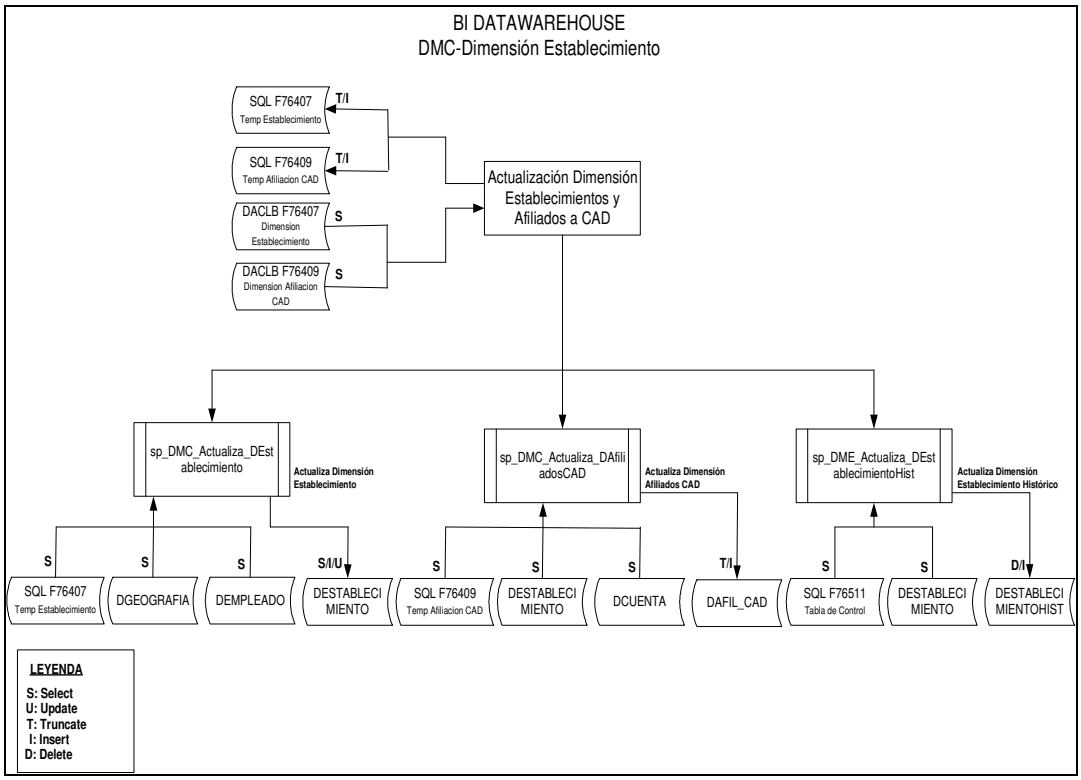


Figura Nº 4.18:"SQL Dimensión ESTABLECIMIENTO"

4.4.2.6 Fact Table Ajustes

El proceso tiene por finalidad cargar la Fact Table Ajustes.

Paquete: "*DME-FactTable Ajustes*"

Secuencia: "Actualiza Fact table Ajustes"

Pasos:

1.- "Borra tabla Temporal" Tablas: F76431 (SQL)

2.- "Transfiere temporal Fact Table Ajustes" Tablas F76431 (SQL), F76431 (DACLB)

3.- "Cambio de código tabla temporal"

Procedure: sp_DME_Actualiza_Cambio_Código_FAjustes

Tablas: F76510 (SQL), F76431 (SQL)

4.- "Actualiza Fact Table Ajustes"

Procedure: sp_DME_Actualiza_FactTable_Ajustes

Tablas: F76431 (SQL), DCUENTA, DESTABLECIMIENTO, FAJUSTES

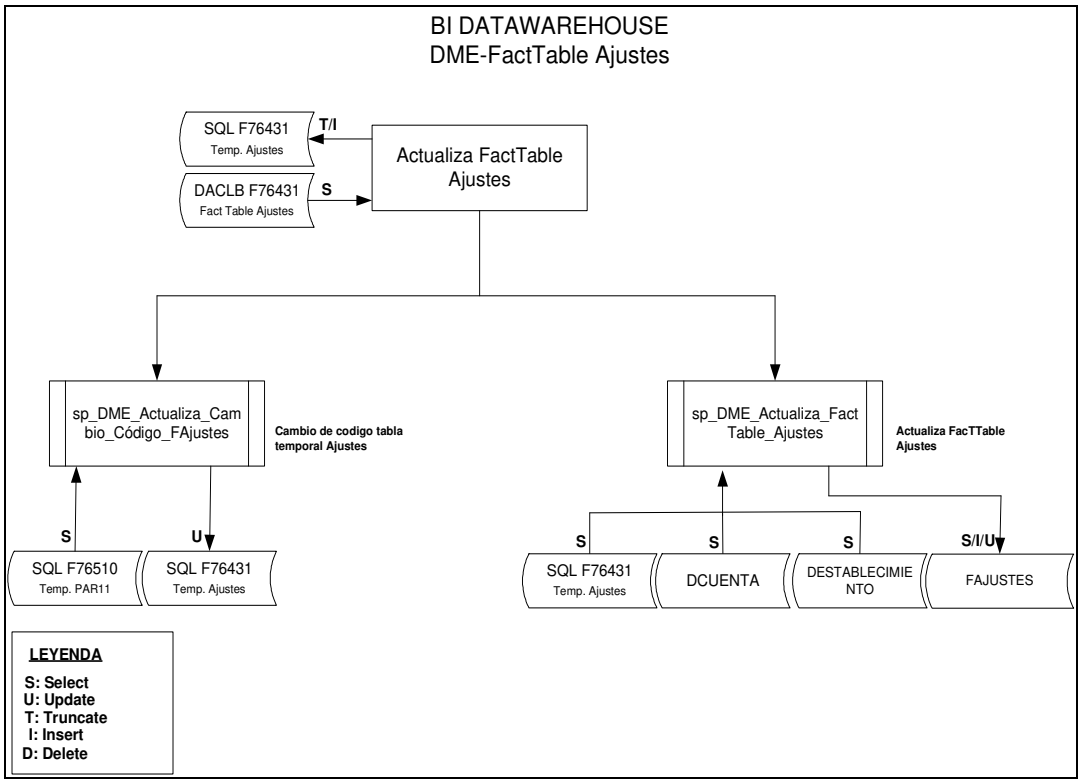


Figura Nº 4.19:"SQL Fact Table AJUSTES"

4.4.2.7 Fact Table Pagos

El proceso tiene por finalidad cargar la Fact Table Pagos.

Paquete: "*DME-FactTable Pagos*"

Secuencia: "Actualiza Fact Table Pagos"

Pasos:

1.- "Borra Temporal" Tablas: F76432 (SQL)

2.- "Transferencia Pagos a Establecimientos" Tablas F76432 (SQL), F76432 (DACLB)

3.- "Actualiza Fact Table Pagos"

Procedure: sp_DME_Actualiza_FactTable_Pagos

Tablas: F76432 (SQL), DFORMAPAGOEST, DESTABLECIMIENTO, FPAGOS

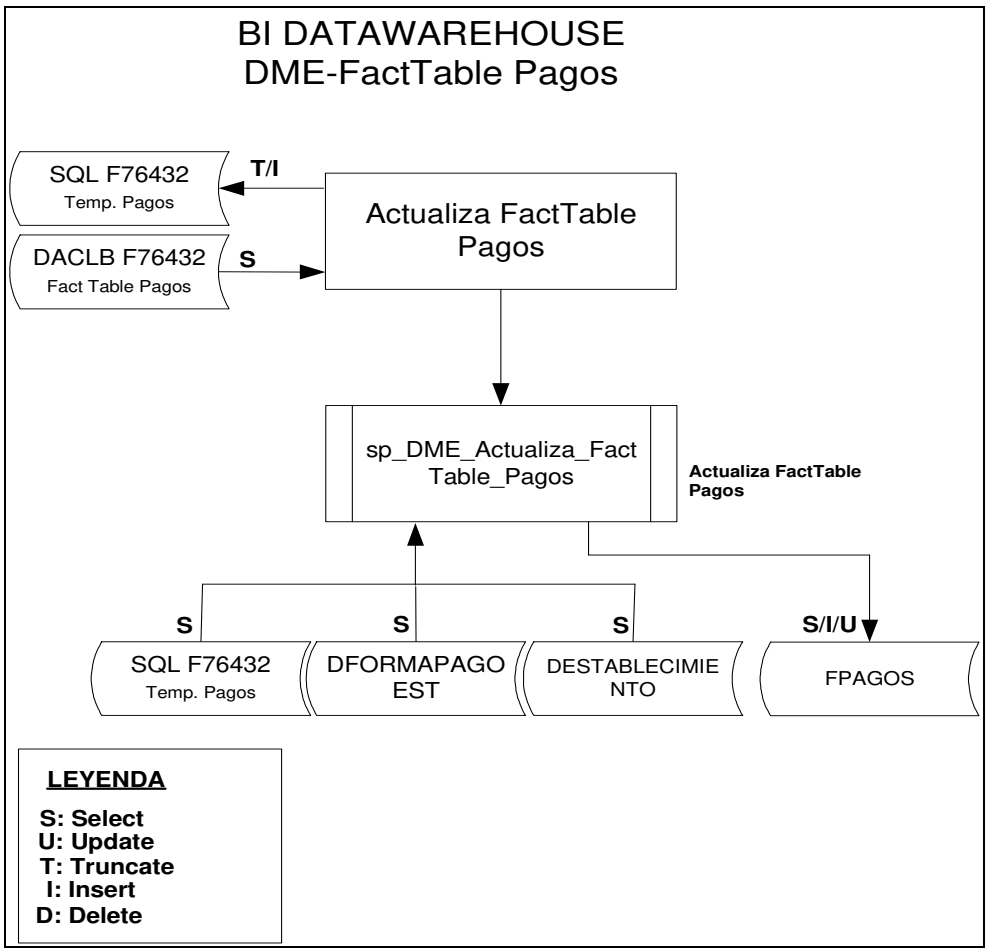


Figura Nº 4.20:"SQL Fact Table PAGOS"

4.4.2.8 Fact Table Visitas

El proceso tiene por finalidad cargar la Fact Tables de Visitas y Entrega de Materiales.

Paquete: "*DME-FactTable Visitas*"

Secuencia: "Actualiza FactTable Visitas"

Pasos:

1.- "Borra Temporal" Tablas: F76429 (SQL), F76430 (SQL)

2.- "Transferencia Visitas a Establecimientos" Tablas F76429 (SQL), F76430 (SQL), F76429 (DACLB), F76430 (DACLB)

3.- "Actualiza FactTable Visitas"

Procedure: sp_DME_Actualiza_FactTable_Visitas Tablas: F76429 (SQL) DESTABLECIMIENTO, DREPRESENTANTE, DEMPLEADO, DTIPOVISITA, FVISITAS

4.- "Actualiza FactTable Entrega de Materiales"

Procedure: sp_DME_Actualiza_FactTable_Entrega_Materiales

Tablas: F76430 (SQL), DARTICULO, FENTREGA_MATERIALES

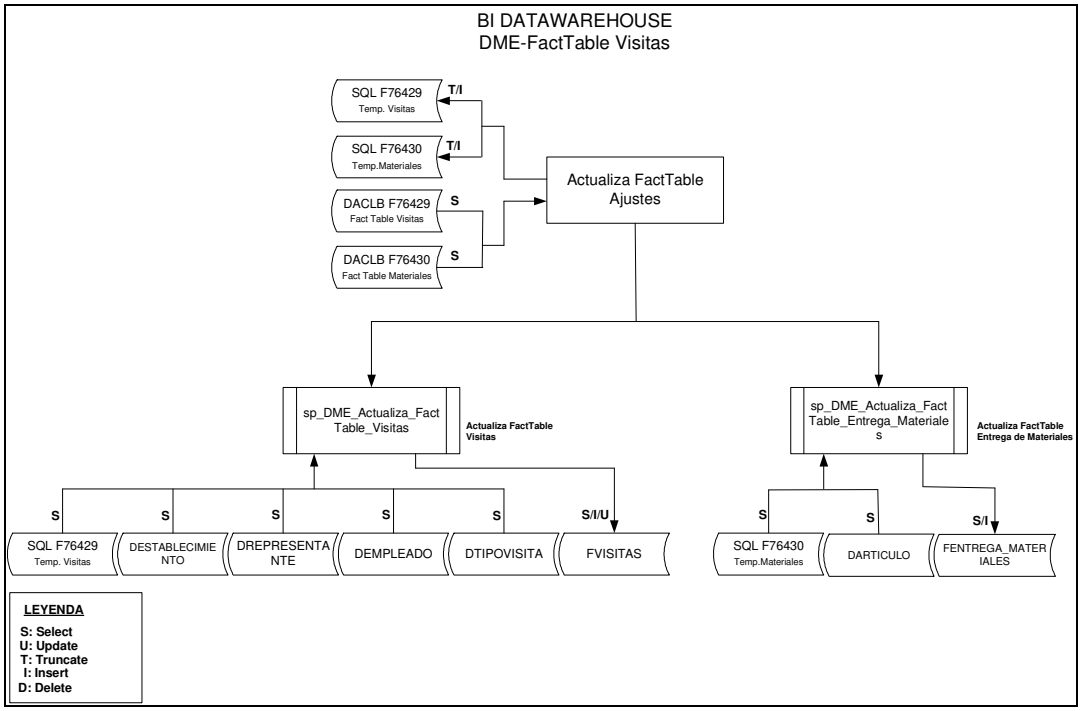


Figura Nº 4.21:"SQL Fact Table VISITAS"

4.4.2.9 Fact Table Proyección de Afiliaciones

El proceso tiene por finalidad cargar la Fact Table de Proyección de Afiliaciones de Establecimientos.

Paquete: "*DME-FactTable Proyeccion de Afiliaciones*"

Secuencia: "Proyeccion de Afiliaciones"

Pasos:

1.- "Borrar Temporal" Tablas: F76434 (SQL)

2.- "Transfiere Metas de Afiliación" Tablas F76434 (SQL), F76434 (DACLB)

3.- "Actualiza Proyección de Afiliación"

Procedure: sp_DME_Actualiza_FactTable_FProyecciones_Afiliacion

Tablas: F76434 (SQL), DGEOGRAFIA, FPROYECCION_AFILIACION

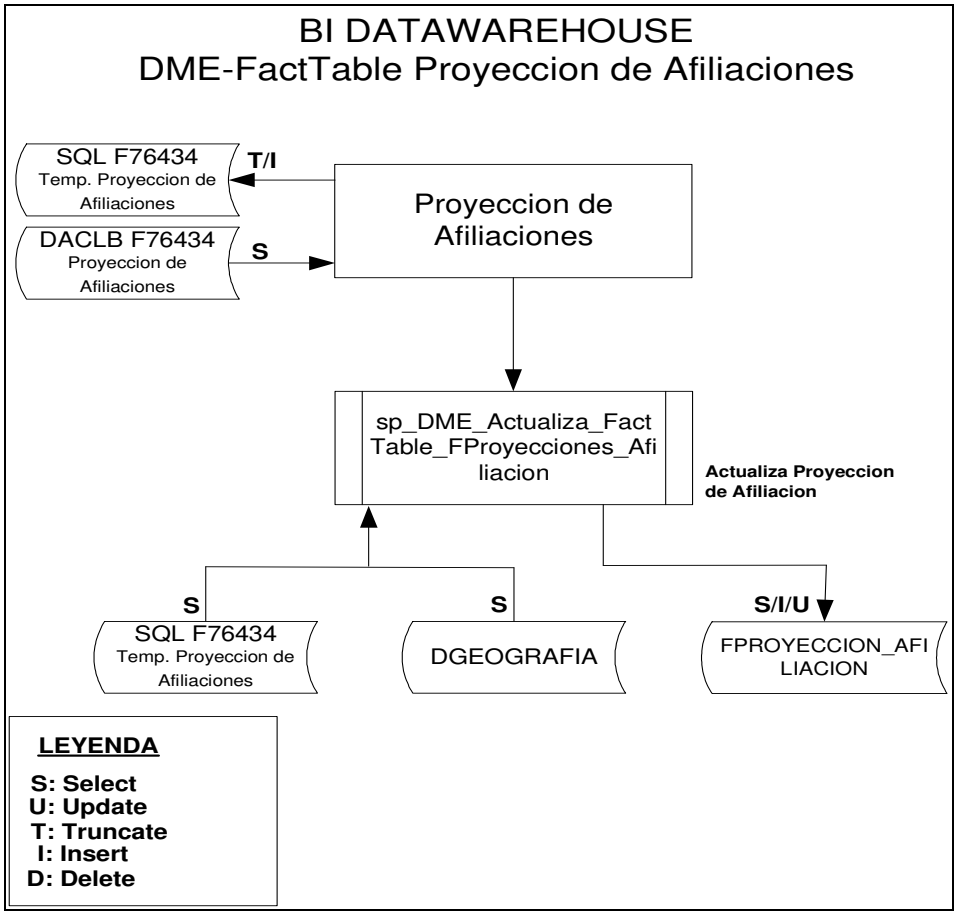


Figura Nº 4.22:"SQL Fact Table PROYECCIÓN DE AFILIACIONES"

4.4.2.10 Fact Table Proyección de Visitas

El proceso tiene por finalidad cargar la Fact Table de Proyección de Visitas.

Paquete: "*DME-FactTable Proyeccion de Visitas*"

Secuencia: "Proyección de Visitas"

Pasos:

1.- "Borrar Temporal" Tablas: F76433 (SQL)

2.- "Transfiere Metas de Visita" Tablas F76433 (SQL), F76433 (DACLB)

3.- "Actualiza Proyección de Visitas"

Procedure: sp_DME_Actualiza_FactTable_FProyecciones_Visita

Tablas: F76433 (SQL), DEMPLEADO, DTIPOVISITA, FPROYECCION_VISITA

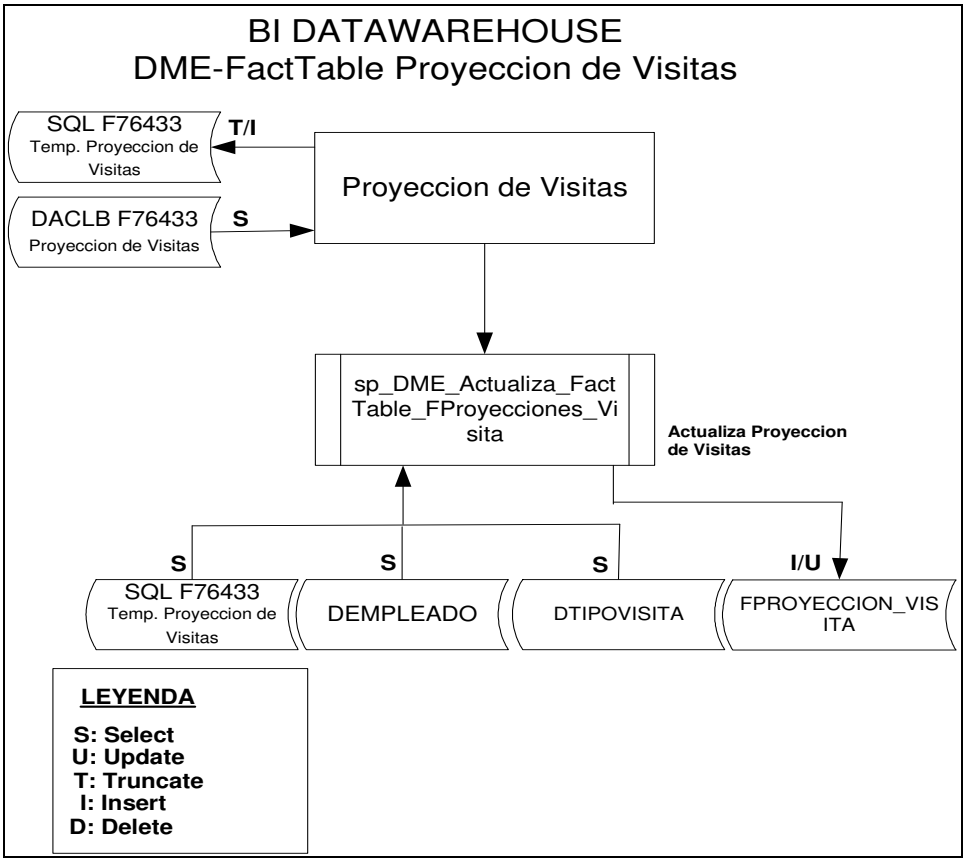


Figura Nº 4.23:"SQL Fact Table PROYECCIÓN DE VISITAS"

4.4.3 Creación de Tablas

```
ALTER TABLE DESTABLECIMIENTO ADD
    REPRESENTANTE_LEGAL  varchar(60) NULL,
    EJECUTIVO_AFILIO     char(4) NULL,
    FORMA_PAGO_CONTADO_SOLES varchar(30) NULL,
    BANCO_PAGO_CONTADO_SOLES varchar(35) NULL,
    TIPO_CTA_CONTADO_SOLES varchar(30) NULL,
    MONEDA_CTA_CONTADO_SOLES varchar(7) NULL,
    NRO_CTA_CONTADO_SOLES varchar(20) NULL,
    FORMA_PAGO_PLAZOS_SOLES varchar(30) NULL,
    BANCO_PAGO_PLAZOS_SOLES varchar(35) NULL,
    TIPO_CTA_PLAZOS_SOLES varchar(30) NULL,
    MONEDA_CTA_PLAZOS_SOLES varchar(7) NULL,
    NRO_CTA_PLAZOS_SOLES varchar(20) NULL,
    FORMA_PAGO_CONTADO_DOLARES varchar(30) NULL,
    BANCO_PAGO_CONTADO_DOLARES varchar(35) NULL,
    TIPO_CTA_CONTADO_DOLARES varchar(30) NULL,
    MONEDA_CTA_CONTADO_DOLARES varchar(7) NULL,
    NRO_CTA_CONTADO_DOLARES varchar(20) NULL,
    FORMA_PAGO_PLAZOS_DOLARES varchar(30) NULL,
    BANCO_PAGO_PLAZOS_DOLARES varchar(35) NULL,
    TIPO_CTA_PLAZOS_DOLARES varchar(30) NULL,
    MONEDA_CTA_PLAZOS_DOLARES varchar(7) NULL,
    NRO_CTA_PLAZOS_DOLARES varchar(20) NULL,
    FLAG_RETENCION_LEGAL char(2) NULL,
    MOTIVO_DESAFILIACION varchar(30) NULL,
    CEP_SOLES          char(10) NULL,
    CEP_DOLARES        char(10) NULL,
```

```
    FLAG_POS            char(2) NULL,
    COD_IATA            char(3) NULL

CREATE TABLE FAJUSTES (
    ID_AJUSTE           varchar(20) NOT NULL,
    PK_ESTABLECIMIENTO   int NOT NULL,
    FECHA_INGRESO        char(8) NOT NULL,
    PK_CUENTA            int NOT NULL,
    ORIGEN               varchar(20) NULL,
    TIPO_MEMO            varchar(30) NULL,
    RECAP                varchar(10) NULL,
    ITEM                 numeric(3) NULL,
    NRO_TICKET           varchar(7) NULL,
    FECHA_ACTIVACION     char(8) NULL,
    MONEDA               varchar(10) NULL,
    TIPO_AJUSTE          varchar(10) NULL,
    COMPROBANTE_PAGO     varchar(15) NULL,
    ESTADO               varchar(10) NULL,
    ID_MEMO              varchar(7) NULL,
    MOTIVO               varchar(30) NULL,
    IMPORTE_AJUSTE_SOLES decimal(9,2) NULL,
    IMPORTE_AJUSTE_DOLARES decimal(9,2) NULL,
    SALDO_SOLES          decimal(9,2) NULL,
    SALDO_DOLARES        decimal(9,2) NULL,
    TIPO_CAMBIO          decimal(5,3) NULL,
    FECHA_CARGA          char(8) NULL,
    FECHA_ULTACT         char(8) NULL
)
```

```
ALTER TABLE FAJUSTES
    ADD PRIMARY KEY (ID_AJUSTE, PK_ESTABLECIMIENTO, FECHA_INGRESO)

CREATE TABLE DFORMAPAGOEST (
    PK_FORMAPAGOEST      int NOT NULL,
    COD_FORMAPAGOEST     char(4) NULL,
    DESC_FORMAPAGOEST    varchar(30) NULL
)
ALTER TABLE DFORMAPAGOEST
    ADD PRIMARY KEY (PK_FORMAPAGOEST)

CREATE TABLE FPAGOS (
    ID_PAGO              char(7) NOT NULL,
    PK_ESTABLECIMIENTO   int NOT NULL,
    PK_FORMAPAGOEST      int NOT NULL,
    FECHA_PAGO           char(8) NOT NULL,
    BANCO                varchar(15) NULL,
    TIPO_CUENTA          varchar(30) NULL,
    NUMERO_CUENTA_CHEQUE varchar(45) NULL,
    FECHA_PROGRAMADA     char(8) NULL,
    MONEDA_PAGO          varchar(10) NULL,
    COMPROBANTE          varchar(15) NULL,
    FECHA_ENTREGA_CHEQUE char(8) NULL,
    ESTADO               varchar(15) NULL,
    IMPORTE_PAGO_SOLES   decimal(9,2) NULL,
    IMPORTE_PAGO_DOLARES decimal(9,2) NULL,
    CARGO_SOLES          decimal(9,2) NULL,
```

```
    CARGO_DOLARES        decimal(9,2) NULL,
    ABONO_SOLES          decimal(9,2) NULL,
    PAGO_NETO_SOLES      decimal(9,2) NULL,
    PAGO_NETO_DOLARES    decimal(9,2) NULL,
    TIPO_CAMBIO          decimal(5,3) NULL,
    FECHA_CARGA          char(8) NULL,
    FECHA_ULTACT         char(8) NULL
)


ALTER TABLE FPAGOS
    ADD PRIMARY KEY (ID_PAGO, PK_ESTABLECIMIENTO, PK_FORMAPAGOEST,
        FECHA_PAGO)

CREATE TABLE DARTICULO (
    PK_ARTICULO          int NOT NULL,
    COD_ARTICULO         varchar(20) NULL,
    NOMBRE_ARTICULO      varchar(70) NULL,
    UNIDAD_MEDIDA        char(3) NULL,
    COSTO_SOLES          decimal(11,3) NULL,
    COSTO_DOLARES        decimal(11,3) NULL
)

ALTER TABLE DARTICULO
    ADD PRIMARY KEY (PK_ARTICULO)

CREATE TABLE DREPRESENTANTE (
    PK_REPRESENTANTE     int NOT NULL,
```

```
    COD_REPRESENTANTE    char(4) NULL,
    NOMBRE_REPRESENTANTE varchar(45) NULL
)

ALTER TABLE DREPRESENTANTE
    ADD PRIMARY KEY (PK_REPRESENTANTE)


CREATE TABLE DTIPOVISITA (
    PK_TIPOVISITA        int NOT NULL,
    COD_TIPOVISITA       varchar(4) NULL,
    DESC_VISITA          varchar(30) NULL
)


ALTER TABLE DTIPOVISITA
    ADD PRIMARY KEY (PK_TIPOVISITA)

CREATE TABLE FENTREGA_MATERIALES (
    NUMERO_VISITA        char(7) NOT NULL,
    PK_ARTICULO          int NOT NULL,
    PK_ESTABLECIMIENTO   int NOT NULL,
    PK_EMPLEADO          int NOT NULL,
    FECHA_VISITA         char(8) NOT NULL,
    PK_TIPOVISITA        int NOT NULL,
    DESCARGA             varchar(10) NULL,
    CANTIDAD_ENTREGADA   int NULL,
    COSTO_ARTICULO_SOLES decimal(11,3) NULL,
```

```
    COSTO_ARTICULO_DOLARES decimal(11,3) NULL,
    TIPO_CAMBIO         decimal(5,3) NULL
)

ALTER TABLE FENTREGA_MATERIALES
    ADD PRIMARY KEY (NUMERO_VISITA, PK_ARTICULO)



CREATE TABLE FVISITAS (
    NUMERO_VISITA        char(7) NOT NULL,
    PK_ESTABLECIMIENTO   int not NULL,
    PK_EMPLEADO          int not NULL,
    FECHA_VISITA         char(8) NOT NULL,
    PK_TIPOVISITA        int NOT NULL,
    PK_REPRESENTANTE     int NOT NULL,
    CONTACTO             varchar(30) NULL,
    ACCION_REALIZADA     varchar(250) NULL,
    SUGERENCIAS          varchar(250) NULL,
    ESTADO               varchar(10) NULL,
    ESTADO_ATENDIDO      varchar(10) NULL,
    FECHA_GENERACION     char(8) NULL,
    FECHA_DIGITACION     char(8) NULL,
    MOTIVO_AFILIACION_PENDIENTE varchar(30) NULL,
    PRIMERA_VISITA       char(2) NULL,
    EJECUTIVO_SOLICITO   int NULL
)

ALTER TABLE FVISITAS
```

```
    ADD PRIMARY KEY (NUMERO_VISITA, PK_ESTABLECIMIENTO,
        PK_EMPLEADO, FECHA_VISITA, PK_TIPOVISITA)

CREATE TABLE FPROYECCION_VISITA (
    PK_EMPLEADO         int NOT NULL,
    PERIODODIA          char(8) NOT NULL,
    PK_TIPOVISITA       int NOT NULL,
    NRO_VISITAS_PROYECTADA int NULL
)

ALTER TABLE FPROYECCION_VISITA
    ADD PRIMARY KEY (PK_EMPLEADO, PERIODODIA, PK_TIPOVISITA)


CREATE TABLE PROYECCION_AFILIACION (
    PK_EMPLEADO         int NOT NULL,
    PERIODODIA          char(8) NOT NULL,
    NRO_AFILIACION_PROYECTADO int NULL
)

ALTER TABLE PROYECCION_AFILIACION
    ADD PRIMARY KEY (PK_EMPLEADO, PERIODODIA)
```

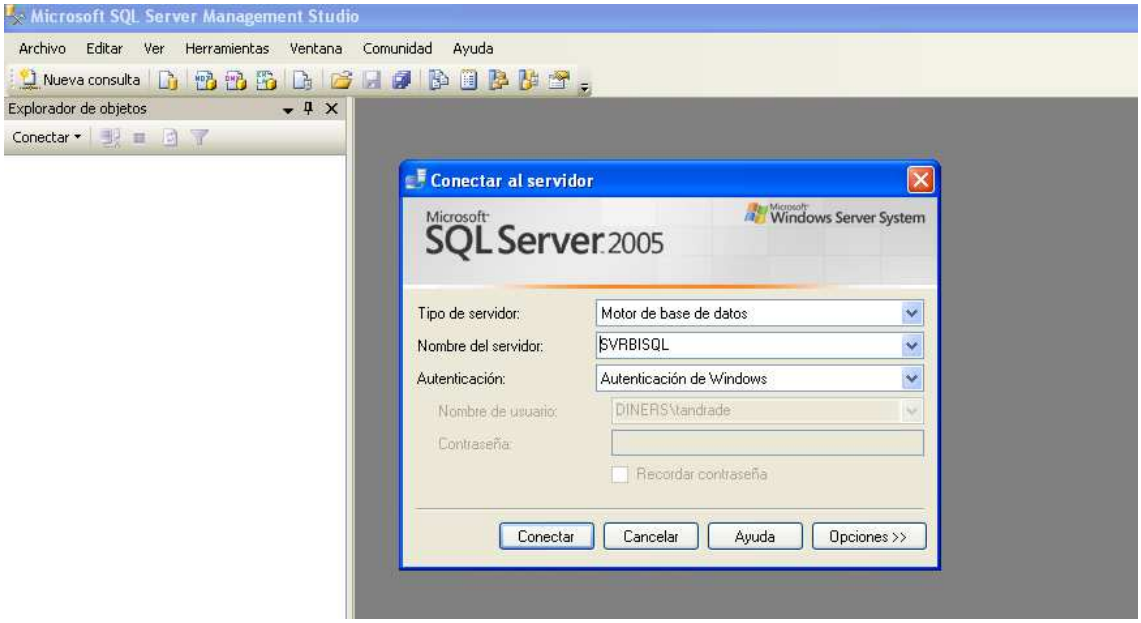


Figura Nº 4.24:"SQL SERVER 2005"

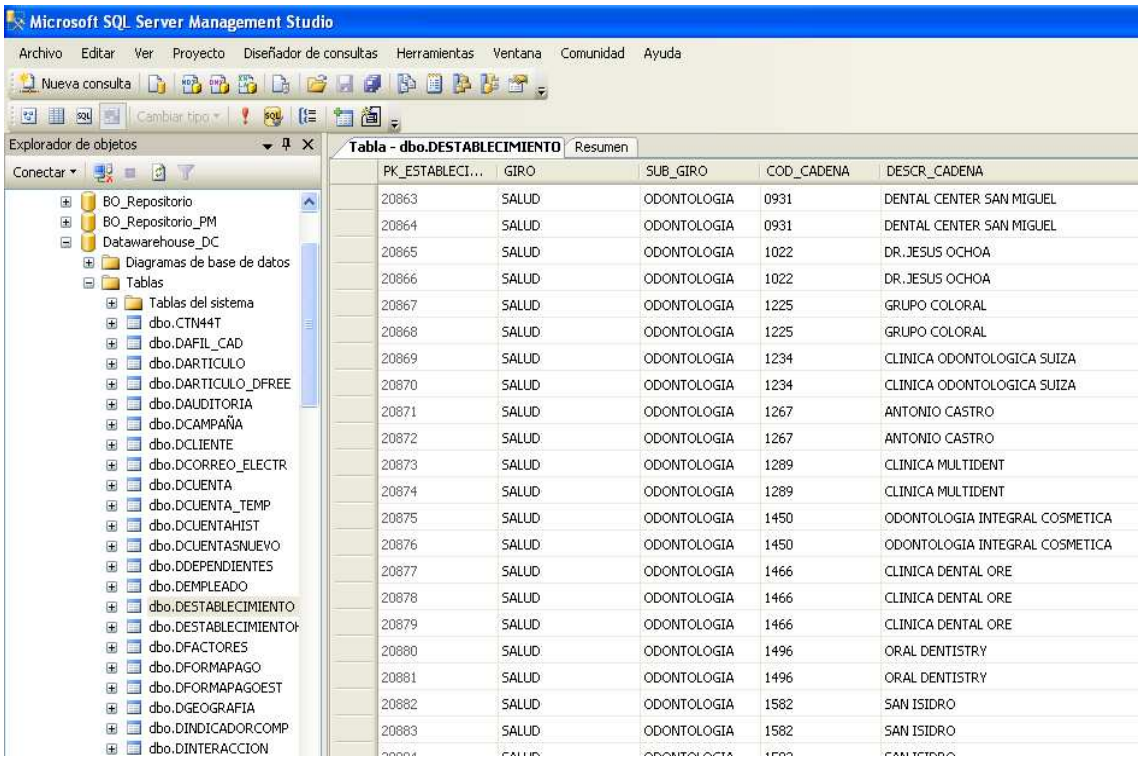


Figura Nº 4.25:"Base de datos SQL"

4.4.4 Integration Services

Para los procesos de carga se utiliza Integration Services de Microsoft Studio 2005.

4.4.4.1 Actualiza Dimensión Establecimiento

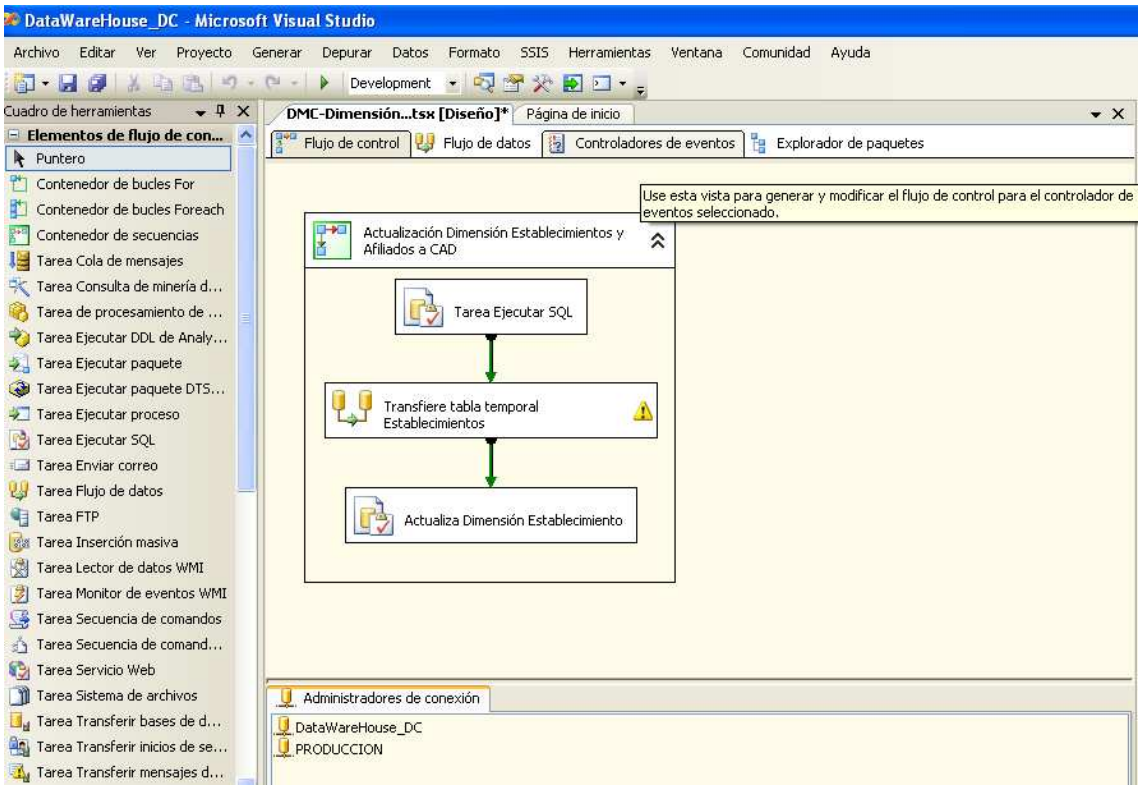


Figura Nº 4.26:"Integration Services Actualiza Dimensión Establecimiento"

4.4.4.2 Actualiza Tabla de Hechos Visita

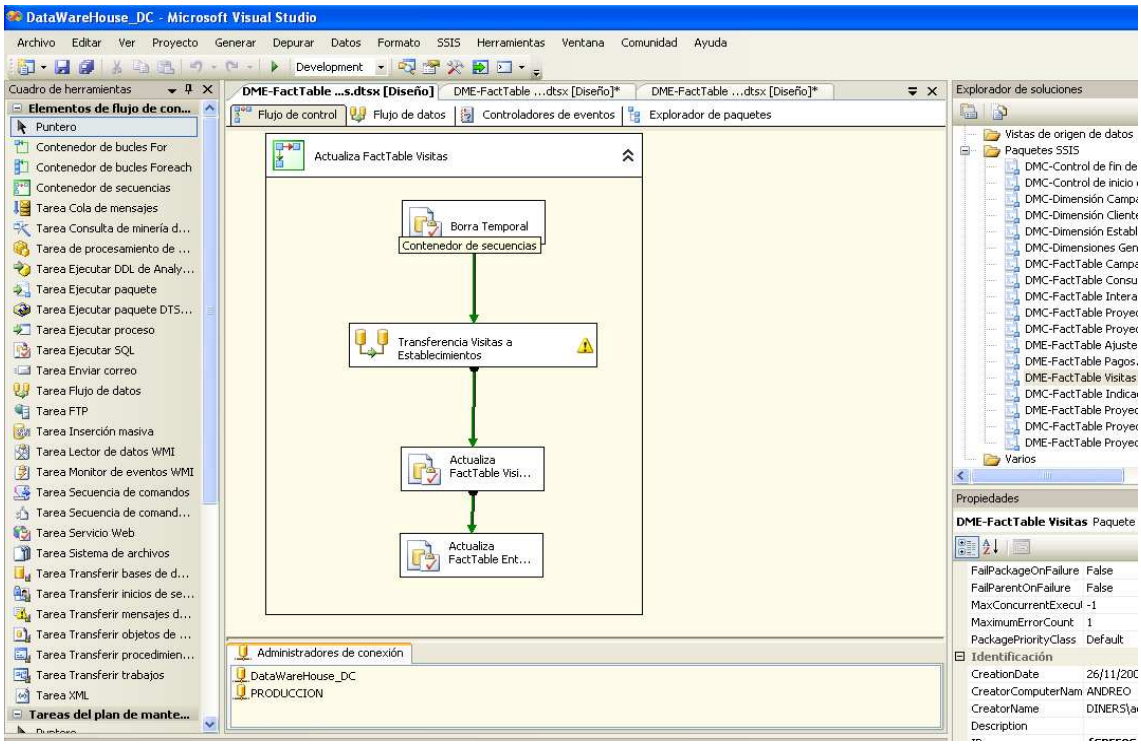


Figura Nº 4.27:"Integration Services Actualiza Tabla de Hechos Visita"

4.4.4.3 Actualiza Tabla de Hechos Ajustes

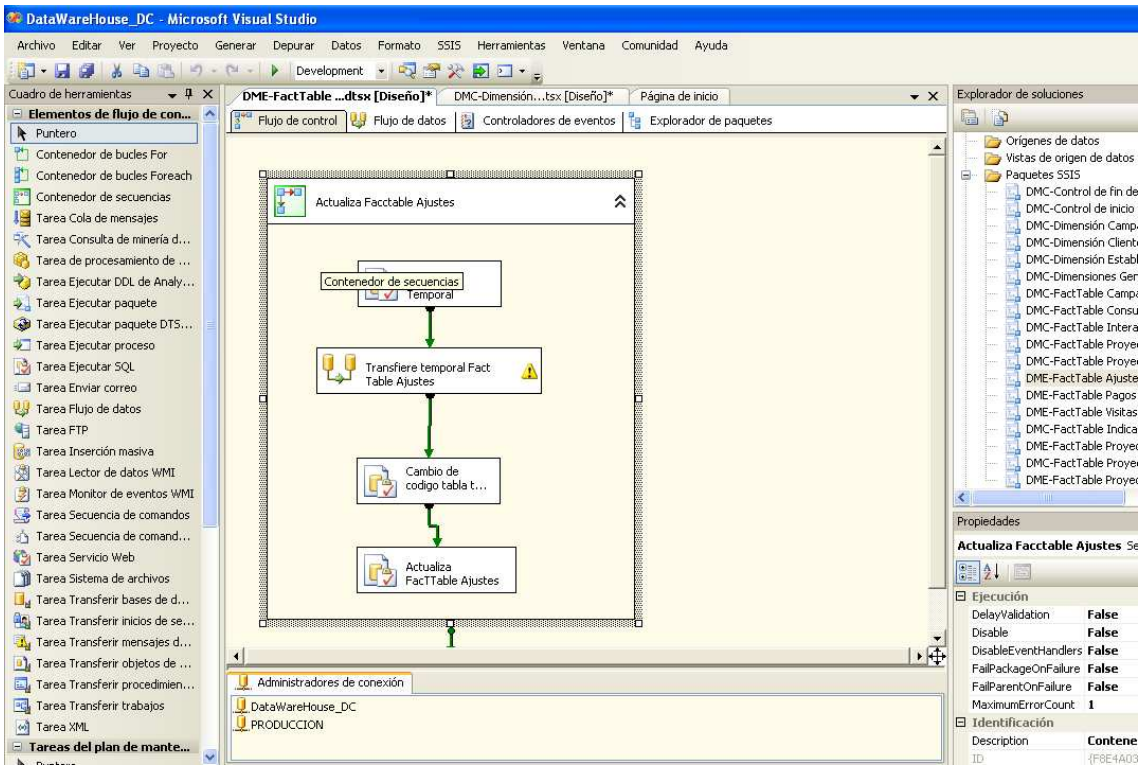


Figura Nº 4.28:"Integration Services Actualiza Tabla de Hechos Ajustes"

4.4.4.4 Actualiza Tabla de Hechos Pagos

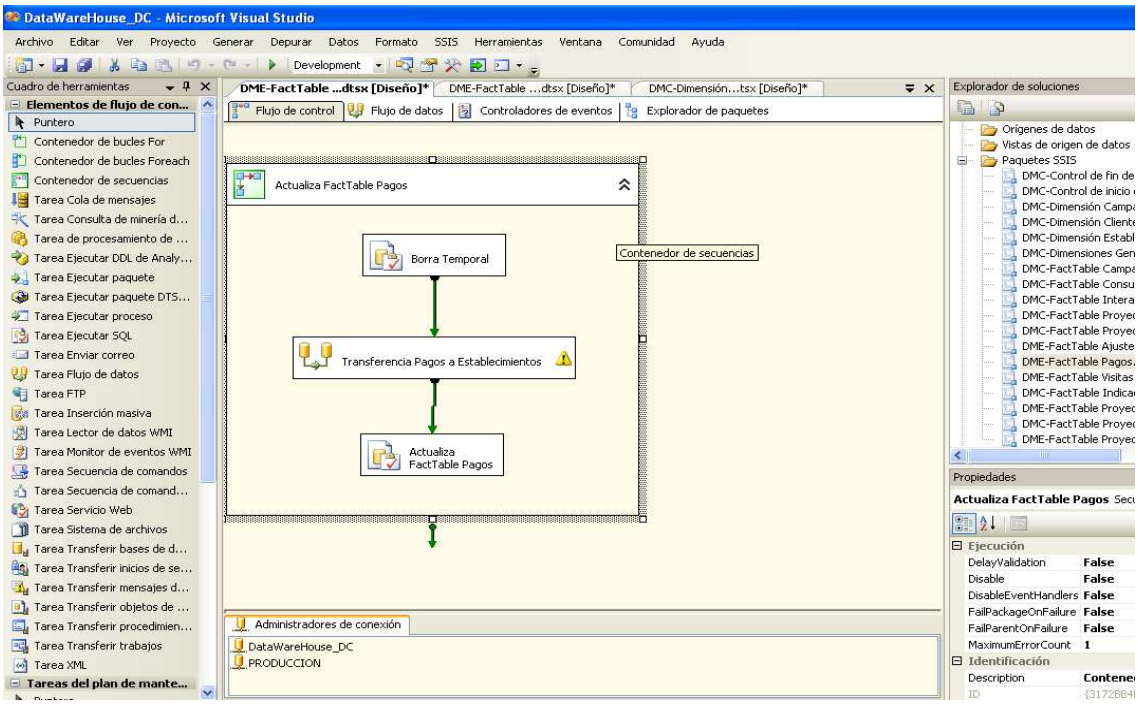


Figura Nº 4.29:"Integration Services Actualiza Tabla de Hechos Pagos"

4.4.4.5 Actualiza Tabla de Hechos Proyección de Visitas

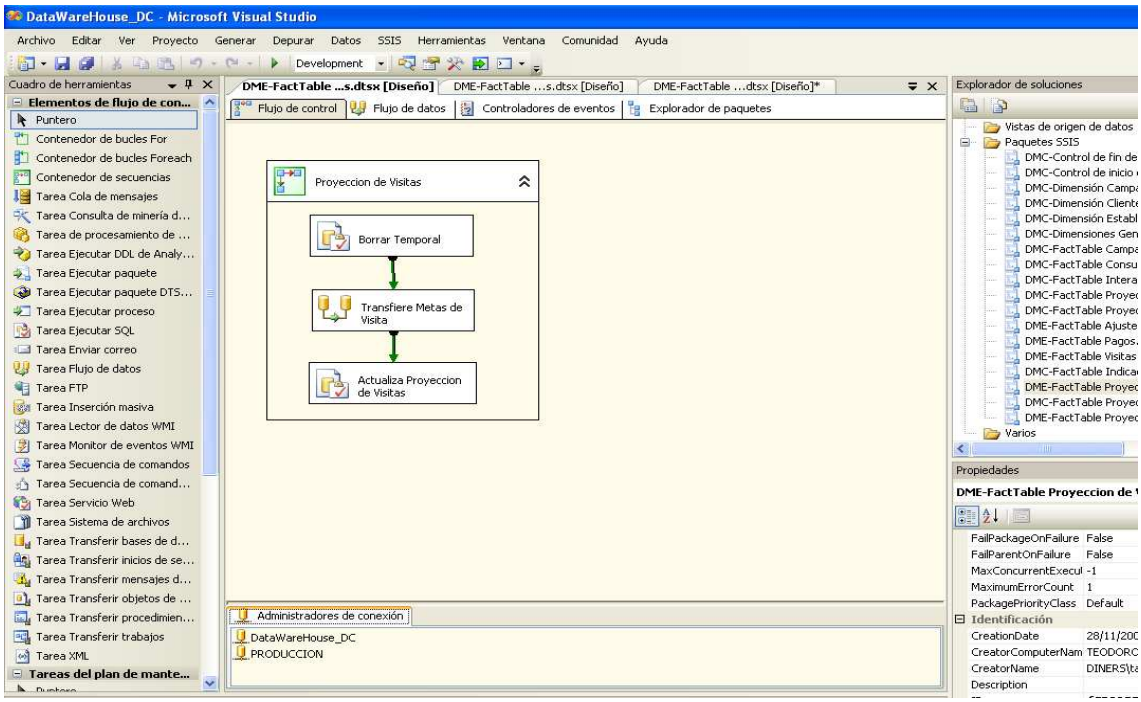


Figura Nº 4.30:"Integration Services Actualiza Proyección de Visitas"

4.4.4.6 Actualiza Tabla de Hechos Proyección de Afiliaciones

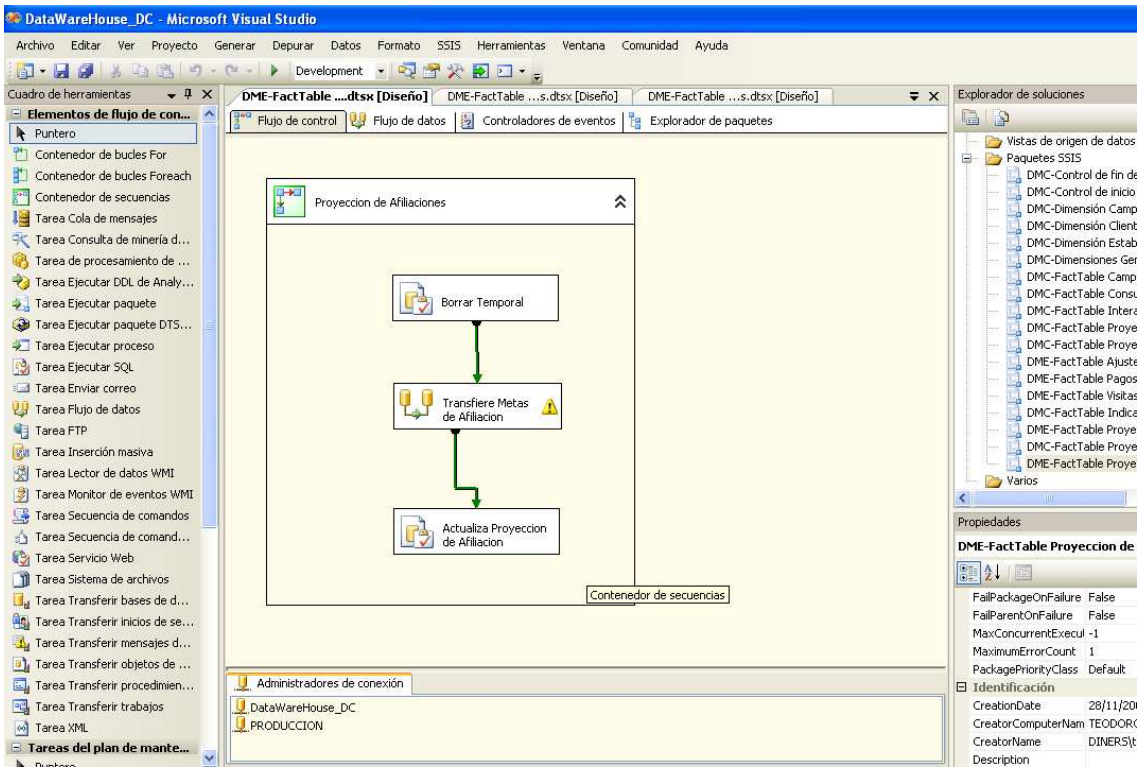


Figura Nº 4.31:"Integration Services Actualiza Proyección de Afiliaciones"

4.5 Explotación de la Información

En esta etapa se diseña e implementa el ambiente a través del cual, el usuario final accederá a la información almacenada en el Data Mart. Se definen los reportes base y los estándares a seguir en la elaboración de los mismos. En esta etapa se capacita a los usuarios en el uso de la solución y se implementan los mecanismos de seguridad para el acceso a los datos.

Figura Nº 4.32:"Explotación de Información"

Objetivos

- Desarrollar el ambiente de acceso a la información
- Desarrollar los reportes base definidos
- Capacitar a los usuarios

Tareas

- Definición de formatos de reportes a ser desarrollados
- Construcción de la capa semántica
- Creación de usuarios y asignación de perfiles de usuario
- Restricción de universos de información y de reportes en función a perfiles de usuario
- Desarrollo de reportes básicos
- Comprobación de resultados
- Capacitación
- Soporte Post – Implementación

Para la explotación de la información y el desarrollo de aplicaciones se emplea el producto Business Objects XI Release 2, esta plataforma nos ofrece un conjunto de herramientas para la administración de desempeño, administración de la información, generación de informes y consultas de análisis, todo esto integrado forma la plataforma de inteligencia de negocios de Diners Club S.A.

Para ello se construye el universo de datos y la capa semántica que se alimenta de la base de datos Datawarehouse_DC.

Figura Nº 4.33:"Ingreso al Diseñador"

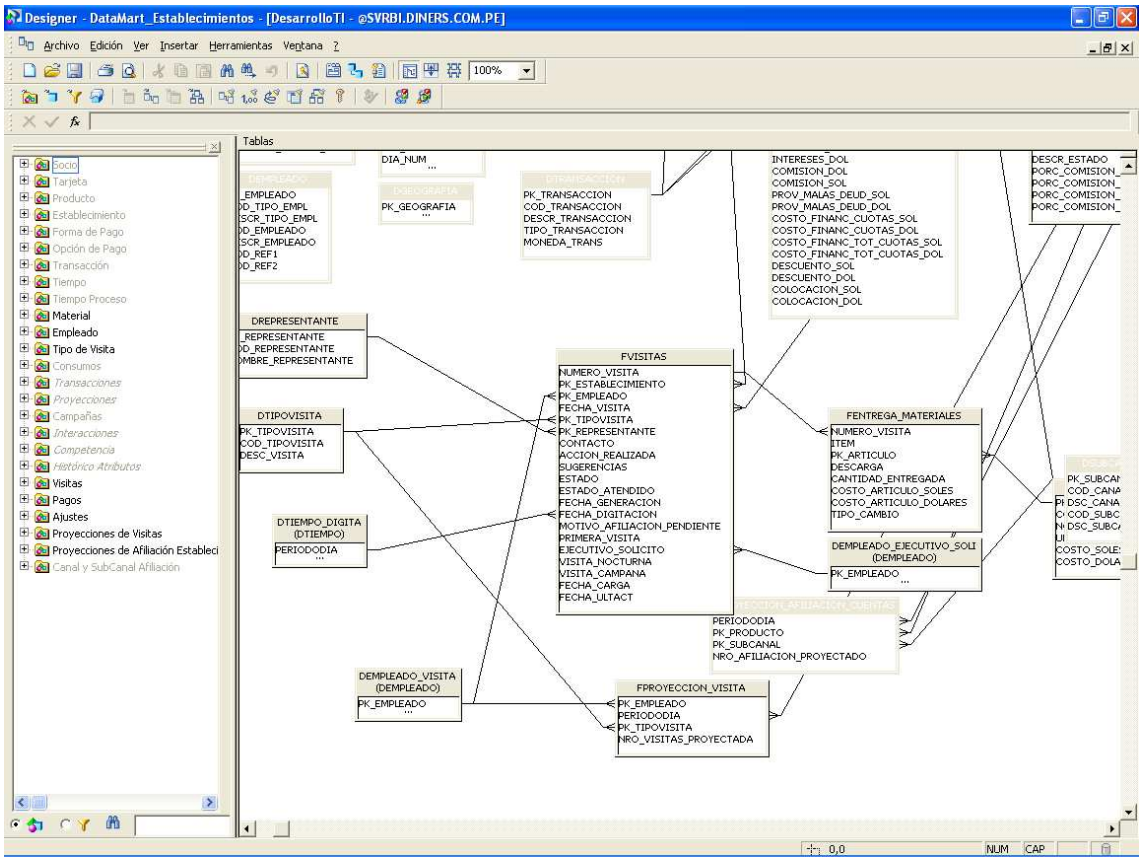


Figura Nº 4.34:"Creación Universo"

Durante esta etapa se lleva a cabo la capacitación al personal con talleres para la construcción de consultas de análisis y construcción de reportes.

Por otra parte se realiza la configuración de los permisos a las carpetas tanto públicas como privadas, se programa los envíos de los reportes desarrollados, formatos de entrega y administración de seguridad y servicios.

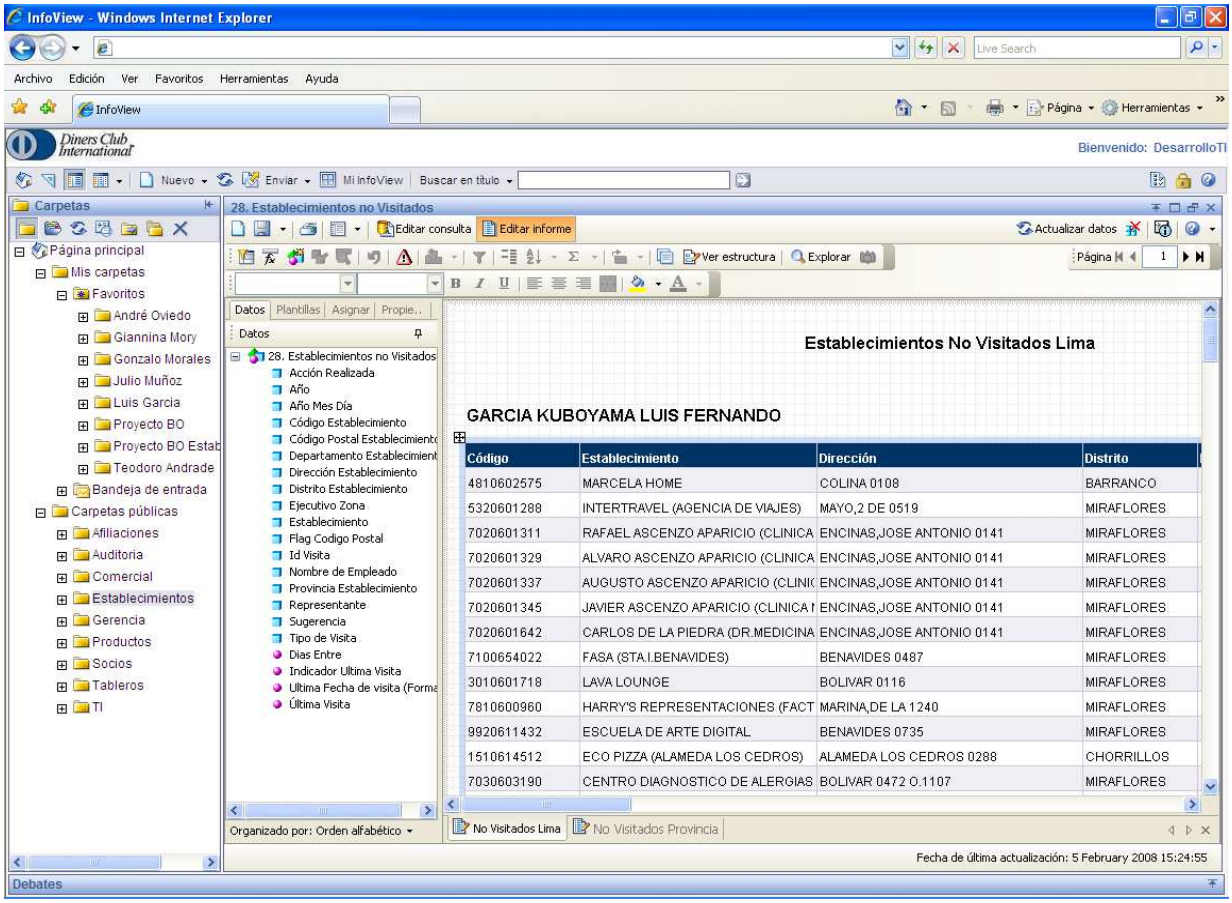


Figura Nº 4.35:"Generación de Reporte"

Capítulo 5: Solución Tecnológica

Soporte Tecnológico

Para la implementación de la solución tecnológica se cuenta con la siguiente infraestructura:

Un servidor de Base de Datos

Características:

Nombre: SVRBISQL Servidor SQL 2005

Modelo: IBM eServer BladeCenter HS21 -[8853L4U]-

Sistema Operativo: Windows 2003 Server

Memoria Física: 2048 Mb

Procesador: Intel(R) Xeon(R) CPU 5140 @ 2.33GHz HT

Disco Duro: 160 GB

Un Servidor BO

Nombre: SVRBI Servidor Business Objects

Modelo: IBM eServer BladeCenter HS21 -[8853L4U]-

Sistema Operativo: Windows 2003 Web Server

Memoria Física: 2048 Mb

Procesador: Intel(R) Xeon(R) CPU 5140 @ 2.33GHz HT

Disco Duro: 80GB

15 licencias Business Objects para el uso de la herramienta.

Presentación de la Solución

La solución está hecha en entorno Web por lo tanto debe ingresar la dirección http://biserver:5030/businessobjects/enterprise315/desktop/InfoView/logon/logo.do,

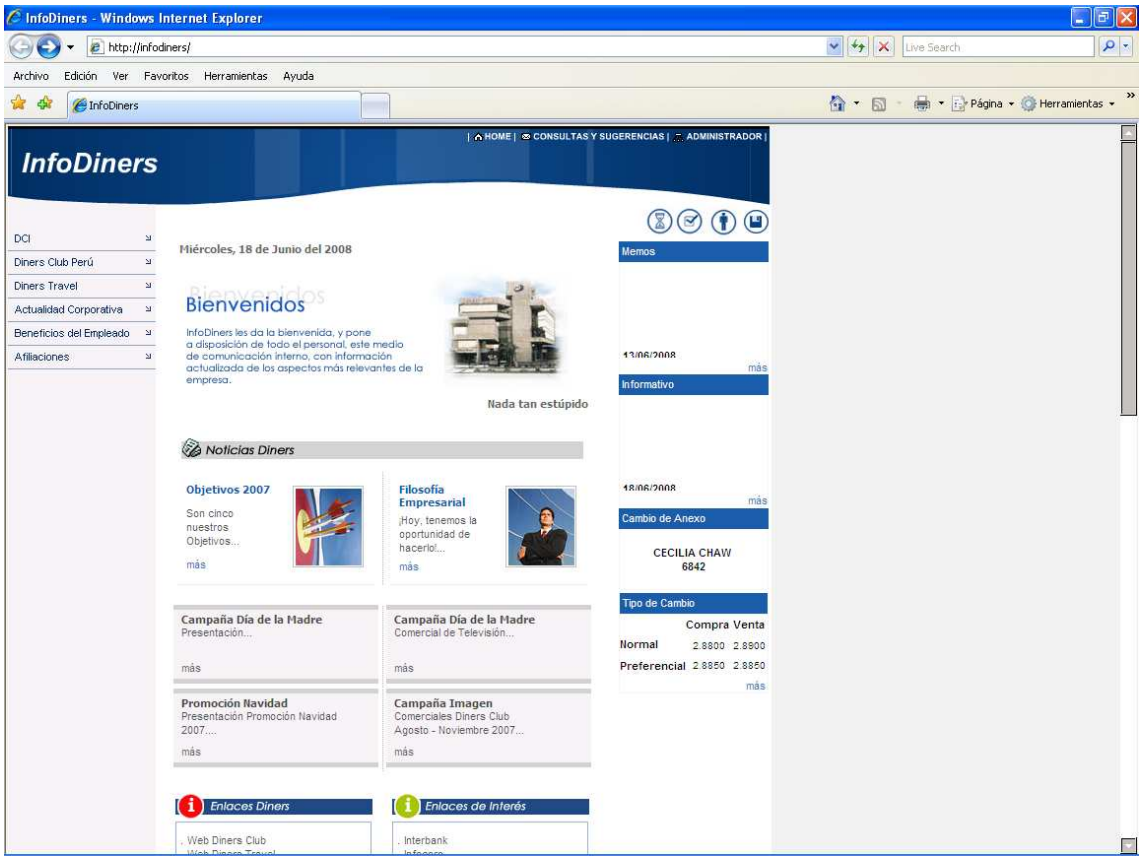


Figura Nº 5.1:"Entorno Web"

Se visualiza la siguiente pantalla, para acceder a la aplicación debe ingresar el usuario y contraseña respectivamente.

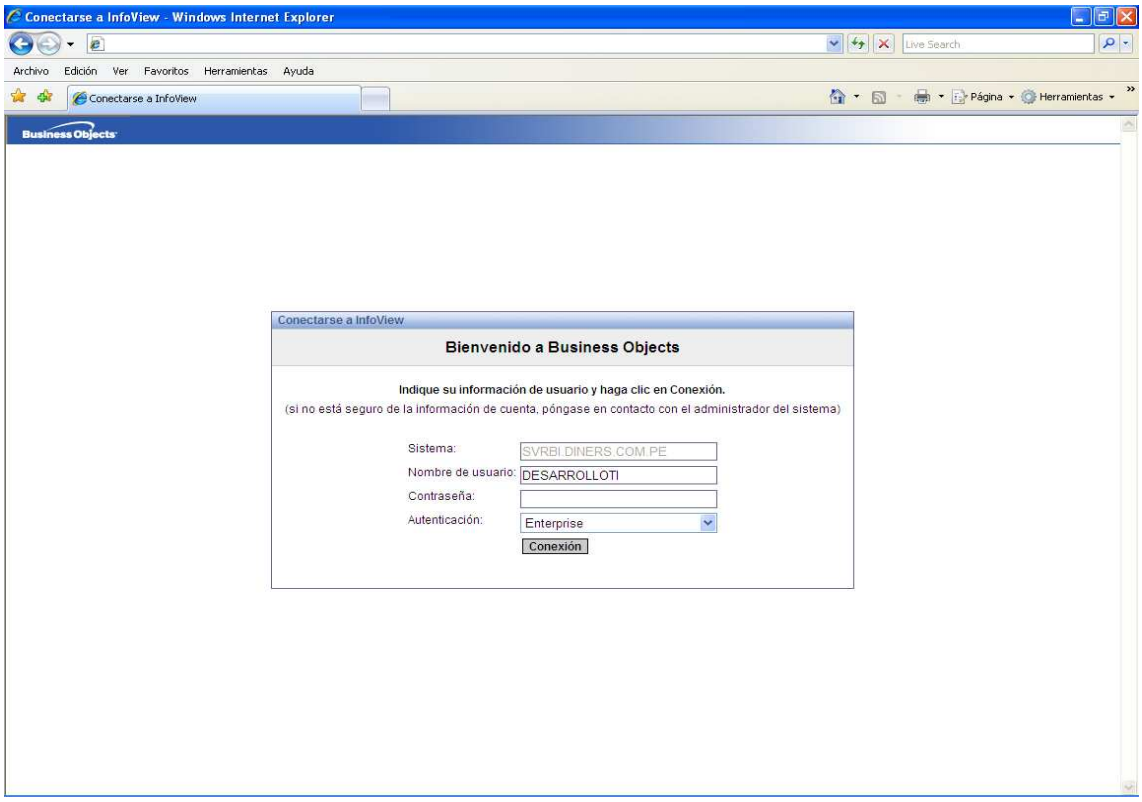


Figura Nº 5.2:"Ingreso a la Aplicación Business Objects"

Se debe elegir el datamart al cual se va a acceder para realizar el datawarehousing.

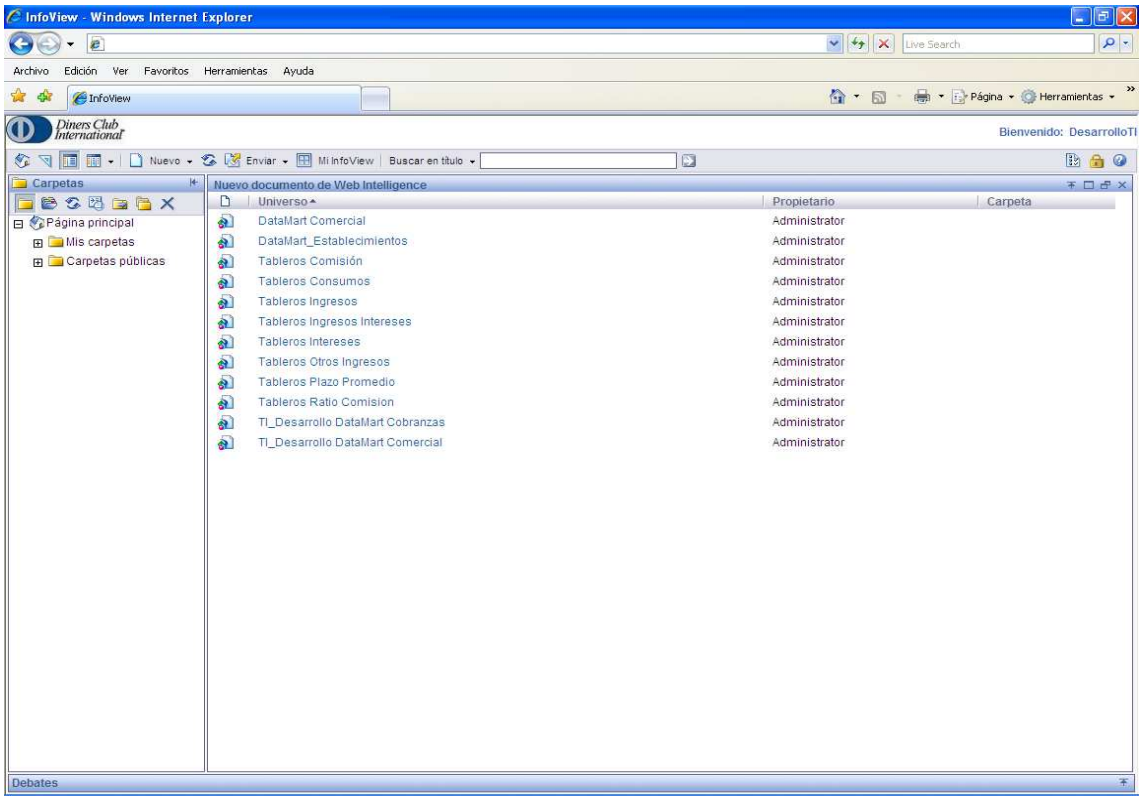


Figura Nº 5.3:"Lista de Datamarts y Tableros"

Desde ese momento se puede realizar la generación de informes y consultas de análisis, para ello se tienen las clases que contienen los atributos e indicadores definidos en el modelo.

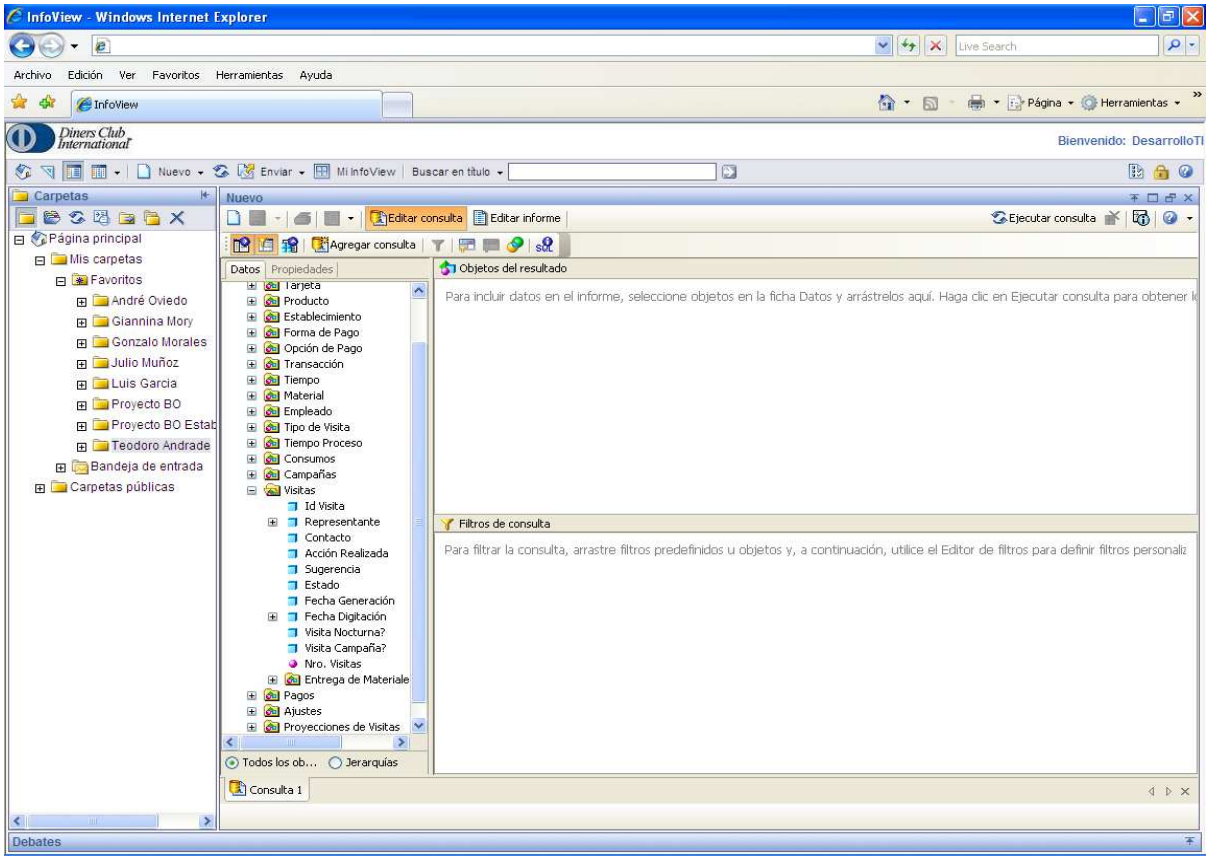
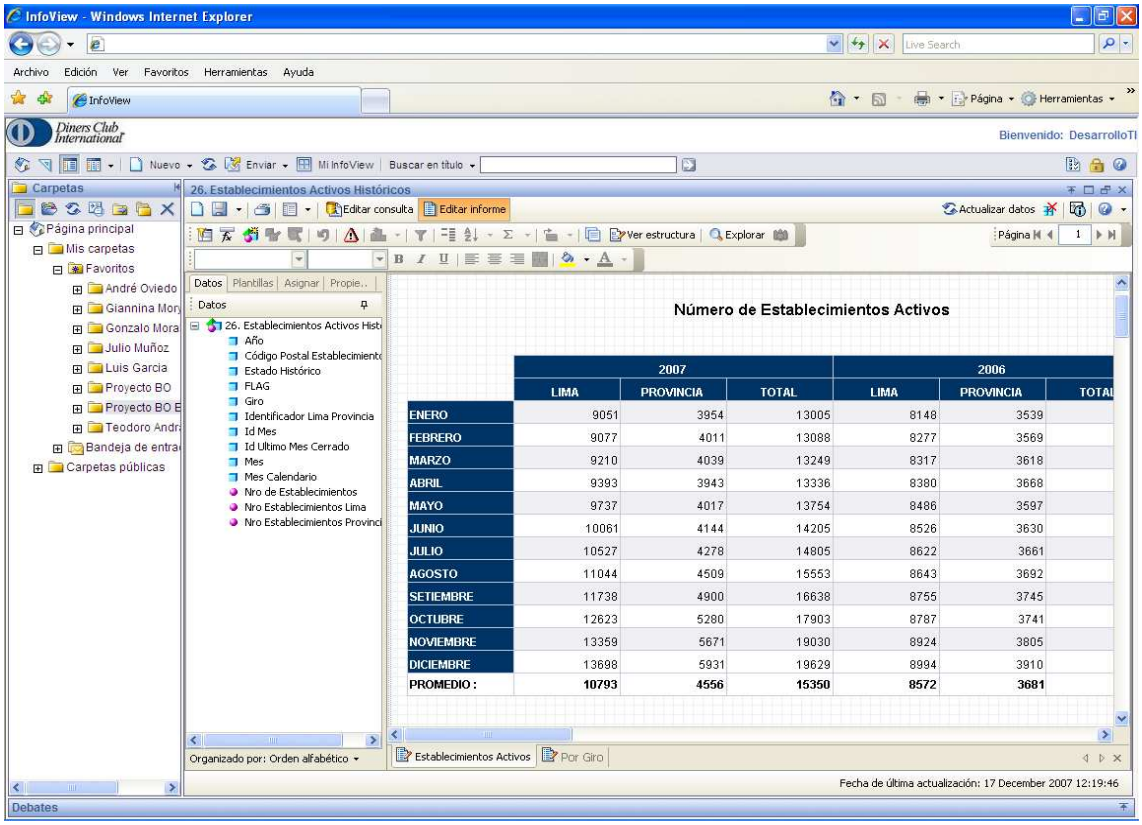


Figura Nº 5.4:"Construcción de Consultas"

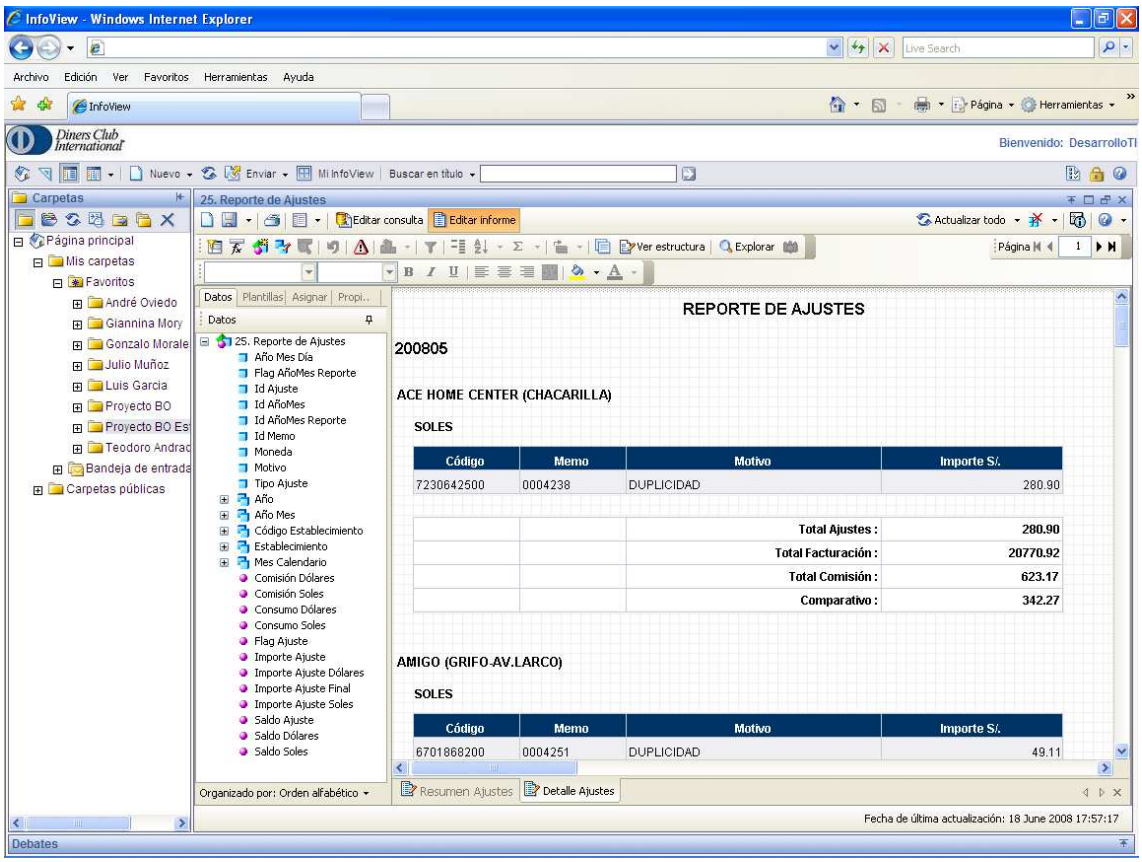
Los siguientes son algunos de los ejemplos de algunos reportes realizados como parte del proyecto.



Número de Establecimientos Activos

	2007			2006		
	LIMA	PROVINCIA	TOTAL	LIMA	PROVINCIA	TOTAL
ENERO	9051	3954	13005	8148	3539	
FEBRERO	9077	4011	13088	8277	3569	
MARZO	9210	4039	13249	8317	3618	
ABRIL	9393	3943	13336	8380	3668	
MAYO	9737	4017	13754	8486	3597	
JUNIO	10061	4144	14205	8526	3630	
JULIO	10527	4278	14805	8622	3661	
AGOSTO	11044	4509	15553	8643	3692	
SETIEMBRE	11738	4900	16638	8755	3745	
OCTUBRE	12623	5280	17903	8787	3741	
NOVIEMBRE	13359	5671	19030	8924	3805	
DICIEMBRE	13698	5931	19629	8994	3910	
PROMEDIO :	**10793**	**4556**	**15350**	**8572**	**3681**	

Figura Nº 5.5:"Reporte I"



REPORTE DE AJUSTES

200805

ACE HOME CENTER (CHACARILLA)

SOLES

Código	Memo	Motivo	Importe S/.
7230642500	0004238	DUPLICIDAD	280.90
		Total Ajustes :	**280.90**
		Total Facturación :	**20770.92**
		Total Comisión :	**623.17**
		Comparativo :	**342.27**

AMIGO (GRIFO-AV.LARCO)

SOLES

Código	Memo	Motivo	Importe S/.
6701868200	0004251	DUPLICIDAD	49.11

Figura Nº 5.6:"Reporte II"

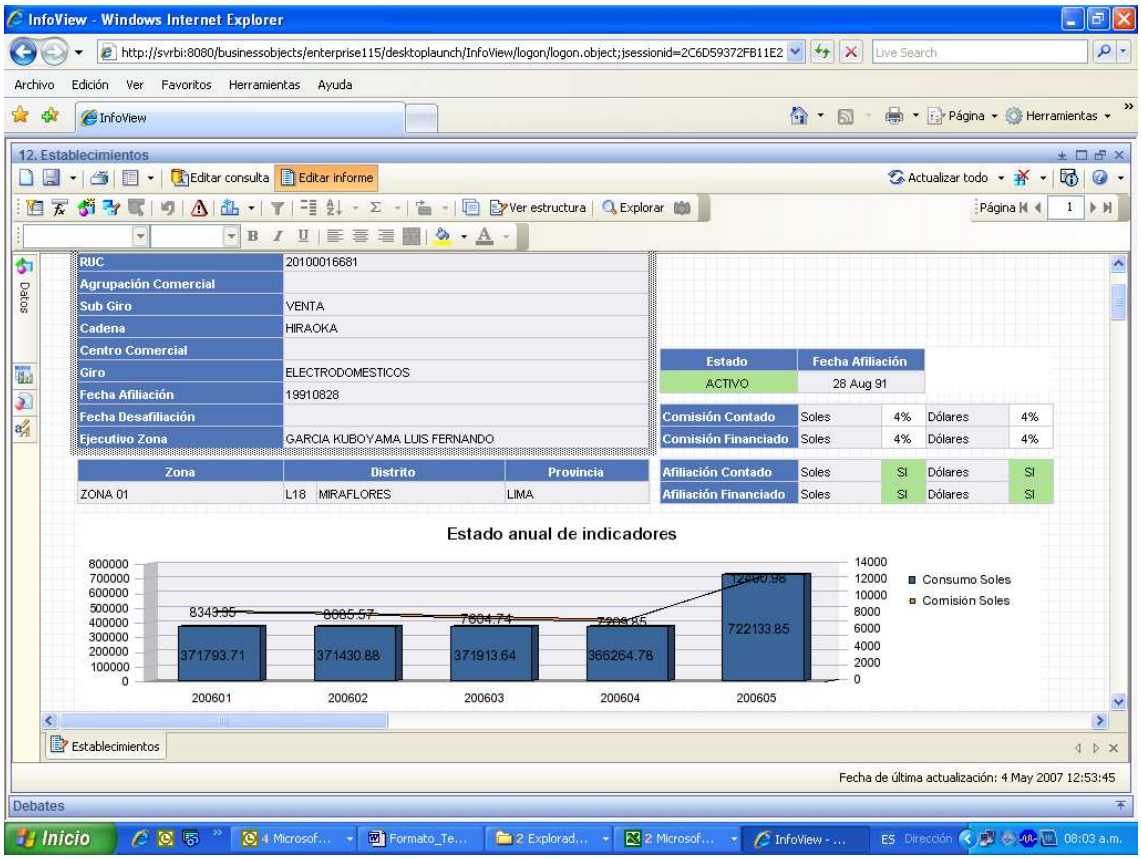


Figura Nº 5.7:"Reporte III"

Capítulo 6: Conclusiones y trabajos futuros

6.1 Conclusiones

Durante todas las etapas de desarrollos pudimos concluir lo siguiente:

Inteligencia de negocios es un proceso de negocios, no un proyecto de tecnología se apoya en ella.

La tecnología adecuada debe soportar el proceso, así como una metodología de desarrollo.

La experiencia en este tipo de desarrollos es fundamental, comenzar un proyecto como este debe contar con apoyo de la gerencia, debe comprometer a los usuarios, establecer la arquitectura idónea para el desarrollo.

La selección de la herramienta es decisiva para el éxito del proyecto.

Mala información origina malas decisiones.

Un buen modelamiento contribuye en gran medida a una buena performance.

Un proyecto de gestión como este requiere la utilización de metodología de desarrollo y de manejo de proyecto.

Para los gerentes de alto mando, el tiempo y la rapidez en los procesos son de vital importancia. Es por eso que el desarrollo de la tecnología ha ido orientada a agilizar el trabajo en las empresas.

En un principio los usuarios de la empresa no tenían mayores conocimientos en materia de análisis multidimensional en lo que respecta a las diferentes maneras en que se puede analizar su información. Se sorprendieron de cómo se realiza la obtención de información de manera rápida y confiable que apoya la toma de decisiones y permite realizar una mejor gestión.

6.2 Trabajos Futuros

Como trabajo futuro en la empresa se deben integrar otros departamentos y nuevos sistemas para un análisis global de la organización. Actualmente se viene planificando

el desarrollo de Datamart Cobranzas, el mismo que debe integrarse a los ya desarrollados Comercial y Establecimientos.

Desarrollar los Dashboards para monitorear y definir los objetivos estratégicos además del control de su cumplimiento.

También en un futuro hacer posible que este Data Warehouse pueda ser accedido a través de Internet y así no sea necesario estar dentro de la red institucional para obtener la información necesaria.

Para finalizar hacer Data Mining para predecir futuras tendencias y comportamientos, y tomar decisiones pro-activas para conducirlas por un conocimiento acabado de la información.

REFERENCIAS BIBLIOGRÁFICAS

[DIA2002] DÍAZ, WLADIMIRO, http://www.uv.es/~diaz/,Departament d e Informática, Universidad de Valencia, Valencia, España.

[FER2001] FERRANDO, SANTIAGO. CARREIRA, IVÁN, “Soluciones Data Warehouse bajo el nuevo modelo de organización a-business: La necesidad de integrar la información” Telefónica Investigación y Desarrollo

[FRA1997] FRANCO, JEAN-MICHEL, "El Data Warehouse el Data Mining",

Primera Edición, Ediciones Gestión 2000, 1997, pp33-34.

[HER2003] HERNANDEZ, JOSE,

http://www.dsic.upv.es/~jorallo/cursoDWDM/dwdm-I.pdf Departamento de Sistemas Informáticos y Computación Universidad Politécnica de Valencia, España.

[INC2001] INCO INSTITUTO COMPUTACIÓN,

http://www.fing.edu.uy/inco/grupos/csi/esp/Cursos/cursos_act/2003/DAP_SistDW/index.html, Universidad de la Republica de Montevideo, Uruguay.

[INM2002] INMON, W.H, Building the Data Warehouse, Tercera Edición, John Wiley & Sons, 2002.

[KIM1996] KIMBALL, RALPH, The Data Warehouse Toolkit,

http://www.dwinfocenter.org/ defined.html, 1996.

[MAR2001] MARCH, ALAN, LEGUIZA, JORGE, “Tecnologias OLAP en la gestión de servicios de salud: su aplicación en estudios del gasto en medicamentos”, Facultad de Ciencias de la Administración, Universidad del Salvador, Buenos Aires, Argentina.

[OST1993] SUSAN OSTTERFELDT, "Data Warehousing, Data Modeling and Desing", MicroStrategy Education, nov 96.

[PER2001] PERALTA, VERONICA, http://www.fing.edu.uy/~vperalta/, Facultad de Ingeniería, Universidad de la República, Montevideo, URUGUAY.

[ROB2000] ROBBINS, STEPHEN; Coulter Mary “Administración” sexta edicion, Prentice-Hall, 2000.

[SAS2003] SAS INSTITUTE EUROPEAN OFFICE, http://www.sas.com, SAS Institud Inc. World Headquartes.

[ULL1999] ULLMAN, J.D.; Widom, J. “Introducción a los Sistemas de Bases de

Datos”. Prentice Hall, 1999.

[YDI2004] YDIRÍN, PRÉSTAMO,

http://140.148.3.250/u_dl_a/servlet/mx.udlap.ict.tales.html,

Escuela de Ingeniería, Universidad de las Américas, Puebla, México.

[Bed97] Jeff Bedell, “Data Warehousing, Data Modeling and Design”, DSSTraining, MicroStrategy, Inc., 1997.

[Gol99] Matteo Golfarelli y Stefano Rizzi, 1999. A Methodological Framework for Data Warehouse Design. University of Bologna. ACM.

[Inm92] William H. Inmon, 1992. “Building the Data Warehouse”, Wiley-QED John Wiley & Sons, Inc..

[Kim92] Ralph Kimball, 1992. The Data Warehouse Toolkit, Wiley Computer Publishing

[Kim95] Ralph Kimball, 1995. Is ER Modeling Hazardous to DSS?, DBMS Magazine October.

[Kim98] Ralph Kimball, Laura Reeves, Margy Ross, Warren Thornthwaite, 1998. The Data Warehouse Lifecycle Toolkit, Wiley Computer Publishing.