

**UNIVERSIDAD AMERICANA
FACULTAD DE INGENIERIA**



**GUÍA METODOLÓGICA PARA LA
DEFINICIÓN Y DESARROLLO DE UN DATA
WAREHOUSE**

Br. ERICKA GRACIELA SEVILLA BERRÍOS

Monografía para optar al grado de
INGENIERO EN SISTEMAS

Profesor Tutor
ING. CARLOS ALBERTO ZEPEDA S.

Managua, Nicaragua, Marzo 2003

A Ericka Daniela,
el ser que ha sido capaz de llenar
con su esencia y presencia cada elemento de mi vida.

Gracias
por mostrarme el significado de
vivir y amar

AGRADECIMIENTOS

A Ronaldo Sevilla y Vida Berríos, mis amados padres, quienes me dieron la vida y hasta el día de hoy han guiado mi camino con su amor y dedicación, porque siempre han estado allí para mí, con sus brazos abiertos llenos de comprensión y la preocupación en el corazón, compartiendo mis alegrías y tristezas, infinitas gracias y aún es poco.

A mi querido hermano Ronaldo, siempre dispuesto, perseverando y acompañándome en el andar de la vida y para la construcción de esta monografía.

A mi tutor Carlos, por su disponibilidad y ayudarme a delimitar y esclarecer mis ideas, a Flavia e Irene y demás amigos por todos sus consejos y su tiempo, y a mis compañeros y colegas que juntos recorrimos la senda del saber.

A todas aquellas personas que no han escatimado sus palabras ni acciones para apoyarme, darme ánimos, ideas, etc y que por razones de espacio no puedo enumerar.

CONTENIDO

INTRODUCCIÓN.....	i
I. ANTECEDENTES	iii
II. OBJETIVOS	v
A. General:	v
B. Específicos:.....	v
DESARROLLO DE LA TESIS.....	1
I. MARCO TEÓRICO	1
A. El Data Warehouse.....	1
B. Metodologías más utilizadas para construir un Data Warehouse	6
II. ANÁLISIS CUALITATIVO DE LAS METODOLOGÍAS	40
A. Barry Devlin.....	41
B. William Inmon.....	43
C. Ralph Kimball	45
D. Síntesis	48
III. CONTEXTO NICARAGÜENSE.....	50
IV. ESTUDIO DE FACTIBILIDAD.....	53
A. Factibilidad Técnica	54
B. Factibilidad Operativa	57
C. Factibilidad Económico-Financiera	61
D. Factibilidad Legal.....	67
E. Sugerencia para el proyecto.....	69
V. GUÍA METODOLÓGICA PARA LA CONSTRUCCIÓN DE UN DATA WAREHOUSE	70
A. Etapa 1: Análisis y Comprensión del Entorno Analítico	72
B. Etapa 2: Definición de una Arquitectura.....	74
C. Etapa 3: Enriquecimiento de Datos	85
D. Etapa 4: Construcción en Incrementos, Data Marts.....	91
CONCLUSIONES.....	99
RECOMENDACIONES.....	101
ANEXO A: INDICES DE ILUSTRACIONES Y TABLAS	
ANEXO B: BIBLIOGRAFÍA	
ANEXO C: COTIZACIONES DE EQUIPOS Y CRONOGRAMA	
ANEXO D: ENTREVISTAS	

ANEXO E: REQUERIMIENTOS DE HERRAMIENTAS SOFTWARE

ANEXO F: MODELOS DE DATOS Y SCRIPTS

ANEXO G: CORRESPONDENCIA FUENTE DESTINO Y RUTINAS ETL

ANEXO H: REPORTES OBTENIBLES

ANEXO I: GLOSARIO

INTRODUCCIÓN

El presente estudio monográfico, satisface la necesidad de un marco de referencia para la construcción de un Data Warehouse, unificado y adaptado a profesionales o estudiantes de Informática que inician su aprendizaje sobre ésta tecnología. El resultado obtenido es una guía metodológica que apoya la definición y enseña el desarrollo de un Data Warehouse que reúne y unifica las metodologías más aceptadas y que además concuerda con el conocimiento que sobre este proceso tienen los principiantes que desean por primera vez introducirse en la construcción de un Data Warehouse. Todo esto con el propósito de que quien utilice y ponga en práctica la guía encuentre una referencia sólida y comprensible que le dirija a través del proceso de desarrollo de un Data Warehouse.

Un Data Warehouse es una colección almacenada de datos corporativos originados en diversos sistemas transaccionales dentro de una organización. Un Data Warehouse es un repositorio único de información cuyo propósito específico es soportar la toma de decisiones en un negocio, no las operaciones del mismo, por ello se encuentra organizado de una forma muy distinta a las bases de datos tradicionales.

En el contexto nacional, las empresas nicaragüenses que desean expandir su mercado más allá de nuestras fronteras o mejorar su rendimiento en el sector en el que se desempeñan, deben ser competitivas y obtener ventajas de todos sus recursos para asegurar la efectividad de sus estrategias y su correspondiente remuneración. Uno de los recursos más importantes con los que cuenta cualquier organización es la información. La explotación de la información de la empresa habilita a los ejecutivos tomar decisiones más adecuadas y efectivas, fundamentalmente reconocer oportunidades y posibilidades de crecimiento.

Sin embargo, el uso apropiado que se haga de la información de una empresa estará en dependencia de la calidad de los datos que la originan, la claridad del significado que tienen para los tomadores de decisiones, su estructura lógica, organización y la disposición a tiempo y en forma. Si los datos son veraces y adecuados pero son extemporáneos difícilmente servirán para la toma de decisiones, ya que la problemática que se le presentó a la empresa no pudo ser resuelta de la mejor manera posible. La velocidad

en la evaluación de diferentes alternativas es un elemento primordial para la supervivencia en el mundo de los negocios actuales.

El escenario anterior puede ser resuelto apropiadamente con la construcción de un Data Warehouse. Cuando la implementación de un Data Warehouse es exitosa aumentan las visitas de los usuarios al Data Warehouse y es en este aspecto en donde se produce el retorno de la inversión, ya que a través de las decisiones tomadas a partir de la información consultada se incrementa la eficiencia de la organización, elevando los márgenes de rentabilidad de la misma.

Las empresas que tienen conocimiento de todas las ventajas que se obtienen de un Data Warehouse tienen entre sus planes estratégicos la construcción de uno. Sin embargo, el desarrollo de un Data Warehouse es una inversión cara y delicada, por el tipo de herramienta y plataforma informática que se necesita para su construcción. Además, el personal especializado que puede asesorar la implementación exitosa es escaso en el país, lo que hace costoso su reclutamiento. Los factores anteriores son los que más han detenido a proyectos de Data Warehouse de ser implementados.

Esta guía adaptada a profesionales o estudiantes de Informática que inician su aprendizaje sobre esta tecnología logrará hacer más fáciles las iniciativas de construcción de Data Warehouses ya se cuenta con un marco teórico de referencia sólido y sencillo que guiará a todo aquel que desee realizar una primera experiencia de desarrollo de Data Warehouse sin necesidad de comprometerse en un proyecto de gran envergadura sin previa experiencia, proporcionando de esta manera personal con buenas bases y nociones de lo que es Data Warehousing a empresas que inicien el desarrollo de un Data Warehouse.

Esta guía potencia la capacidad de las personas para planificar, de manera consistente, la implementación del Data Warehouse dentro de la empresa y es dirigida al personal de Tecnología de Información que pertenece a la organización en cada una de las etapas del proyecto de Data Warehouse, tomando lo mejor de las metodologías más utilizadas y adaptándolo a principiantes, pero que puede ser fácilmente retomada por el director del proyecto como referencia de apoyo.

I. ANTECEDENTES

El término Arquitectura de una Organización se refiere a la colección de componentes tecnológicos y sus interrelaciones que integrados nos permiten satisfacer las necesidades de información en una compañía. Aquí se introduce el concepto de Arquitectura de Tecnología de la Información de una Organización con la intención de proporcionar un base de referencia sobre la tecnología de Data Warehousing como uno de muchos componentes de una arquitectura de Tecnología de la Información (IT).

El concepto de Data Warehouse se desarrolla basándose en la necesidad de encontrar una estructura óptima para analizar los datos dejando el enfoque tradicional de estructuras orientadas al registro de transacciones; con ello la tecnología de data Warehousing se establece como un modelo que evoluciona paralelamente al cambio en las necesidades actuales de construir sistemas que faciliten el proceso de toma de decisiones.

Según la definición de W. H. Inmon, “Un Data Warehouse es un conjunto de datos integrados, orientados hacia una materia, que varían con el tiempo y que no son transitorios, los cuales apoyan el proceso de toma de decisiones de una administración”[INMON92]. En otras palabras, un Data Warehouse es el almacenamiento de datos con fines estratégicos de negocio, concepto muy opuesto al objetivo de las bases de datos operativas. La importancia reside en que son fuente de materia prima selecta en los sistemas de soporte para la toma de decisiones.

Los profesionales de IT deben asegurar una infraestructura IT que soporte adecuadamente una gran cantidad de requerimientos de los diferentes usuarios de negocios, cada uno de los cuales tiene necesidades constantes y diferentes. Además, continuamente deben aprender nuevas tecnologías, evaluar nuevas herramientas y mantenerse en contacto permanentemente con sus asociados tecnológicos. No todas las últimas tecnologías son útiles, de modo que es necesario escoger de entre el “rompecabezas” e integrar las distintas piezas para formar un todo coherente. Este problema es resuelto a través de la construcción de un Data Warehouse.

Ante esta situación los profesionales de IT han desarrollado estrategias y metodologías para la construcción de Data Warehouses según entorno y propósito de

desempeño específicos. No existe, un punto de vista único que guíe la construcción y adaptación de un Data Warehouse a los requerimientos de las organizaciones en Nicaragua, debido a que las necesidades de cada una de ellas son diferentes, al igual que su contexto. Sin embargo, es posible establecer un marco de referencia unificado que brinde, a cada participante novel de un proyecto de Data Warehouse, el apoyo necesario para su definición y desarrollo.

II. OBJETIVOS

A. General:

Diseñar una guía metodológica para que apoye la definición y muestre el desarrollo de un Data Warehouse, mediante la elaboración de un estudio para adaptar los enfoques y estrategias de construcción de las metodologías más utilizadas hasta el año 2001, a Tecnólogos de la Información noveles en Data Warehousing.

B. Específicos:

- i.* Proponer una estrategia para la planificación, definición y desarrollo de un Data Warehouse, a través de un análisis cualitativo de las metodologías más utilizadas, para lograr un método base que sea fácil, entendible, práctico y adaptado a noveles en Data Warehousing.
- ii.* Precisar la viabilidad de la puesta en práctica de la guía, a través de la realización de un análisis de Costo-Beneficio, para la factibilidad de construcción de un Data Warehouse utilizando la guía metodológica propuesta.
- iii.* Esbozar una propuesta de diseño de Data Warehouse, que utilice los criterios formulados en esta guía metodológica dentro de una institución financiera, a través del uso de Modelado Dimensional de datos, para comprobar la funcionalidad de la guía.

DESARROLLO DE LA TESIS

I. MARCO TEÓRICO

A. El Data Warehouse

La concepción generalizada que se tiene acerca de lo que es un Data Warehouse, implica algún tipo de Base de Datos que proporciona varias funciones a varios usuarios, proviene más bien de la implementación del Data Warehouse que de su definición conceptual. La divergencia de criterios sobre una el concepto en sí se cimienta en el hecho de que no exista una definición oficial de Data Warehouse, esto es, una definición estándar soportada por un comité de estándares. A pesar de esto, las características del entorno analítico que supone ser el Data Warehouse son ampliamente reconocidas y aceptadas. Ambos, definición y características a continuación.

1. Definición

Data Warehouse es un anglicismo que, como muchas de las expresiones que predominan en el mundo de la Tecnología de Información, no tiene traducción exacta al castellano, la mejor aproximación que se puede hacer sería llamarle Almacén de Datos. No obstante, si el Data Warehouse fuese exclusivamente un almacén de datos, los datos encontrados dentro de él no servirían para brindar soporte a la toma de decisiones dentro de una organización.

Al verificar fuentes bibliográficas se observa que diferentes libros y artículos atribuyen a diferentes autores el origen de la definición de Data Warehouse, sin embargo las tendencias que más predominan (las referencias más utilizadas) indican principalmente dos fuentes distintas.

Una de estas fuentes señala a Barry Devlin y Paul Murphy, quienes en 1988 escribieron "An Architecture for a Business and Information System"[Devlin88] que fue publicado en el IBM Systems Journal, y el cual es uno de los primeros artículos que describe la arquitectura de un Data Warehouse, que fue desarrollado entre 1985 y 1986 para uso interno de IBM Europa en Dublín. Ellos hicieron uso del término Information Warehouse para referirse al proyecto, dicha expresión en la actualidad es ampliamente usada para referirse a un Data Warehouse.

Posteriormente [Devlin97,p20] definió que “Data Warehouse es un único, completo y consistente almacén de datos obtenido de una variedad de fuentes y puesto a la disposición de los usuarios finales de manera que ellos puedan entender y usar en el contexto del negocio”.

Sin embargo, la definición más difundida y aceptada de un Data Warehouse pertenece a William H. Inmon quien fue el primero en acuñar el término. “Un Data Warehouse es un conjunto de datos orientados hacia una materia, integrados, no transitorios y que varían con el tiempo, los cuales apoyan el proceso de toma de decisiones de una administración” [Inmon96, p33] Inmon es considerado el padre del Data Warehouse y a escrito una serie de libros junto a otros colaboradores donde abordan ampliamente los diferentes tópicos y niveles de la arquitectura Data Warehouse.

Otros autores como Rob Mattison en [Mattison96] quien identifica cuatro características de un Data Warehouse –estar organizado como un área de almacén de datos neutral, es usado para “data-mining” y otras aplicaciones, llena un conjunto específico de requerimientos de negocio y es usado para satisfacer un conjunto predefinido de criterios de negocio– en su fondo coinciden plenamente con lo expresado por Inmon, en su ya clásica y comúnmente usada definición.

La tecnología de Data Warehouse, define en sí misma un nuevo entorno de trabajo en el que hay profusión de nuevos términos, los cuales hacen referencia a los conceptos que sirven de fundamento al Data Warehouse, por ello es sumamente necesario tener en claro el significado de cada uno de ellos, para poder desenvolverse con soltura y propiedad en el ambiente de los Data Warehouse.

2. Características del Data Warehouse

Al analizar un poco cada término utilizado en la definición de Data Warehouse de Inmon, nos indican claramente que un Data Warehouse se caracteriza por ser:

- **Integrado:** los datos almacenados en el Data Warehouse deben integrarse en una estructura que elimine las inconsistencias existentes entre los diversos sistemas operacionales, que sirven de fuentes de datos. La información suele estructurarse

también en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.

- **Temático:** sólo los datos necesarios para generar información del negocio se integran desde el entorno operacional. Los datos se organizan por temas, no por aplicación, así se facilita el acceso y entendimiento por parte de los usuarios finales a los datos contenidos en el Data Warehouse. Esta característica permite realizar análisis y minería de datos (data mining).
- **Histórico:** el tiempo es parte implícita de la información contenida 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, el Data Warehouse se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones y análisis de tendencias, entre otras cosas.
- **No volátil:** el Data Warehouse se construye para ser leído, y no modificado. La información que existe en un Data Warehouse es permanente, su actualización consiste en la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin alteración a los datos que ya existían.

Además se puede agregar que un Data Warehouse tiene otras características, quizá menos trascendentales, pero que lo diferencian de otras tecnologías:

- **Contiene datos diversos:** es un repositorio unificado de información. Los datos de toda la organización aunque pertenezcan a aplicaciones disímiles son integrados en el Data Warehouse
- **Optimizado para la consulta masiva:** el diseño físico tiene un objetivo distinto al de las bases de datos transaccionales, mejorar los tiempos de respuesta de consultas masivas de información, sin perder de vista la orientación temática y la facilidad de entendimiento que se debe brindar al usuario final
- **La interfaz de usuario esta dirigida a los ejecutivos:** las aplicaciones que se construyen sobre un Data Warehouse y que sirven para acceder a su información, en su mayoría serán utilizadas por analistas de decisiones y altos ejecutivos que descubran las bondades del Data Warehouse, por lo que deben ser amigables e intuitivas.
- **Gran volumen:** Cuando se habla de Data Warehouse el espacio de almacenamiento se mide en Gigabytes y Terabytes, como consecuencia de la información sumariada –el Data Warehouse admite redundancia– y de el largo tiempo de vida de la información –usualmente entre 5 y 10 años–.

Todo lo expuesto presenta las características sobresalientes del nuevo entorno analítico-informático que nace producto de la incorporación de la tecnología de Data Warehousing en una organización y que lo distingue del entorno operativo –los sistemas transaccionales. También se puede caracterizar un Data Warehouse haciendo un contraste de cómo los datos de un negocio almacenados en un Data Warehouse, difieren de los datos operacionales usados por las aplicaciones de producción. [INEI97]

La siguiente figura muestra las diferencias fundamentales entre el ambiente operacional y el Data Warehouse que se han discutido en esta sección.

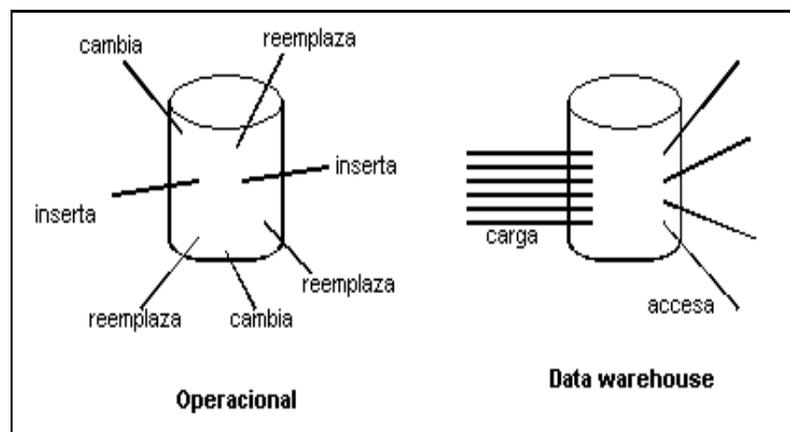


Ilustración I-1 Diferencias entre Sistemas Operacionales y Data Warehouse

Estas diferencias representan un cambio radical en cuanto el enfoque que se le da a los sistemas de información y al rol que desempeña el equipo de tecnología de la información dentro de una organización. Ellas se encuentran resumidas en la tabla IV-1

Las ventajas que presenta la existencia de un Data Warehouse no pueden ser conocidas a priori (en avance), es decir antes de tener lista la primera etapa de implementación porque un Data Warehouse es usado en una forma completamente distinta a otros datos y sistemas construidos por los sistemas de información. Esto hace que no sea posible aplicar análisis del tipo “Retorno a la Inversión” (ROI) para justificar la inversión en un Data Warehouse. [Inmon96, p67]

Tabla I-1.- Diferencias entre Data Warehouse y las Bases de Datos Estándar

Base de Datos Estándar	Data Warehouse
En su mayor parte son actualizaciones	En su mayor parte son lecturas
Muchas transacciones pequeñas	Las consultas son largas y complejas
Mb – Gb de Datos	Gb – Tb de datos
Snapshots Actuales	Historia
Datos Crudos	Datos resumizados y consolidados
Los usuarios son oficinistas	Tomadores de decisiones, los usuarios son analistas

Por sobre todo, las bondades del Data Warehouse no son cuantificables de manera cuantitativa sino cualitativa -¿Cuánto vale una buena decisión, oportuna y a tiempo?- Los beneficios que brinda un Data Warehouse derivan directamente de los nuevos usos de la la información existente o de la disponibilidad de información que difiere en sustancia, calidad o detalle de la información que se tienen disponible y se usa en ese momento. Es a partir de esa información que las ventajas competitivas fluyen [Porter85].

Otras ventajas más tangibles son:

- Alto desempeño de las consultas, el diseño del Data Warehouse lo permite.
- La información se encuentra disponible todo el tiempo, aun si los sistemas fuentes no lo están.
- El procesamiento local en los sistemas fuentes no se afecta, todos los análisis necesarios para soporte a la toma de decisiones se trasladan hacia el Data Warehouse y este se actualiza cuando los sistemas operacionales están fuera de línea, sin ocasionar interrupciones innecesarias.
- Información extra en el Data Warehouse. Mucha información ya se calculó en avance, existen muchos datos que ya se encuentran resumizados. La información que más se necesita ya está lista cuando se requiere.
- Consultas sólo del ámbito del Data Warehouse. En el Data Warehouse se puede realizar consultas que fuera de él sería imposible de conseguir, por restricciones de recursos y tiempo.

B. Metodologías más utilizadas para construir un Data Warehouse

“Cualquiera que construya un Data Warehouse necesita consultar referencias adicionales de una mayor profundidad. Yo sugiero la lectura de Building the Data Warehouse de Bill Inmon, The Data Warehouse Toolkit de Ralph Kimball y Data Warehouse: From Architecture to Implementation de Barry Devlin” [Porter02]. Así se expresa, acerca de quienes son los principales exponentes de las metodologías Data Warehouse, John Porter quien es el Administrador de Datos y Patrocinador del proyecto de Data Warehouse de la Universidad del Estado de Arizona, uno de los primeros Data Warehouse empresariales en la Educación Superior. La mayoría de los expertos coinciden con Porter en que es conveniente para todo individuo que se inicia en el estudio de Data Warehousing, iniciarse en esta tecnología con el conocimiento que proporciona el enfoque de Bill Inmon, continuar con las instrucciones y herramientas de Ralph Kimball e incluir también las enseñanzas de Barry Devlin.

Bill Inmon fue el primero en definir y defender el concepto de Data Warehouse, por lo cual se le conoce como el padre del Data Warehouse, además de proponer estructuras como el ODS o la Fábrica de Información Corporativa (Corporate Information Factory) y de una estrategia corporativa para la construcción de Data Warehouses. Ralph Kimball, por su lado merecidamente puede ser llamado con el título de padre de la Inteligencia de Negocio (Business Intelligence), pues él ha codificado el esquema estrella y las estructuras de datos llamadas copo de nieve, ha definido los conceptos de Data Mart, jerarquías de dimensiones, métricas base y agrupadas, drilling, entre otras y en corto plazo ha desarrollado la ciencia detrás de las modernas herramientas de reportes analíticos. Ambos han hecho inmensurables contribuciones al campo del Data Warehousing. [Lee02]. Barry Devlin, fue pionero al proclamar la construcción de sistemas inteligentes para soporte a la toma de decisiones, de una forma más orientada al análisis, a los que denominó Information Warehouse, y que en realidad es un Data Warehouse, esto ocurrió desde los años ochenta, aún cuando su libro fuera publicado hasta en 1997, varios años después de los libros de Inmon y Kimball, Devlin sigue siendo un innovador en sus enseñanzas.

Un enfoque exitoso de Data Warehousing debe contemplar mucho más que el proceso de diseño. Las situaciones envuelven la arquitectura de datos del Data Warehouse, la cultura organizacional, los roles y responsabilidades cambiantes, la calidad de datos, integración de datos y demás, por lo que es imperativo reconocer que no existe un único método aplicable a toda empresa u organización pues cada una es única, si no más bien

conocer y comprender cuales son las filosofías y metodologías más exitosas para utilizarlas y obtener mayores beneficios, en lugar de enfocarse en una sola y disminuir la perspectiva.

A continuación se abordan ligeramente estas tres metodologías ya mencionadas.

1. Las Enseñanzas de Barry Devlin

Barry Devlin es uno de los gurús del Data Warehouse. Él plantea en su libro [Devlin97] una metodología para la construcción de un Data Warehouse que se propone satisfacer la necesidad de un enfoque generalizado e implementación racional que necesitan las empresas en la actualidad. Su propuesta es una metodología que trata de cubrir por entero todo el proceso de implementación, en ella resaltan los elementos que se discutirán luego.

Sin embargo, su metodología no está claramente definida paso a paso como lo hace W. Inmon o Ralph Kimball. Devlin en cambio parte de una base de conceptos que se abordan a continuación:

Arquitectura de datos conceptual en tres capas

Un atributo importante de la metodología de Devlin es la exposición de tres distintas arquitecturas de datos para los datos en el Data Warehouse, cada una con una, dos y tres capas respectivamente.

Devlin concluye que la opción ganadora, para la mayoría de los proyectos, es la Arquitectura de datos conceptual en tres capas [Devlin97, p63], sin embargo acepta que la arquitectura de dos capas es muy difundida, ya que se adapta particularmente bien a los ambientes computacionales bastante homogéneos que tienen un número relativamente pequeño de plataformas de hardware y software y a aquellos que utilizan el enfoque de Data Marts, propuesto por diversos autores, para la construcción del Data Warehouse.

1. *Arquitectura de una capa:* Cada elemento de dato se almacena una sola vez, por lo que todos los datos se gestionan como si fueran de tiempo real, incluso los datos derivados.

2. *Arquitectura de dos capas:* La capa más baja es utilizada por los sistemas operacionales, son datos en tiempo real. La capa superior es utilizada por los sistemas informacionales, son datos derivados [Devlin97, p67].
3. *Arquitectura de tres capas:* La primera capa es la de datos reales y la segunda capa de la arquitectura anterior se subdivide en dos: la capa de datos reconciliados y la de datos derivados. En este punto la capa de datos reconciliados es una base de datos normalizada que constituye la realización física del modelo de datos corporativo [Devlin97, p70]. Esta arquitectura es recomendada en todas las situaciones donde los negocios requieren una vista empresarial de los datos. En esta arquitectura Devlin denomina Business Data Warehouse (BDW) a la segunda capa y utiliza el término Business Information Warehouse (BIW) para referirse a la capa de datos derivados.

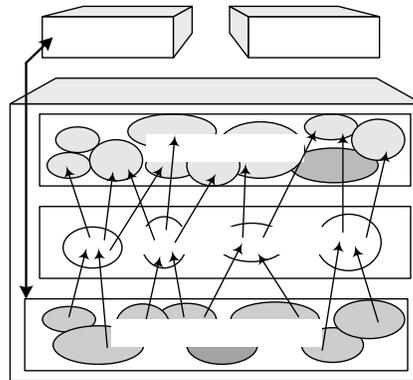


Ilustración I-2. Arquitectura de tres capas

Replicación de Datos

Devlin la define como un conjunto de técnicas que proveen soporte, de una forma comprensiva, para copiar y transformar los datos de la localidad fuente hacia la destino de una manera consistente, repetible y bien entendida, y que consta de los siguientes tres componentes de ejecución claves:

- **Captura:** es un componente de la replicación de datos que interactúa con una fuente de datos almacenados para obtener una copia de algunos o todos los datos, o un registro de los cambios que han ocurrido en el almacén.

- **Transformación:** es el componente que convierte datos entre diferentes estructuras lógicas o físicas de acuerdo a reglas predefinidas. Este componente acepta los datos en el formato de la fuente desde el componente de captura y lo cambia al formato que el componente de aplicación utilizará en el destino.
- **Aplicación:** es un componente que aplica los datos capturados y transformados al almacén destino de los datos.

1.1. Creación del Activo de Datos

Este primer apartado se refiere a utilizar los datos existentes para crear un activo del activo de datos para la organización- un activo que debe ser administrado y protegido, razón por la cual existe el BDW. En la concepción de Devlin el BDW es rigurosamente definido y en mayor parte centralizado que el BIW.

Existen dos etapas que deben ser tomadas en consideración:

Diseño del BDW

Según el enfoque de Devlin, el BDW debe ser diseñado con considerable cuidado, debido a que es un componente central del Data Warehouse. La etapa de diseño puede además subdividirse en las siguientes partes:

- I. **Modelado del BDW:** este proceso consta de dos grandes iteraciones.
 - A. *Definición del alcance del modelo entidad relación:* reducir el modelo entidad relación de la organización a solamente aquellos aspectos que se aplican a los sistemas informacionales. Se debe enfatizar en los aspectos relacionados con el tiempo, los cuales permiten el mantenimiento de datos periódicos.
 - B. *Diseño físico del BDW:* En esta segunda etapa existen tres áreas de particular interés.
 1. *Mapeo hacia los sistemas operacionales:* Establece una correspondencia entre los sistemas fuentes de datos y el BDW y resuelve los problemas de inexistencia de modelos de datos en los sistemas operacionales.
 2. *Restricciones de fuentes de datos:* Se refiere a la disponibilidad física de los datos en los sistemas fuente, algunos datos ya no se encuentran disponibles u otros deberían estar unificados pero no existen elementos que permitan relacionarlos.

3. *Restricciones de desempeño:* Se refiere a la discordia entre desempeño y volúmenes de datos, ambos factores son influenciados por el hecho que la población del Data Warehouse opera en fuentes de datos fijas y con significativas restricciones operativas; y porque la población del Data Warehouse en sí misma tienen restricciones operativas similares.
- II. **Estructura de Datos Periódicos:** cada instancia de una entidad del negocio, es representada por un conjunto de registros, cada uno de los cuales tiene la misma clave pero que se diferencian por las marcas de tiempo utilizadas como parte de la llave. Se debe crear una línea guía de entre los diferentes esquemas de marca de tiempo para maximizar la eficiencia de las mismas.
 - III. **Archivo y recuperación:** este proceso debe diseñar una función que pueda proveer almacenamiento manejable y eficiente y recuperación de datos entre los diferentes niveles en la jerarquía de almacenaje.

Población del BDW

La población del BDW ha sido considerada una de las partes técnicamente más demandantes de la construcción del Data Warehouse, lo que hace necesario diseñar un proceso de obtención de datos que permita que los datos fluyan sin problemas y de forma confiable desde los sistemas operacionales a el BDW.

- I. **Captura-** de datos desde los sistemas operacionales. Se debe definir la captura para la carga del BDW tomando en cuenta los requerimientos del sistema en cuanto a desempeño y la necesidad de transformar datos transitorios o semi- periódicos a datos periódicos.
- II. **Transformación-** de datos entre los dos ambientes. Existen una serie de funciones de transformación de las que deben seleccionarse las que más satisfagan las necesidades del negocio con el objetivo de normalizar y alinear los datos entrantes al Modelo de Datos de la Organización.
- III. **Aplicación-** Existen 4 modos de aplicación de datos y debido a sus características unos u otros pueden ser usados en la creación del BDW o en su mantenimiento, ambas situaciones requieren diferentes modos de aplicación de los datos.

1.2. Liberación del activo de datos para el usuario final

En este apartado Devlin describe como el Warehouse soporta al usuario final, brindándole información en lugar de datos y como la empresa alcanzará beneficios a través

del uso del Business Information Warehouses (BIWS). Los desarrolladores necesitan estructurar los BIW para permitir el acceso de los usuarios tan lejos como sea posible pues en su vasta mayoría dicho acceso de parte de los usuarios ocurrirá en este nivel [Devlin97, p245].

Diseño de BIW

El diseño de BIW debe proveer un buen desempeño sobre un rango de las formas más probables en que los usuarios utilizarán los datos en el BIW, puesto que el comportamiento de los usuarios es impredecible y no se puede conocer en avance.

En general, los BIW son sustancialmente denormalizados en relación al BDW, esto con el objetivo de incrementar el desempeño de las consultas y para reducir la complejidad de los datos desde el punto de vista de los usuarios [Devlin97, p231]. El diseño de BIWS tiene mucho en común con el diseño del BDW en cuanto a las partes en las que se divide, pero se diferencia en los enfoques con que se aborda algunas partes:

- I. **Modelado de BIW:** la justificación para modelar BIW proviene de la audiencia del mismo, si ésta es pequeña el modelado no es apropiado pues no se justifica el esfuerzo, si por otro lado el BIW servirá a todo un departamento o serán datos de uso público el modelado de datos es vital para asegurar la consistencia de uso y entendimiento de los datos.
- II. **Datos Históricos en el BIW:** conjuntos de datos históricos deben ser generados según se requiera para necesidades analíticas específicas y retenidos sólo por los períodos de tiempo requeridos.
- III. **Archivo y Recuperación:** consideraciones similares al diseño de BDW.

Población de BIW

Como en la población del BDW, la función genérica usada para poblar los BIW es la replicación de datos en sus tres pasos previamente identificados [Devlin97, p247]:

- I. **Captura-** de datos desde el BDW. Los BIW en su mayoría utilizan datos provenientes de más de una tabla del BDW, por lo que es necesaria la sincronía entre dichas tablas del antes de comenzar con la población del BIW.
- II. **Transformación-** de datos entre los dos ambientes. Se limita a aquellas funciones que utilizan primordialmente el estándar SQL para realizar las transformaciones entre las

que se destacan: agregación y enriquecimiento, las cuales envuelven cálculos sobre los datos.

- III. **Aplicación-** Cualquiera de los cuatro modos de aplicación de datos pueden utilizarse, en dependencia de las circunstancias. En el BIW co-existen snapshots y datos periódicos los cuales deben ser tratados de forma distinta por el componente de aplicación.

Acceso del Usuario a la Información

Cualquier aplicación que permita al usuario acceder y manipular los datos del negocio en el Data Warehouse contiene elementos de una Interfaz de Información del Negocio (Business Information Interface - BII). La BII provee las funciones necesarias para seleccionar los datos, realizar análisis, guardar los resultados y producir reportes en cualquier formato requerido. Este proceso finalmente transforma los datos en información [Devlin97, p261]. La estructura del BII debe ser como sigue:

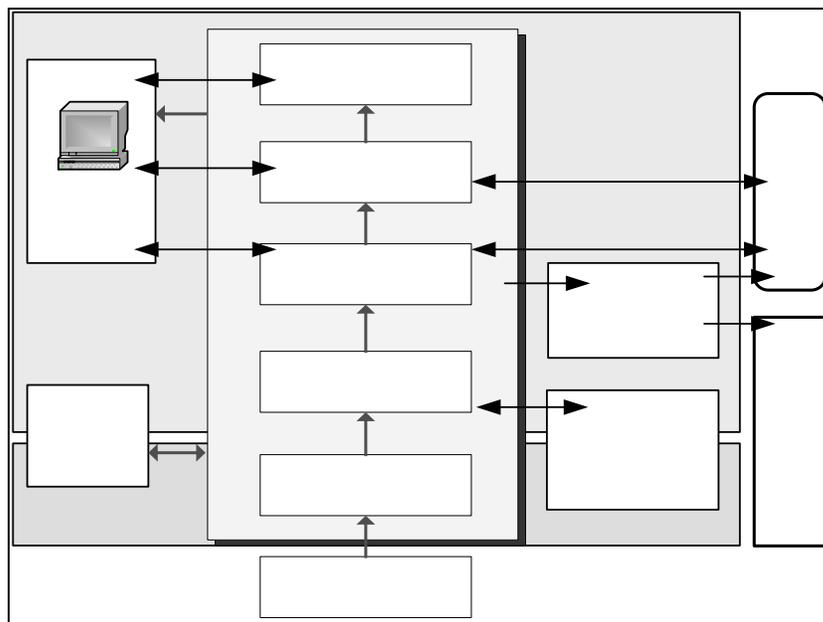


Ilustración I-3. Estructura BII.

Sólo a través de este proceso y de las aplicaciones construidas alrededor del BII, el negocio obtiene valor real a partir de la inversión en un Data Warehouse. El BII debe soportar consultas, manipulación y reporte de datos y también proveer los medios para guardar y re-usar los resultados, tanto a nivel personal o departamental de la organización [Devlin97, p274].

Información - Datos en Contexto

Los Metadatos brindan soporte al usuario final para encontrar datos relevantes y entender el significado de los mismos, por lo tanto el acceso y uso de los metadatos depende en gran parte de la funcionalidad y estructura que estos presenten, lo que convierte la población de metadatos en la mayor dificultad que se les asocia [Devlin97, p275].

El componente de Metadatos se complementa con el BII para lograr liberar el activo de datos a los usuarios finales que estarán en capacidad de percibir y por tanto obtener valores de los datos. Estos componentes proveen el enlace entre los datos fundamentales almacenados y administrados en el Warehouse y la información requerida por el usuario final, juntos transforman los datos en información.

La estructura del componente de Metadatos, denominado BIG (Business Information Guide), es muy análoga a la del BII, incluso se puede decir que el BIG es una forma muy especializada del BII, optimizada para el uso con Metadatos [Devlin97, p296]. La estructura es como sigue:

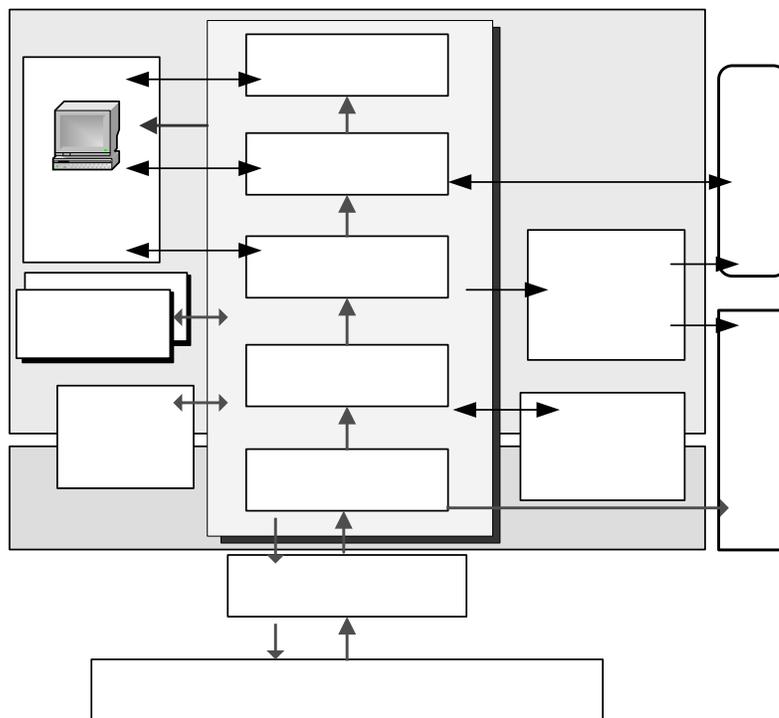


Ilustración I-4. Estructura BIG

1.3. Implementación del Data Warehouse

Este apartado incluye la información que permite estructurar de forma integral el Data Warehouse.

Obstáculos a la Implementación

Aborda los problemas de las áreas claves y su posible resolución:

Tamaño y alcance del Data Warehouse

Según [Devlin97, p304] un verdadero Data Warehouse envuelve a la empresa por completo, esto implica que el tamaño del Data Warehouse es enorme. Con el objetivo de evitar desastres potenciales en la implementación que puedan poner en peligro la realización del proyecto Devlin plantea dividir la implementación completa en piezas alcanzables que puedan ser implementadas en un proceso de etapas. El resultado es un proceso que se expande a varios años pero que entrega valor al negocio en ciclos de seis o doce meses, en paralelo con una implementación progresiva y en etapas de la arquitectura del Data Warehouse.

Colocación del BDW y BIW en la empresa

Esta etapa se refiere a la arquitectura física del Data Warehouse y se deben considerar aspectos de la estructura organizacional, geográfica y de las plataformas físicas usadas por la compañía, decidir si implementar de forma centralizada o distribuida. Esta etapa es altamente dependiente del ambiente de la empresa porque no existe una única forma de implementación.

Planeación de la Implementación

El aspecto más importante de la planeación de la implementación es que este proceso debe armonizar juntos valor para el negocio y la infraestructura de entrega. El resultado es una serie de proyectos cada uno entregando valor al negocio e infraestructura de una manera controlada y en etapas. La prioridad de la secuencia de entregas en este proceso es determinada por una combinación de las necesidades del negocio y de aspectos técnicos, siendo la influencia más importante el modelo de datos de la empresa [Devlin97, p333].

Para segmentar el Data Warehouse se debe seguir los siguientes pasos.

1. *Definir el modelo de alto nivel de la empresa:* este paso se realiza una sola vez.
2. *Modelar un subconjunto del BDW:* definir la estructura y las relaciones de las aplicaciones operacionales que son fuente de datos, este es el paso más complejo.
3. *Modelar el primer BIW:* definir su estructura y sus relaciones con el BDW.

Se debe planear una implementación en etapas que entreguen la infraestructura requerida para asegurar el máximo re-uso de código y esfuerzos realizados previamente y una asociación cercana con las aplicaciones informacionales específicas priorizadas. Cada etapa combina tareas relacionadas con la infraestructura genérica y las aplicaciones específicas del negocio, éstas son sus tres mayores componentes [Devlin97, 312]:

1. *Desarrollo de la infraestructura:* las áreas de infraestructura que normalmente se presentan son: implementación del DBMS, población del BDW y del BIW, creación del BIG y componentes de administración.
2. *Modelado Empresarial:* Pasar del modelado lógico al nivel físico de una área definida del negocio.
3. *Aplicación del Negocio:* La parte del trabajo que debe ser hecha para soportar los requerimientos de negocios específicos de los usuarios, incluye: obtención de requerimientos para el uso de información, modelado del BIW a nivel lógico y físico, documentación de los términos del negocio, definiciones de datos y demás, para la captura y uso del catálogo del Data Warehouse .

Todo proceso de implementación de un Data Warehouse debe iniciar con un proyecto piloto diseñado para probar la viabilidad de todo el enfoque.

Estructura Física del Data Warehouse

El diseño lógico de la arquitectura de datos de tres capas lleva a un conjunto de opciones muy flexibles y poderosas para el diseño físico del Data Warehouse. Estas opciones van desde una decisión basada en una implementación centralizada hasta a un enfoque completamente distribuido basado en una mezcla de plataformas de distinto tamaño [Devlin97, p378].

Hay dos formas de dividir el BDW en subconjuntos:

Segmentación del BDW: Segmentar el BDW en conjuntos de tablas relacionadas muy estrechamente, donde cada conjunto corresponde a áreas del modelo de datos que muestra muchas relaciones internas pero pocas externas.

Partición del BDW: el cual puede ser dentro del segmento consiste en particionar los datos en el segmento de acuerdo a conjuntos de valores de datos para que reflejen distintas áreas del negocio o de desempeño según el nivel que ocupen.

Al considerar el modelado de datos, los requerimientos de uso de los datos y sus restricciones y el tráfico en la red, los diseñadores del Warehouse pueden determinar una óptima distribución física.

Administración del Data Warehouse

De los componentes de administración del Data Warehouse, es importante prestar atención especial al componente de administración de replicación, utilizado para definir como opera la función de replicación de datos, mantiene los metadatos- fuente, destino y mapeo de funciones – requeridos para la replicación.

También hay dos componentes más que son utilizados para soportar las principales actividades de crear mantener y utilizar el Data Warehouse:

- Administración de procesos: coordina las funciones que operan entre ambientes y cuyas acciones deben ocurrir en una secuencia particular.
- Transferencia de Datos: es responsable del movimiento de datos a todos los niveles de granularidad, desde un registro solo a un archivo completo, entre sistemas físicamente separados.

2. El enfoque de W. H. Inmon

William H. Inmon es considerado el padre del Data Warehouse, pues fue el primero en acuñar el término, además su libro con el cual dio a conocer su teoría sobre el Data Warehouse, es hoy en día, una década después de su primera publicación, una de las referencias más completas y utilizadas por los profesionales que ingresan al mundo de la tecnología del Data Warehouse. Él ha escrito más de 32 libros solo o en compañía, con lo que ha consolidado una serie de obras que abarcan cada aspecto del diseño, construcción y gestión de un Data Warehouse.

Inmon presenta dos descripciones del camino a seguir para la construcción de un Data Warehouse. Un plan de migración y una metodología. Según describe el propio William Inmon en [Inmon96, p288] la metodología que él propone difiere del plan de migración en varias maneras. El plan de migración describe actividades generales dinámicamente. La metodología describe actividades específicas, los resultados de esas actividades y el orden de las actividades, pero las dinámicas iterativas de crear un Data Warehouse no son descritas. En otras palabras el plan de migración describe un plan impreciso en tres dimensiones, mientras la metodología describe un plan detallado en una dimensión. Juntos ellos forman un retrato completo de lo que se requiere para construir el Data Warehouse.

2.1. Data-Driven Development Methodology

Uno de los aspectos sobresalientes de la metodología dirigida por los datos es que ella construye sobre esfuerzos previos- construye sobre ambos código y procesos que han sido desarrollados con anterioridad. La única forma en que el desarrollo sobre esfuerzos previos puede ser alcanzado es a través del reconocimiento de aspectos comunes. Antes que el desarrollador inicie su trabajo, él o ella necesita saber que es lo que realmente existe y como eso afecta el proceso de desarrollo. Esta es una de las esencias de la metodología dirigida por los datos.

Una metodología dirigida por los datos tiene cuando menos dos características distintivas [Inmon96, p290]:

1. Se enfoca en utilizar el código y los datos que han sido elaborados previamente como base para la nueva construcción en lugar de construir alrededor de ellos. Para realizar esto, es imperativo el reconocimiento de aspectos comunes en los datos y el procesamiento, la clave para este reconocimiento es el modelo de datos.
2. Hay un énfasis en el almacén central de datos –El Data Warehouse- como la base para el procesamiento de soporte a la toma de decisiones, lo que reconoce que el procesamiento de soporte a la toma de decisiones tiene un ciclo de vida de desarrollo sumamente diferente a los sistemas operacionales.

El desarrollo de sistemas operacionales se forma alrededor de un ciclo de vida de desarrollo que inicia con los requerimientos y termina con el código, lo que se diferencia grandemente del procesamiento de soporte a la toma de decisiones, el cual inicia con los datos y termina con los requerimientos.

La metodología propuesta por Inmon expone los resultados que se deben obtener y describe el orden en que las cosas deben ser hechas. Cómo son alcanzados los resultados requeridos se deja enteramente al criterio del desarrollador. La metodología la dividió en tres partes componentes, incluye las actividades clásicas del desarrollo de sistemas operacionales para que sea una metodología dirigida por los datos [Inmon96, p288].

A continuación la esencia de la metodología propuesta por Inmon en [Inmon96, p317], Apéndice:

METH-1: Desarrollo de Sistemas Operacionales

- M1. *Actividades Iniciales del Proyecto*: Obtención de los requerimientos preliminares del sistema a través de Entrevistas, recopilación de datos, Diseño de Aplicaciones Conjuntas (JAD), análisis del plan estratégico de negocios, requerimientos de parte de los sistemas existentes para los nuevos sistemas.
- M2. *Uso de código y datos existentes*: Usar cuanto código y datos sean posibles y preparar para futuros proyectos que usaran código y datos desarrollados en el proyecto actual.
- M3. *Determinación de tamaño y fases*: Luego de la obtención de los requerimientos generales se deben de determinar su tamaño y dividir el desarrollo del proyecto en fases que funcionen como unidades pequeñas y manejables.
- M4. *Formalización de los requerimientos*: asegurar que los requerimientos sean completos, organizados, leíbles, comprensibles y a un nivel de detallan que permita ser efectivos, que no sean discordantes o se traslapen. Se debe separar los requerimientos operacionales de los de soporte a la toma de decisiones.

Modelado de Datos

- D1. *Diagrama Entidad Relación*: De la especificación formal de los requerimientos resulta la necesidad de identificar las áreas de mayor interés que constituirán el sistema y las relaciones y cardinalidad entre ellas.
- D2. *Conjunto de Elementos de Datos*: Cada tema es dividido (en términos de nivel de detalle) en conjunto de elementos de datos (CED). Los CED contiene los atributos

- de los datos, agrupamiento de los atributos, llaves, tipos de datos, conectores y agrupamiento secundario de los datos. Solo los datos primitivos se manejan aquí.
- D3. *Análisis de Desempeño*: Se resuelve el asunto de la denormalización física de los datos que permita un ingreso y actualización eficientes, para el caso de grandes cantidades de datos o de procesamiento.
- D4. *Diseño físico de la Base de Datos*: Obtención de las tablas y bases de datos diseñadas físicamente, luego de transformar todas las consideraciones lógicas de diseño de datos, desempeño, actualización, ingreso, disponibilidad, etc.

Especificaciones de Proceso

- P1. *Descomposición Funcional*: Resulta de tomar todas las funciones amplias a ser alcanzadas por el sistema y dividir las en una serie de funciones pequeñas sucesivas. Es la descripción de todas las actividades a ser realizadas durante el desarrollo desde un nivel alto hasta un nivel bajo.
- P2. *Nivel de Contexto 0*. Corresponde a D1, Diagrama Entidad-Relación, en la especificación de proceso.
- P3. *Nivel de Contexto 1-n*: Los niveles restantes de la descomposición funcional describen más detalladamente las actividades que ocurren, de manera ordenada, organizada, completa y en concordancia con el flujo de actividades.
- P4. *Diagramas de Flujo de Datos (DFD)*: Existe un DFD para cada nivel de contexto n, indica la entrada de un proceso, la salida del proceso, el almacenamiento de datos necesario para establecer el proceso y una breve descripción del proceso.
- P5. *Especificaciones Algorítmicas, Análisis de Desempeño*: Es el bosquejo del procesamiento actual paso por paso. Los procesos de cada DFD se dividen en especificaciones algorítmicas detalladas, tomando en cuenta los aspectos de desempeño que deben ser resueltos en el diseño de los programas.
- P6. *Pseudocódigo*: Los algoritmos y especificaciones de programas se refinan en pseudocódigo, el cual debe incluir completitud, orden de ejecución, todos los casos requeridos, todas las contingencias (manejo de errores, condiciones de excepción), estructura de la codificación. Corresponde a D4 en modelado de datos.
- P7. *Codificación*: Construcción del código fuente. La traducción completa y eficiente de pseudocódigo en código, incluyendo la documentación en código.
- P8. *Caminata*: Explicación verbal del código a los colegas, para encontrar y corregir la mayor cantidad posible de errores antes de las pruebas.

P9. *Compilación*: El código fuente es compilado y se corrigen todos los errores encontrados.

P10. *Pruebas unitarias*: Pruebas del código a diferentes niveles.

P11. *Implementación*: Esta etapa incluye en sí muchas actividades como:

- Descarga inicial de datos
- Conversiones de datos
- Establecimiento de utilidades de monitoreo
- Escritura de la documentación
- Establecimiento de procedimientos de recuperación y respaldo.

METH-2: Desarrollo Del Data Warehouse

Este es el componente de la metodología que se ocupa del desarrollo de sistemas y procesamiento de soporte a la toma de decisiones.

DSS1. *Análisis del Modelo de Datos*: Confirmación que el modelo de datos de la organización es sólido y que contiene la identificación de los temas de mayor interés, cada tema tiene separada su propia definición de datos: subtipos de datos, atributos, relaciones claramente definidas, identificación de llaves entre otros.

DSS2. *Análisis Breadbox*: Permite la determinación del tamaño –estimación bruta- del entorno de los sistemas de soporte a la toma de decisiones. Simplemente proyecta, en términos crudos, que cantidad de datos mantendrá el Data Warehouse. Con esto es posible determinar si será necesario considerar múltiples niveles de granularidad.

DSS3. *Valoración Técnica*: Contiene definiciones técnicas que tienen la habilidad de manejar grandes cantidades de datos, permitir que los datos sean ingresados flexiblemente, organizar los datos de acuerdo al modelo de datos, recibir y enviar datos a una amplia variedad de tecnologías, contemplar descargas de datos masivas, acceder conjunto de datos simultáneamente o registro por registro.

DSS4. *Preparación del entorno técnico*: Instalación, ubicación y desarrollo de los componentes técnicos que recibirán los datos: la red, el almacenamiento secundario, su sistema operativo, la interfaz hacia y desde el Data Warehouse el software administrador del Data Warehouse y el Data Warehouse en sí.

DSS5. *Análisis de los temas del Data Warehouse*: Determinación del tema que será el primero en implementarse (poblarse). Debe ser grande lo suficiente para tener sentido y pequeño lo suficiente para permitir una rápida implementación.

DSS6. *Diseño del Data Warehouse*: Algunas de las características del diseño incluyen:

- Acomodación de los diferentes niveles de granularidad, si existen.
- Orientación de los datos a los principales temas de la organización.
- La presencia de solo datos primitivos y datos derivados públicamente.
- La ausencia de datos que no apoyan los sistemas de soporte a las decisiones.
- Variabilidad de tiempo en cada registro de datos.
- Denormalización física de los datos donde sea aplicable
- Adaptación de los datos del entorno operacional al analítico- Data Warehouse.

DSS7. *Análisis de los Sistemas Fuente*: Identificación del sistema de registro, es decir el mapeo de los datos del ambiente operacional al ambiente analítico. Se deben resolver los siguientes aspectos relacionados a la integración de los datos:

- La estructura y resolución de llaves al pasar al ambiente analítico.
- Atribuciones: Elección de una fuente de datos entre muchas, que hacer cuando no hay fuente de datos, las transformaciones a realizar.
- Creación de la variabilidad de tiempo en los registros a partir de datos actuales.
- Creación de la estructura analítica a partir de la estructura operacional.
- Reflejo de las relaciones del entorno operacional en el ambiente analítico.

DSS8. *Especificaciones*: Formalización de la interfaz entre los ambientes operacionales y analíticos en términos de especificaciones de programas, que permitan la extracción e integración de los datos lo más eficiente y simple posible. Verifica que datos operacionales deben ser obtenidos y como guardarlos.

DSS9. *Programación*: Programas de transformación que permiten la extracción, integración y ubicación en perspectiva de tiempo de los datos. Incluye todas las actividades estándar de programación como: elaboración de pseudocódigo, codificación, compilación, pruebas. Esto asegura que el código generado sea eficiente, documentado, con capacidad de cambio rápido, eficaz y completo.

DSS10. *Población*: Ejecución de los programas desarrollados en las etapas anteriores. Se deben resolver los aspectos de frecuencia de población, reglas de purga, envejecimiento de los datos poblados, administración de múltiples niveles de granularidad y refrescamiento. Con este paso final se obtiene un Data Warehouse poblado y funcional, accesible y comprensible que sirve las necesidades de la comunidad de los sistemas de soporte a la toma de decisiones.

Meth-3: Procesamiento Heurístico

Esta componente de la metodología de lo que Inmon llama entorno arquitectónico describe el uso del Data Warehouse para propósitos de análisis.

- DEPT1. *Repetición del Desarrollo Estándar*: Para la obtención de reportes estándares, el procesamiento analítico repetitivo debe seguir el procesamiento normal descrito en METH-1, exceptuando el modelado de datos, porque la fuente de datos es el Data Warehouse.
- IND1. *Determinación de los datos necesarios*: Selección de datos para análisis posteriores y uso potencial en la satisfacción de los requerimientos de reportes.
- IND2. *Programas para extraer datos*: Escritura de un programa para acceder y recuperar los datos seleccionados.
- IND3. *Combinar, Fusionar, Analizar*: Edición, combinación con otros datos y refinamiento de los datos obtenidos para que los datos sean utilizables en análisis.
- IND4. *Análisis de datos*: Revisión de los resultados obtenidos para asegurar que satisfagan las necesidades del analista de datos, si no se inicia el proceso iterativo.
- IND5. *Respuesta a la cuestión*: Producción del reporte final luego del proceso iterativo.
- IND6. *Institucionalización*: Si existe necesidad de elaborar un reporte de forma repetitiva, entonces se debe convertir éste en un conjunto de requerimientos a ser satisfechos mediante una operación de ocurrencia regular.

Todos los procesos IND ocurren a nivel individual, pues se realizan a través de procesamiento realmente puros para satisfacer necesidades individuales en la organización. Los IND usualmente deben ser hechos varias veces antes de obtener un resultado satisfactorio, por lo que los programas escritos deben tener como característica esencial su fácil modificación para poder ser ejecutados en numerosas ocasiones.

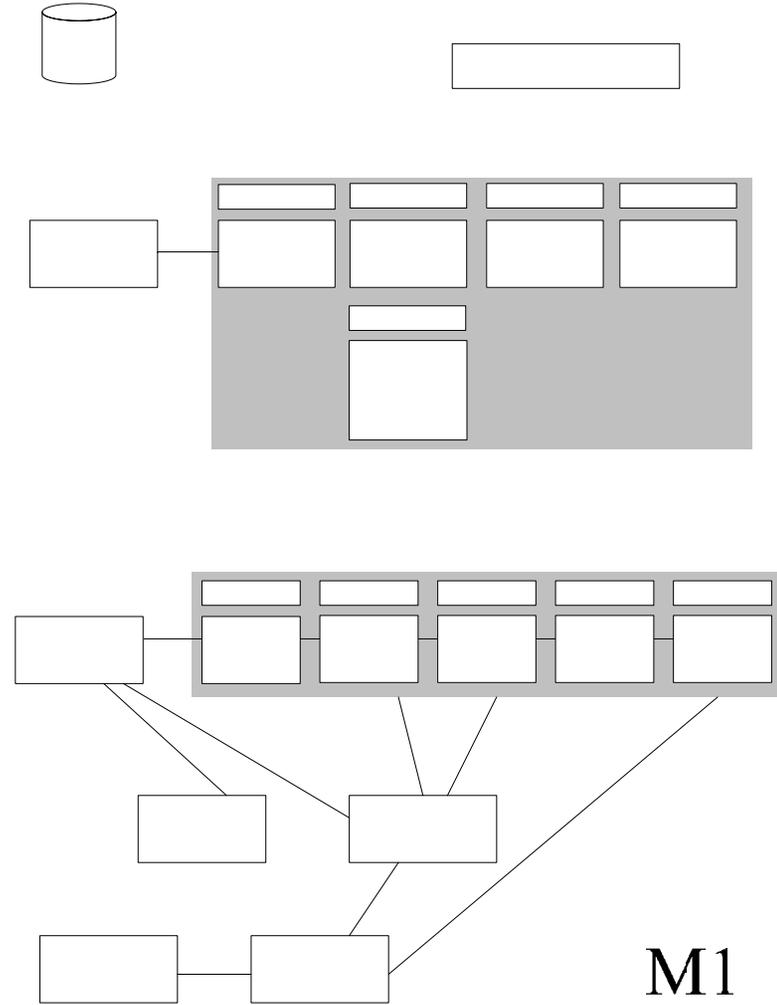
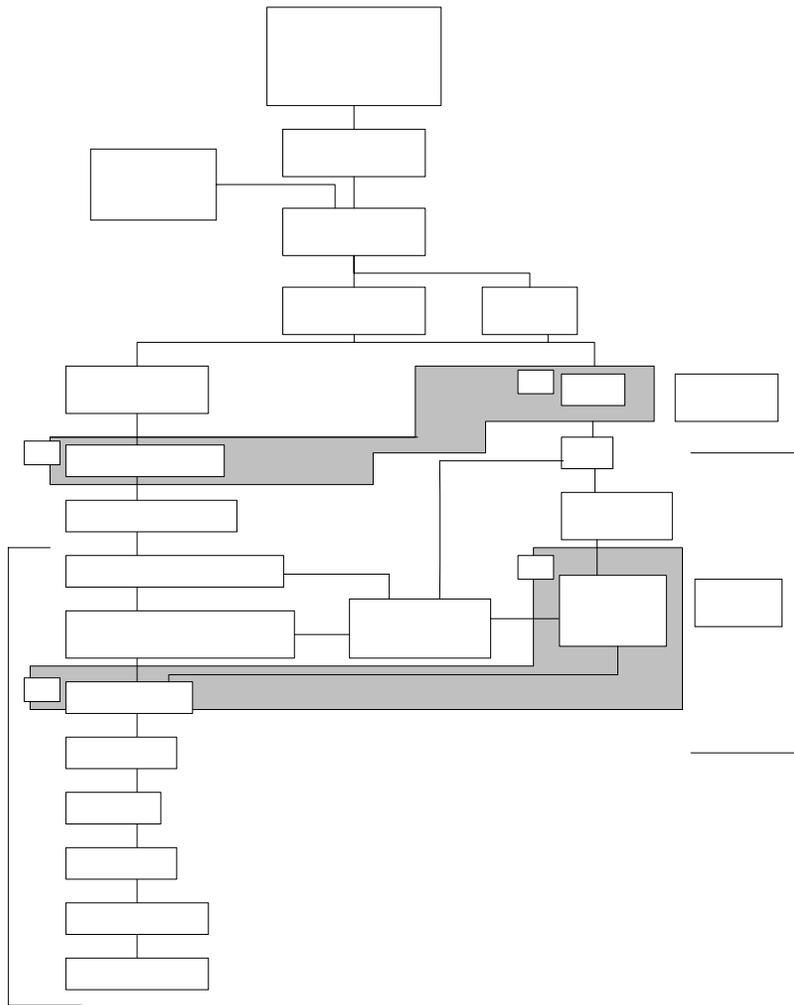


Ilustración I-5. Metodología Data Warehouse de Inmon

M1

Entrevi
Recopi
Sesione
Plan Es
Sistema

S

2.2. Migration Path

El punto inicial del plan de migración es el modelo de datos. El modelo de datos representa las necesidades de información de la organización, no necesariamente lo que realmente tiene.

El modelo de datos debe identificar cuando menos lo siguiente [Inmon96, p276]:

- Los temas de interés principales de la organización
- Las relaciones entre estos temas principales
- El agrupamiento de las llaves y atributos que representen más plenamente los principales temas incluyendo:
 - Atributos
 - Las llaves
 - Los grupos repetitivos de atributos y llaves
 - Conectores entre las áreas de los temas principales
 - Relaciones de subtipos.

Luego que se tiene el modelo de datos en su lugar es necesario definir el sistema de registros, el cual se define en términos de los sistemas existentes que la organización ya tiene. El sistema de registros es la identificación de la mejor información que tiene la organización, determinada en función del modelo de datos.

La determinación de la mejor fuente existente de datos utiliza los siguientes criterios:

- Datos más completos
- Datos más oportunos
- Datos más precisos
- Datos más cercanos a la fuente de entrada
- Datos más cercanos a la estructura del modelo de datos en términos de llaves, de atributos, de agrupamiento de atributos de datos juntos.

El siguiente paso es diseñar el Data Warehouse. Si el modelo de datos está elaborado correctamente el diseño del Data Warehouse sólo requiere que pocos aspectos del modelo se cambien para convertirlo en un diseño de Data Warehouse, principalmente se debe:

- Agregar un elemento de datos a la estructura de llaves, si esta no lo presenta.
- Eliminar los datos puramente operacionales.
- Las relaciones de integridad referencial deben ser cambiadas por otras fabricadas
- Los datos derivados usados frecuentemente deben ser agregados al diseño
- La estructura de datos debe ser alterada cuando sea apropiado para:
 - Agregar arreglos de datos
 - Agregar datos redundantes
 - Futura separación de datos en condiciones adecuadas
 - Mezcla de tablas cuando sea apropiado
 - Realizar análisis de la estabilidad de las necesidades de datos.

Entre las áreas de los temas de interés debe haber muchas tablas separadas, cada una de las cuales es conectada por una llave común.

Después del diseño de datos, el siguiente paso es diseñar y construir las interfaces entre el sistema de registro y el Data Warehouse, las cuales poblarán el Data Warehouse en periodos de tiempo regulares. Las interfaces realizan una serie de actividades vitales de modificación de los datos antes de cargar el Data Warehouse:

- Integración de los datos
- Alteración de las bases de tiempo de los datos
- Condensación de los datos
- Exploración eficiente del ambiente de los sistemas existentes.

La etapa final de este plan de migración es iniciar la población del Data Warehouse con el primer tema de interés, es importante que solo una fracción del Data Warehouse sea poblada inicialmente para asegurar que los inevitables cambios que sobrevienen y que deben ser hechos puedan llevarse a cabo de manera fácil y rápida, lo que garantiza la flexibilidad de la primera iteración del Data Warehouse. Una vez que el usuario final tiene acercamiento a los datos y proporciona retroalimentación al arquitecto de datos, entonces puede ser sano poblar cantidades mayores de datos.

Los procesos de población y retroalimentación continúan indefinidamente, mientras el Data Warehouse continúa su proceso evolutivo incorporando cambios producto de éstos dos procesos. Sin embargo, no se debe olvidar que las actividades y asuntos del

entorno del Data Warehouse son completamente independientes de las actividades y asuntos del entorno de los sistemas operacionales existentes.

El ciclo de Retroalimentación

“En el corazón del éxito en el desarrollo de un Data Warehouse se encuentra el ciclo de retroalimentación entre el arquitecto de datos y el analista de sistemas de soporte a la toma de decisiones” [Inmon96, p283].

El Data Warehouse es poblado desde los sistemas operacionales. El analista de sistemas de soporte a la toma de decisiones usa el Data Warehouse como base para sus análisis. Al encontrar nuevas oportunidades el analista de sistemas de soporte a la toma de decisiones transmite esos requerimientos al arquitecto de datos, quien realiza los ajustes apropiados.

No se debe pasar por alto ninguno de los siguientes aspectos para asegurar una implementación exitosa del ambiente del Data Warehouse [Inmon96, p284]:

1. El analista de sistemas de soporte a la toma de decisiones actual bajo las bases de obtener lo que desea para luego establecer que es lo que realmente necesita.
2. Cuanto más corto sea el ciclo de retroalimentación, más exitoso será el esfuerzo de construcción del Data Warehouse.
3. Cuanto más grande sea el volumen de datos que deba ser cambiado más largo será el ciclo de retroalimentación.

3. El ciclo de Vida del Data Warehouse de Ralph Kimball.

El primer libro de Ralph Kimball que vio luz fue *The Data Warehouse Toolkit* [Kimball95], en el cual este autor mostraba como usar el modelado dimensional para diseñar Data Warehouses usables y efectivos. Sin embargo, la presentación y explicación de una metodología que utilice estas técnicas para la construcción de completos Data Marts y Data Warehouses aparece hasta en una publicación posterior *The Data Warehouse Lifecycle Toolkit* [Kimball98]. La expresión Ciclo de Vida se refiere a todos los pasos del proceso completo de desarrollo de software: planeación, diseño, codificación, prueba, implementación y administración, el ciclo de vida de Kimball es una metodología paso a paso para diseñar, desarrollar y desplegar Data Marts y Data Warehouses, y que es

organización para afrontar dicho proyecto. También se debe elaborar el plan para el proyecto, así como gestionar la puesta en marcha del mismo, definiendo y manteniendo su alcance.

Obtención de Requerimientos

Cada organización es única en sí misma, cada vez que se inicia un Data Warehouse, es imposible conocer en avance los requerimientos de tal instrumento de apoyo a la toma de decisiones, por tanto se debe de hacer uso de entrevistas o sesiones con facilitador para lograr obtener datos de la información necesaria en la empresa para poder definir de manera correcta el contenido y utilización del Data Warehouse.

3.2. Diseño de Datos o Modelado Dimensional

Este es el corazón de la metodología de Kimball, él defiende el modelado dimensional con el argumento de que proporciona grandes ventajas en desempeño e inteligibilidad además de ser siempre aplicable. Aquí se diferencia de los autores ya expuestos pues ellos no se inclinan por ningún tipo de modelado, sino que solo señalan que es imperativo contar con un adecuado modelo a partir del cual fundamentar el Data Warehouse.

El Modelado Dimensional, según su creador Ralph Kimball, es el diseño físico y lógico que transformará las antiguas fuentes de datos en las estructuras finales del Data Warehouse, a través de una técnica que busca la presentación de los datos en un marco de trabajo estándar que es intuitivo y permite un acceso de alto desempeño. Cada modelo dimensional está compuesto de una tabla que tiene una llave compuesta llamada tabla de hechos y un conjunto de tablas más pequeñas llamadas dimensiones. Cada tabla dimensión tiene una llave primaria simple, que corresponde exactamente a una de las partes de la llave compuesta en la tabla de hechos. Esta estructura característica es usualmente llamada esquema estrella [Kimball98, p144].

Los pasos necesarios para convertir un Diagrama Entidad-Relación (ERD) a un conjunto de diagramas de modelado dimensional son [Kimball98, p146]:

1. Separar el ERD de la organización en procesos de negocios discretos y modelar cada uno separadamente.

2. Seleccionar las relaciones muchos a muchos en el modelo entidad-relación que contengan cantidades numéricas y aditivas que no pertenezcan a la llave y designarlos como tablas de hechos.
3. Denormalizar todas las tablas restantes en tablas planas con llaves simples que conectan directamente a las tablas de hechos. Estas tablas se convierten en las dimensiones. En los casos en los que una tabla dimensión se conecte a más de una tabla de hechos, se representa esta misma tabla dimensión en ambos esquemas, a estas tablas dimensión se les llama conformadas entre los dos modelos dimensionales.

El objetivo de aplicar modelado dimensional a una estructura como el Data Warehouse se deriva de las fortalezas de este método [Kimball98, p148], el cual proporciona:

- Un marco de trabajo predecible, que se obtiene al utilizar dimensiones conformadas, esto mejora la presentación y también el desempeño, ya que un joint de N-tablas se traduce en una operación sort-merge sobre las dimensiones que luego se aplican a las tablas de hechos.
- Soporta cambios inesperados en el comportamiento del usuario.
- Es fácilmente extensible para acomodar nuevos elementos de datos inesperados y nuevas decisiones de diseño.

La Arquitectura de Bus del Data Warehouse

“Un Data Warehouse está hecho de las unión de muchos de sus Data Marts” [Kimball98, p19]. Cada Data Mart debe ser representado por un modelo dimensional y dentro de un único Data Warehouse, todos estos Data Marts deben ser construidos a partir de dimensiones conformadas y hechos conformados. Esta es la base de la arquitectura de bus del Data Warehouse.

La planeación del Data Warehouse debe iniciarse con un fase pequeña de arquitectura de datos que cubra todo y que a la vez tenga metas muy finitas y específicas, luego se debe seguir esta arquitectura con una implementación paso a paso de Data Marts separados, donde cada implementación se adhiere a la arquitectura predefinida [Kimball98, p153]. Cada Data Mart para que sea práctico debe ser basado en los datos más granulares (atómicos) que sea posible coleccionar y almacenar.

Cualquier implementador de Data Warehouse que quiera ser exitoso inevitablemente realiza los siguientes pasos:

1. Crear una arquitectura circundante que defina el alcance e implementación del Data Warehouse completo.
2. Supervisar la construcción de cada pieza del Data Warehouse completo.

Sin una adherencia estricta a las dimensiones conformadas un Data Warehouse no puede funcionar como un todo integrado, ya que a través de las dimensiones conformadas se logra [Kimball98, p157]:

1. Una tabla dimensión sola puede ser usada contra múltiples tablas de hechos en el mismo espacio de base de datos
2. Las interfaces de usuario y los datos que contienen son consistentes en cualquier instante en que se utilice la dimensión.
3. Existe una interpretación consistente de los atributos y por ende de los rollups a través de los Data Marts.

Las dimensiones y hechos conformados juegan el rol de un bus en el Data Warehouse. Al definir una interfase estándar de bus para un entorno de Data Warehouse dado, un nuevo Data Mart puede ser introducido en el Data Warehouse de modo que el puede coexistir útilmente con lo que ya se encuentra en su lugar [Kimball98, p164].

Técnicas de Modelado Dimensional

La idea fundamental del modelado dimensional es que casi cada tipo de datos de negocio puede ser representado como un tipo de cubo de datos, donde cada celda del cubo contiene un valor medido y las aristas del cubo definen las dimensiones naturales de los datos [Kimball98, p165].

Los elementos esenciales del modelado dimensional son:

- **Hechos:** Un hecho es algo que usualmente no se conocen en avance. Casi cualquier campo con valor numérico es un hecho, aunque algunos pueden ser valores de texto. Un hecho es una observación del mercado de trabajo.
- **Atributos:** Usualmente son campos de texto que describen las características de algo tangible, como las descripciones de producto.
- **Dimensiones:** Los atributos textuales que describen los objetos y que tienen una alta correlación entre ellos, son organizados dentro de las dimensiones.

Método de Diseño de cuatro pasos para diseñar una tabla de hechos individual

El diseño lógico detallado de un esquema dimensional es dirigido por los siguientes cuatro pasos [Kimball98, p 194]:

1. *Escoger el Data Mart*: En las primera implementaciones es aconsejable mantener la situación lo más simple posible por lo que es una adecuada decisión escoger una única fuente de datos para el Data Mart.
2. *Declarar la Granularidad de la tabla de Hechos*: Consiste en una clara definición de que se considera una tabla de hechos en el diseño dimensional propuesto, que estilo de tabla de datos se utilizara: transacciones individuales, snapshots de alto nivel o elementos individuales de documentos de control como facturas.
3. *Escoger las Dimensiones*: Seleccionar de todas las dimensiones posibles aquellas con la menor granularidad que tomen un valor único en el contexto de un conjunto dado de medidas del negocio.
4. *Escoger los Hechos*: Los hechos deben ser siempre específicos a la granularidad de la tabla de hechos. No se deben mezclar hechos de otros períodos de tiempo u otras agregaciones para hacer los cálculos convenientes.

Construcción de Modelos Dimensionales

Aquí Kimball expone que el conjunto de los modelos dimensionales constituyen el diseño lógico del Data Warehouse, pero es necesario decidir que modelo, es decir que Data Mart, se construirá primero.

Según Kimball se debe comenzar con un enfoque de planeación de arriba-abajo al que llama Matriz de Arquitectura del Data Warehouse, en la cual se nombran todos los Data Marts que sea posible construir junto con las dimensiones implicadas por dichos Data Marts y luego se relacionan ambos para determinar las dimensiones que son utilizadas por más de un Data Mart y que deben ser conformadas para asegurar un Bus de Data Warehouse.

En la siguiente etapa, una vez que ya se han identificado los Data Marts y sus dimensiones asociadas, se procede con los diseños físico y lógico detallados de las tablas individuales, se utiliza el método de cuatro pasos para diseñar cada tabla de hechos.

Luego ya será posible la elaboración del borrador inicial, el cual permitirá rastrear e identificar los hechos base y los hechos derivados que deban ser incluidos, así como solucionar aspectos de hechos duplicados y otros, para adaptar mejor el modelo a las características y exigencias del negocio. En este proceso se realizan los cambios necesarios en el modelo, los cuales deben ser plenamente documentados, los realizados y los pendientes, junto a todas las decisiones que se realicen, desde elegir las fuentes de datos hasta determinar las fórmulas de cálculo de los hechos derivados y demás.

El modelado dimensional para la primera implementación, no el modelo de la empresa, se puede considerar finalizado una vez que se han resuelto los aspectos requeridos para alcanzar el objetivo del negocio, se pueden agrupar en las categorías siguientes [Kimball98, p 307]:

- Elementos críticos de datos.
- Elementos críticos de las políticas del negocio: implican decisión acerca de cómo manejar determinadas situaciones
- Futuro/Fuera del Alcance: son elementos interesantes de tener, pero que no justifican el esfuerzo necesario en la primera implementación

3.3. Arquitectura

Una arquitectura bien definida agrega valor al proyecto, al mejorar la comunicación y planeación. Una arquitectura efectiva producirá un incremento en la flexibilidad del sistema, facilita el aprendizaje a los usuarios y mejora la productividad y el re-uso ya que permite elegir adecuadamente herramientas que automaticen partes de los procesos del Data Warehouse, así como reutilizar procesos existentes en lugar de iniciar de nuevo a partir de la nada [Kimball98, p 318].

El planteamiento de arquitectura al que se refiere Kimball, se resume en la tabla que sigue, llamada marco de trabajo de la arquitectura, donde las columnas muestran las principales áreas de la arquitectura: datos, técnica e infraestructura y las filas representan los niveles de detalle en orden creciente: requerimientos del negocio, modelo de arquitectura, modelo detallado, implementación.

Tabla I-2. Marco de Trabajo de la Arquitectura

Nivel de detalle	Datos (qué)	Técnica (cómo)		Infraestructura (donde)
		Back room	Front room	
Requerimientos de Negocio y Auditoría	Datos que dirigen el negocio	Como se trataran los datos	Formas de análisis de los datos	Hardware y software necesario
Modelos de arquitectura y Documentos	El modelo dimensional	Capacidades específicas para tratar los datos	Servicios para que los usuarios utilicen la información	Fuente-Destino de datos. Capacidad de almacenamiento
Modelos detallados y Especificaciones	El Modelo Físico y Lógico	Que estándares y productos proveen dichas capacidades	Especificaciones de reportes. Usuarios posibles. Frecuencia	Interacción con las capacidades y utilidades del sistema.
Implementación	Creación de la BD, índices, etc. Documentar	Programar, automatizar, documentar	Construir el ambiente de reportes y análisis. Documentar	Instalar y probar los componentes nuevos de la infraestructura

Este marco de trabajo de la arquitectura del Data Warehouse ayuda a dividir en piezas manejables la tarea de construir la arquitectura del Data Warehouse.

Arquitectura de Datos

En el área de arquitectura de datos se incluye el contenido del Warehouse, el diseño físico y lógico de los modelos de datos, agregaciones, jerarquías y más, los cuales ya fueron abordados en el apartado anterior.

Arquitectura Técnica

El área de arquitectura técnica cubre los procesos y herramientas que se aplican a los datos [Kimball98, p323]. En el área técnica existen dos conjuntos que tienen distintos requerimientos, brindan sus propios servicios y componentes de almacenaje de datos, por lo que se consideran cada uno aparte: El back room (habitación trasera) y el front room (habitación frontal). El back room es el responsable de la obtención y preparación de los datos, por lo que también se conoce como adquisición de datos. El front room es responsable de entregar los datos a la comunidad de usuario y también se le conoce como acceso de datos. A pesar de esta separación ambos componentes interactúan de manera significativa.

En la arquitectura técnica es importante conocer que existen principios de diseño que deben presentar todas las herramientas, procesos y componentes de las capas de servicio [Kimball98, p331]:

- **Dirigidos por los Metadatos:** los metadatos proveen parámetros e información (un conjunto de información de control acerca del Data Warehouse, su contenido,

sistemas fuentes y su proceso de carga) que permite que la aplicación desempeñe sus tareas correctamente.

- **Capas de Servicio Flexible:** El servicio de consolidación de datos y el servicio de consulta de datos son capas de aplicación que también proveen cierto nivel de indirección lo cual se agrega a la flexibilidad de la arquitectura.

El back room es donde tiene lugar el proceso de preparación de los datos. Es el cuarto de máquinas del Data Warehouse [Kimball98, p335]. Su principal asunto es resolver los problemas específicos de la migración de datos del punto A al punto B, con las transformaciones apropiadas y en el momento justo. En la arquitectura del Back Room se debe analizar los aspectos relativos a: los sitios donde se almacenan los datos (sistemas fuente, área de conciliación de datos, servidores de presentación), los servicios que proporcionara (extracción, transformación, carga, control del trabajo de conciliación) y la administración de los activos del back room (respaldo y recuperación, archivo, seguridad de la extracción y carga y servicios futuros).

El front room es lo que los usuarios finales miran y utilizan para trabajar día a día. Es la cara del Data Warehouse [Kimball98, p373]. Su principal asunto es construir una capa intermedia entre los usuarios y la información que permita esconder algunas de las complejidades propias de la plataforma técnica y que les ayude a encontrar lo que buscan. En la arquitectura del Front Room se debe analizar los aspectos relativos a:

1. Los sitios donde se almacenan los datos: herramientas de acceso a los almacenes de datos, almacenes de datos para reporte estándar como los Data Mart personales, sistemas de carga a otras aplicaciones.
2. Los servicios que proporcionara para el acceso de datos: navegación, acceso y seguridad, monitoreo activo, administración de consultas, localización de consultas, reportes estándar, servicios de acceso futuro, servicios de escritorio, modelación de aplicaciones y Data mining, implicaciones web y enfoques para la arquitectura de herramientas de escritorio.

Arquitectura de Infraestructura y Metadatos

El área de arquitectura de infraestructura se refiere a las plataformas que soportan los datos y procesos. Es la planta física del Data Warehouse y provee los cimientos para todos los elementos de la arquitectura ya descritos. La Infraestructura incluye el hardware, la red y funciones de bajo nivel que los componentes de más alto nivel

no contemplan. Los metadatos son de un tipo distinto pero proveen el mismo tipo de capa base de soporte para las herramientas del back room y del front room [Kimball98, p411].

Para definir correctamente este componente en la arquitectura del Data Warehouse, siempre se debe tomar en consideración: los requerimientos del negocio, los cuales dictaminan la dirección del Warehouse, la evolución que la infraestructura sufrirá en la puesta en marcha del proyecto y evolución, los factores de la infraestructura del back room (tamaño de datos, volatilidad, cantidad de usuarios, hardware, sistema operativo entre otros), factores de infraestructura del front room (consideraciones de los servidores de aplicación, del escritorio), factores de redes y conectividad (ancho de banda, acceso remoto, gateways, transferencia de archivos, conectividad de la base de datos).

En el aspecto relativo a los metadatos los puntos claves en los que pueden ser agrupados son: metadatos de los sistemas fuente, metadatos de conciliación de datos, metadatos del DBMS, metadatos del front room. Finalmente, debe existir un catálogo de metadatos, lo cual es una herramienta que permita catalogar y darles seguimiento a todos los metadatos.

3.4. Implementación

Antes de iniciar la implementación es importante decidir que hechos deben ser agregados con respecto a que dimensiones con el objetivo de mejorar el desempeño general del Data Warehouse, para ello es necesario que:

- Cada agregación en cada nivel distinto sea guardado en su propia tabla de hechos, separados del resto de los datos atómicos.
- Siempre que sea posible las dimensiones adjuntas a las tablas de hechos deben ser versiones reducidas de las dimensiones asociadas con las tablas de hechos base.
- Las tablas base de hechos atómicos y sus tablas de hechos agregadas deben ser relacionadas como una familia de esquemas (permite que el navegador conozca que tablas están relacionadas entre si).
- Se debe forzar a todas las rutinas SQL creadas por cualquier herramienta de usuario final para acceso a los datos o de aplicación, se refieran exclusivamente a la tabla de hechos base y sus dimensiones de tamaño completo.

Otro elemento que debe ser completado es el diseño físico.

Completar el Diseño Físico

Existen varios aspectos que por ser meramente de implementación no se consideraron en el componente de arquitectura que define los modelos y diseño físico. Por lo tanto una vez iniciada la implementación del Data Warehouse es necesario completarlos. Estos detalles son altamente dependientes de los factores individuales de un proyecto: El modelo lógico de datos, el RDBMS del Warehouse, los volúmenes de datos, los patrones de uso y las herramientas de acceso, sin embargo, es posible determinar un proceso de alto nivel para el diseño físico.

La secuencia básica es iniciar con algo de planeación –desarrollo de estándares de nomenclatura y de la base de datos y una estrategia de seguridad. Luego, construir del modelo físico incluyendo un corte en el tamaño preliminar de la base de datos y de las tasas de crecimiento. Revisar el plan de agregaciones que ya se definió. Una vez que ya se conoce cuales son todas las tablas, se puede hacer un corte para la estrategia preliminar de índices. En este punto, uno puede realmente iniciar la parte física, pues se cuenta con suficiente información acerca de lo que ocurre en la base de datos y cuando, lo necesario para iniciar el diseño y construcción de la instancia de base de datos en sus mínimos detalles, e implementar herramientas de monitoreo de uso [Kimball98, p570].

Consolidación de Datos

El proceso de consolidación de Datos se puede dividir en un plan de diez pasos para la creación de la aplicación de consolidación de datos para el Data Mart [Kimball98, p610]:

Plan:

1. Crear un esquema de alto nivel, muy consolidado, del flujo de fuente a destino.
2. Probar, escoger e implementar la herramienta de consolidación de datos.
3. Drill down por cada tabla destino, bosquejar gráficamente cualquier reestructuración o transformación de datos compleja y el proceso de generación de claves arbitrarias. Desarrollar la secuencia preliminar de trabajo.

Carga de Dimensiones:

4. Construir y probar la carga de una tabla de dimensión estática.
5. Construir y probar el proceso de cambio lento para una dimensión.
6. Construir y probar la carga de las dimensiones faltantes.

Tablas de hechos y automatización

7. Construir y probar la carga de las tablas de hechos históricas.
8. Construir y probar el proceso de carga incremental.
9. Construir y probar la carga de las tablas de agregaciones.
10. Diseñar, construir y probar las aplicaciones de automatización del proceso de conciliación.

Es en esta etapa cuando se debe asegurar la calidad de los datos y su limpieza y cuidar los detalles propios de implementación de cada proyecto. Por todo esto, la etapa de conciliación de datos es una de las piezas más difíciles de un proyecto de Data Warehouse.

Construir las Aplicaciones de Usuario Final

El proceso de construcción de estas aplicaciones se puede dividir en dos apartados: Especificación de la aplicación y Desarrollo de la Aplicación [Kimball98, p665].

Este proceso es estándar y análogo al desarrollo de aplicaciones para los sistemas transaccionales, con la salvedad en que típicamente se basa en la herramienta de acceso de datos seleccionada y que debe brindar soporte a usuarios con requerimientos de análisis improvisados.

3.5. Despliegue y Crecimiento

Un despliegue exitoso de un Data Warehouse requiere planeación consistente y coordinación previa a la culminación de los esfuerzos de desarrollo. Un apropiado mantenimiento y crecimiento evidencian el éxito de llevar a cabo un proyecto importante como es un Data Warehouse, una correcta gestión que ponga en primer lugar satisfacer a los usuarios de negocio, sin sacrificar atención al back room y entorno técnico permite asegurar una adecuada evolución del Data Warehouse si es acompañado de mediciones y rastreo en el Data Warehouse y retroalimentación de parte de los usuarios.

Planeación del Despliegue

“Despliegue es la convergencia de tecnología, datos y aplicaciones en los escritorios de los usuarios de negocios, junto con la necesaria educación y estructura de soporte al usuario” [Kimball98, p691].

El despliegue necesita mucho trabajo de planeación y la coordinación de muchas variables, por lo que su preparación no puede iniciarse hasta última hora, sino con mucha antelación. Esta planeación y preparación permite colocar todas las piezas del Warehouse en su respectivo lugar antes de desplegarlo. Mucho de la arquitectura e infraestructura del back room como el HW, el DBMS, el catálogo de Metadatos y demás serán establecidas mucho antes del despliegue puesto, que se necesitan para las actividades del despliegue mismo. La tecnología que reside en los escritorios de los usuarios es la última pieza que debe ser colocada en su lugar puesto que es el punto de inicio del despliegue en si.

La Planeación del despliegue se puede dividir en varias partes. En una de ellas se determina la disponibilidad de instalación en los escritorios de usuario, en otra es necesario desarrollar una estrategia de educación del usuario final, que es clave para el éxito del Data Warehouse, debe ser muy bien integrada e incluir los aspectos del contenido de datos, las aplicaciones de usuario final y de las herramientas de acceso de datos [Kimball98, p693]. Una estrategia de soporte al usuario final es fundamental para terminar de consolidar el proceso de aprendizaje y asegurar la correcta utilización y éxito del Warehouse a través de una fluida comunicación y retroalimentación de parte de los usuarios.

Otra parte del despliegue consiste desarrollar un marco de trabajo para la emisión del despliegue. El proceso utilizado para desplegar el Data Warehouse debe ser evidentemente similar al proceso utilizado para desplegar un nuevo paquete comercial de software, las emisiones son definidas para un desarrollo basado en requerimientos. Cada cambio significativo al Data Warehouse, consistente en nuevos datos y/o contenido de aplicación o nuevos grupos de usuarios, debe ser considerado una emisión [Kimball98, p705]. La parte final del despliegue es por supuesto el desarrollo de una estrategia para documentar el despliegue.

Mantenimiento y Crecimiento del Data Warehouse

Este es el último componente de la metodología de Ralph Kimball.

Para administrar el entorno del Data Warehouse existente es importante enfocarse en los usuarios de negocio, los cuales son el motivo de su existencia, además de

gestionar adecuadamente las operaciones del Data Warehouse, medir y proyectar su éxito y comunicarse constantemente con los usuarios para establecer un flujo de retroalimentación.

Finalmente, es importante sentar las bases para el crecimiento y evolución del Data Warehouse en donde el aspecto clave es manejar el crecimiento y evolución de forma iterativa utilizando el Ciclo de Vida propuesto, y establecer las oportunidades de crecimiento y evolución en orden por nivel prioridad.

II. ANÁLISIS CUALITATIVO DE LAS METODOLOGÍAS.

Para la elaboración de la guía objeto de este estudio, es necesario realizar un análisis cualitativo de las metodologías abordaron, para determinar cual o cuales elementos de una o de todas las metodologías mejor cumplen el objetivo de servir de guía a noveles en Data Warehousing, y que les apoye en la definición y desarrollo de un proyecto de Data Warehouse.

Es necesario recalcar que un enfoque exitoso para la construcción de un Data Warehouse debe contemplar mucho más que el proceso de diseño, pues el diseño es sólo una parte del todo, aún cuando el diseño mismo se encuentre en el núcleo del desarrollo. Las situaciones que se deben contemplar y resolver envuelven la arquitectura de datos del Data Warehouse, la cultura organizacional, la calidad de datos, integración de datos y demás, con tantas variables en juego es imperativo reconocer que no existe un único método aplicable a toda empresa u organización pues cada una es única, si no más bien conocer y comprender cuales son las filosofías y metodologías que han sido más exitosas y tratar de utilizarlas y adaptarlas a las circunstancias específicas que se afrontan, combinando sus mejores puntos, enseñanzas y prácticas en un todo cohesivo, en lugar de enfocarse en una sola de ellas.

Para que una metodología tenga éxito debe reunir ciertos elementos. La metodología para una iniciativa de Data Warehouse debe resolver el amplio alcance arquitectónico y funcional del Data Warehouse. La estrategia de Data Warehouse que se elija se debe enfocar en producir un marco de trabajo conceptual de soluciones integradas para las distintas capas de la arquitectura, mediante la exploración de los requerimientos de la arquitectura de datos, la arquitectura de metadatos, los procesos del back-end (extracción, transformación y aplicación) y las aplicaciones del front-end (DSS), así como la infraestructura de tecnología. El enfoque de la estrategia será amplio, no profundo, ya que esto depende de la situación en cada empresa. El objetivo es asegurar que la arquitectura de información entregue valor de negocio inmediatamente a los encargados de realizar las decisiones que dirigen el negocio y que coloque al negocio en posición para éxitos futuros. Se culmina con la entrega de una arquitectura conceptual consolidada y los planes para una estrategia de emisiones en etapas. Para que una estrategia de Data Warehouse sea alineada con la estrategia de un negocio, la metodología debe explorar todos los aspectos de la estrategia del negocio.

De las metodologías abordadas en el Marco Teórico es posible hacer por separado un análisis cualitativo de los atributos y limitantes de cada una. A continuación el análisis:

A. Barry Devlin

El primer trabajo publicado en el tema de Data Warehouse, pertenece a Barry Devlin, como se dijo en el Marco Teórico. La metodología expuesta por él, es una guía para implementar y construir un Data Warehouse que incluye el análisis racional del negocio que debe ser hecho, la arquitectura técnica con la que se debe contar y el proceso de implementación completo. Esta metodología provee una fundación teórica excelente para diseñar un ambiente integral de Inteligencia de Negocio (cualquier información que pertenece a la historia, estado actual o proyecciones futuras de una organización). Devlin estructuró su punto de vista como una arquitectura comprensiva al integrar el Data Warehouse al desarrollo de los sistemas de información y considerarlos como un todo. Para Devlin el Warehouse es parte del diseño de sistemas a nivel corporativo, lo cual es un enfoque muy apropiado e innovador.

En su visión, la base de datos del Data Warehouse es parte de una solución más grande, no un final en sí mismo. El Data Warehouse es un *activo estratégico* desarrollado para servir los intereses de la comunidad de negocios entera. El Data Warehouse es una colección de datos que será usada por los usuarios de negocio para suplir la mayor parte de las necesidades de acceso y análisis de información. El Warehouse también será un vehículo para incrementar la calidad y disponibilidad de los datos conforme a una naturaleza diversa de necesidades y usuarios, esto permite luego ingresar datos integrados y de calidad en el ciclo de retroalimentación de la información y para que puedan ser usados por otros sistemas corporativos.

En su enfoque, Devlin separa claramente los distintos componentes de una arquitectura eficaz y adecuada para la realización de un proyecto de Data Warehouse, que deben de estructurarse y trabajar en conjunto para asegurar el éxito del proyecto. Además, él proporciona un número de técnicas, sugerencias y tips de cómo implementar y diseñar el ambiente de Inteligencia de Negocio.

El elemento diferenciador de la Arquitectura propuesta por Devlin, es la Arquitectura en tres capas, entre ellas destaca la capa intermedia de datos reconciliados. Su propósito es tomar los datos de sistemas operacionales diversos, heterogéneos, distribuidos geográficamente y combinarlos y enriquecerlos en una imagen única y lógica del *modelo de datos empresarial*. De esta capa los datos pasan a la capa de datos derivados a través de procesos que generan cualquier combinación de datos que los usuarios puedan requerir. Todo esto es parte del Data Warehouse y no una plataforma separada, por lo que Devlin logra ampliar el alcance del DW en la organización.

Esta característica resulta de mucho valor para aquellos ambientes donde existen muchos y diversos sistemas fuentes de datos y algunos sistemas de soporte a la toma de decisiones que no están logrando su objetivo. Entonces, se hace necesario reconciliar los datos procedentes de estos sistemas sin introducir nuevos entornos, porque implican más duplicación de datos, sin forzar los datos existentes para que quepan en los ambientes presentes aunque no sean adecuados para ellos, porque no se lograría suplir las necesidades de información de la empresa, y hacer coincidir todo en un ambiente acorde a los estándares de buenas prácticas de arquitectura de datos que proclaman una clara separación del entorno operativo con el informacional.

Otro de los componentes que Devlin claramente destila y del que aborda las características y estructura que lo constituyen, es la interfaz de usuario, esto se deriva del hecho de que para los usuarios de negocios, el Data Warehouse no es una Base de Datos sino una aplicación. Aunque el Data Warehouse sea inmejorable, si la interfaz es pobre difícilmente será exitoso.

La clave de la importancia de este autor radica en que sus enseñanzas son vitales para poder asimilar de forma clara como debe de ser idealmente un Data Warehouse, sus conceptos teóricos son inmejorables y la evidente diferencia que marca entre uno y otro componente de la arquitectura del Data Warehouse, así como la descripción detallada que hace de los mismos, permite que sea invaluable como metodología de enseñanza, aunque con la debilidad de presentar pocas herramientas de aplicación real en el desarrollo de un proyecto, sino más bien componentes a nivel lógico y conceptual.

Una de las debilidades de la propuesta de Devlin es que no muestra la forma en los distintos componentes de la arquitectura que él describe deben estar ensamblados e interactuar para formar un todo coherente, por lo que deja inconcluso este aspecto metodológico y arquitectónico medular. Para que una metodología sea sólida es de suma importancia, que ésta presente un diagrama completo de la situación y solución descrita, porque ocurre que surgen muchas preguntas y cuestiones que aparecen sin aparente solución. El enfoque de Devlin deja sin puntualizar su propuesta, por lo que se vuelve difusa su efectividad en algunos puntos.

B. William Inmon

La definición de Data Warehouse de Inmon ha puesto en claro una buena síntesis de lo que es un Data Warehouse y ha ayudado inmensamente al desarrollo de la industria, debido a que provee guías concretas para construir un Data Warehouse. Implícitamente, esta definición sustenta uno de los principios fundamentales del desarrollo de un Data Warehouse, el principio que el ambiente de origen de los datos y el ambiente de acceso de datos deben estar físicamente separados en diferentes bases de datos y en equipos separados.

Inmon también identifica la importancia de utilizar un Data Warehouse para guardar datos históricos continuos, ya que uno de los mayores obstáculos para el análisis de información relevante es no contar con datos disponibles sobre un periodo de tiempo extendido. Operacionalmente, se tiende a almacenar solamente una vista actual del negocio, lo cual es un período mucho muy corto para un análisis serio de tendencias.

A Inmon se le asocia frecuentemente con los esfuerzos de Data Warehouse a nivel empresarial, que involucran desde un inicio todo el ámbito corporativo, sin centrarse en un incremento específico hasta después de haber terminado completamente el diseño del Data Warehouse. En su filosofía, un Data Mart es sólo una de las capas del Data Warehouse, los Data Marts son dependientes (obtienen la información) del depósito central de datos o Data Warehouse Corporativo y por lo tanto se construyen luego de él.

El enfoque de Inmon de desarrollar una estrategia de Data Warehouse e identificar las áreas principales desde el inicio del proyecto es necesario para asegurar una solución integral. Esto ayuda a evitar la aparición de situaciones inesperadas en el futuro

cercano del proyecto que le puedan poner en peligro, debido a que se conoce con antelación y bastante exactitud la estructura que presentarán los principales núcleos del desarrollo, lo cual permite enfocar los esfuerzos del desarrollo actual para ser compatible con los subsiguientes.

Inmon es defensor de utilizar el modelo relacional para el ambiente en el que se implementará el Data Warehouse Corporativo, asegura que esta es la alternativa más adecuada para que el almacén central sea más eficiente sin afectar a los usuarios finales ya que la frecuencia de acceso de los mismos es muy escasa en este nivel. Mientras, aplicará al esquema estrella o modelado dimensional a la aplicación Front End que llama Data Mart, y que es donde realmente tiene lugar el acceso de los usuarios en su Arquitectura.

Inmon ciertamente coincide en que el modelado dimensional está bien para los Data Mart, pero hace énfasis en que estos deben ser dependientes del Data Warehouse Corporativo; sin embargo está muy convencido que un diseño basado en Diagramas Entidad Relación es mucho más apropiado para el Data Warehouse central de mayor magnitud. Según Inmon y aquí también coincide Devlin, la estructura ideal que se busca para un Data Warehouse, porque proporciona la manera más efectiva de coleccionar, almacenar y diseminar la Información, es muy probablemente:

- *Datos antiguos*, limpiados en un RDBMS (potencialmente un Data Warehouse Empresarial).
- *Datos reconciliados*, desde el Data Warehouse Empresarial obtienen su información los Data Marts, cubos y otras herramientas para análisis y reportes que utilicen un enfoque multidimensional para mostrar la información.

El problema que trae consigo este enfoque es que es ideal para los propósitos de desarrollo del equipo de Tecnología de Información pero no para las finanzas de la organización. A esta estructura no es posible dividirla en partes modulares que al implementarse comiencen a ser explotadas, sino que es hasta que toda la arquitectura está en su lugar que los usuarios de negocio obtienen beneficio de ella. Es un enfoque de “big bang” que trae consigo mucho riesgo a la compañía que invierte grandes esfuerzos en el desarrollo del DW y no es sino hasta que comienzan a aparecer los Data Marts que realmente comienza a explotar su inversión y a obtener beneficios de ella.

La estrategia de big bang de Inmon es en cierta medida comparable con el ciclo de vida clásico del desarrollo de aplicaciones y comparte con ésta metodología algunos riesgos, el más peligroso para un Data Warehouse es sin duda, que en el marco de que es imposible conocer en avance cuales son las necesidades concretas de información de una empresa, el ambiente dinámico en que se mueve la organización, el cambio de estructura que conlleva el desarrollo de la nueva plataforma y los consiguientes cambios a los sistemas transaccionales que su introducción implica; es muy probable la posibilidad de que luego de un considerable plazo de tiempo y recursos invertidos en el desarrollo del Data Warehouse, una vez finalizado y puesto en explotación el mismo, se hagan evidentes algunos cambios fundamentales que traen consigo altos costos de desarrollo para la organización, poniendo en evidente peligro el éxito de todo el proyecto en sí y que podían ser evitados con una pronta detección en una temprana puesta en explotación de un primer avance del Data Warehouse.

Otra de las restricciones que trabajan en contra de la metodología de Inmon es que consume mucho más tiempo trabajar única y completamente con esta metodología, esto tiene como consecuencia que muchas empresas se inclinen por usar metodologías de la que obtengan resultados tangibles en un espacio menor de tiempo.

C. Ralph Kimball

Kimball difiere de los otros autores abordados en enfoque: “El Data Warehouse no es nada más que la unión de todos los Data Marts que lo constituyen” [K, p19]. En el mundo de Kimball el Data Mart es el Data Warehouse, esto se afirma en el sentido de que Kimball expone que al construir los Data Marts ya se está construyendo el Data Warehouse de una manera incremental. Un Data Mart es un subconjunto de datos organizados, como en el Data Warehouse, para el soporte a la toma de decisiones, pero que sólo representa la visión de un departamento o individuo, por este motivo Kimball es frecuentemente asociado con esfuerzos departamentales y no corporativos.

En la actualidad la mayoría de los proyectos de Data Warehouse implementan el modelo de Data Marts de Kimball en lugar del esquema de Data Warehouse empresarial propuesto por Bill Inmon o de la arquitectura en tres capas de Devlin, esto obedece a motivos de tiempo, costo y el riesgo de fracaso asociados con el desarrollo de los dos últimos [Lee02]. A esta tendencia general se le ha identificado como la aproximación que

pretende garantizar la probabilidad de éxito más grande en la implementación de un Data Warehouse, tanto por la rapidez en la obtención de resultados en períodos cortos (tiempo) con inversiones moderadas (costo) como por la modularidad posible de alcanzar con este enfoque considerando cada Data Mart como un incremento del sistema final, el Data Warehouse (menor riesgo de fracaso) [Wolf99].

El punto central de la metodología de Kimball es el modelado dimensional. Un buen diseño asegura en gran parte el éxito del proyecto. El objetivo primordial que se persigue con un Data Warehouse, servir de soporte a la toma de decisiones, sólo es alcanzado si el diseño del Data Warehouse - Data Mart propone una estructura consistente y adecuada a las necesidades de información de la organización. Por este motivo Kimball pone énfasis en el diseño de los Data Marts, para lo cual utiliza el modelado dimensional en la versión del esquema estrella. Kimball afirma que esta tecnología siempre puede ser aplicada en cualquier proyecto de Data Warehouse y que es el método más adecuado para alcanzar el objetivo ya mencionado. El esquema estrella representa la denormalización óptima de los datos que mejor se adapta a los requerimientos de los usuarios.

El concepto clave que ha popularizado la metodología del señor Kimball es que él aborda el proyecto de Data Warehouse como un proceso de *Implementación Gradual*, Data Mart a Data Mart. Sin embargo, Kimball también pone en claro que lo primero que se debe hacer al comenzar el modelado dimensional es analizar la sólida base que representa el Diagrama Entidad Relación de la empresa y a partir de allí iniciar el modelado dimensional, es decir, primero se debe contemplar toda la organización empresarial para encontrar los procesos discretos del negocio, luego corresponde establecer cuales son todos los posibles Data Marts y de entre ellos seleccionar cual es el más adecuado de implementar, en la correspondiente iteración del Data Warehouse. A continuación ya se puede enfocar en él o los Data Mart que pertenecen a la etapa actual del proyecto y proceder con el ciclo de vida que expone en su metodología.

El ciclo de vida propuesto trae como consecuencia que existan Data Marts que se traslapen, para el caso en que se tienen que contemplar las diferentes vistas que distintos usuarios o departamentos tienen acerca del Modelo de Datos Corporativos, las implementaciones de vistas disímiles deben realizarse en Data Mart separados. Para asegurar la correcta unión y engranaje de los Data Marts y evitar que se conviertan en

conjuntos disjuntos (islas) de información Kimball establece el método de dimensiones conformadas y lo designa como el *Bus del Data Warehouse*.

Todos estos elementos para que funcionen sinérgicamente deben ajustarse en un marco de trabajo sólido, flexible y extensible, que constituye la arquitectura que guiará la implementación del Data Warehouse. Kimball como ya se vio, utiliza una matriz para clasificar tres grandes áreas: Datos, Tecnología e Infraestructura, los cuales tienen cuatro niveles de detalles siendo el más bajo la implementación física del Data Warehouse. Una vez establecida la arquitectura, se procede a implementar los primeros incrementos.

La Implementación por incrementos de Data Marts trae consigo algunas consideraciones importantes:

- La arquitectura Data Warehouse se debe desarrollar al principio del proyecto.
- El primer incremento se desarrolla basándose en la arquitectura.
- La operación del Data Warehouse puede implicar la realización de cambios en la arquitectura.
- Cada incremento adicional puede extender el Data Warehouse.
- Cada incremento puede causar ajustes en la arquitectura.
- La operación continua puede causar ajustes en la arquitectura.

Por estas consideraciones muchos expertos afirman en que el enfoque de Kimball trabaja mejor si primero existe una estrategia de implementación en la organización, pues de esta forma se reduce el número de cambios, que en muchos casos representa una gran parte de los esfuerzos de mantenimiento o de desarrollo del nuevo incremento. Estos cambios son necesarios de realizar para asegurar el adecuado funcionamiento y crecimiento del Data Warehouse. Es decir, que para un proyecto que envuelve la creación de más de un Data Mart es aconsejable que primero se deba desarrollar una estrategia corporativa como esqueleto y luego continuar con la metodología de Kimball.

Otro gran detrimento a la metodología de Kimball, son consideraciones al corazón de la misma, el modelado dimensional. Un esquema estrella se construye obteniendo y asimilando requerimientos de los usuarios, lo que determina la forma y contenido de la estrella. El resultado de la estrella es óptimo para los usuarios que participan en el proceso de obtención de requerimientos. El modelado dimensional es

excelente para representar las vistas de las personal que son de pensamientos similares, pero diferentes grupos de personas querrán su propia estrella que represente sus propias vistas. El esquema estrella se forma alrededor de los requerimientos de usuarios y porque estos requerimientos varían de un tipo de usuarios a los otros no es de sorprender que diferentes estrellas sean óptimas para diferentes tipos de usuarios.

El problema real es cuando existen múltiples ambientes independientes de esquemas estrella, los mismos datos detallados aparecen en cada estrella. No existe reconciliación de datos y las nuevas estrellas requieren la misma cantidad de trabajo para la creación que las antiguas estrellas. Como resultado:

- Las uniones crecen innecesariamente grandes cuando cada estrella necesita datos detallados que otra estrella ya ha obtenido.
- Los resultados de cada estrella son inconsistentes con el resultado obtenido de cada otra estrella y la habilidad de reconciliar las diferencias no es aparente.
- No existen bases para construir nuevas estrellas porque cada una es construida independientemente.
- La interfase para soportar las aplicaciones que alimentan las estrellas se vuelve inmanejable.
- Se genera una gran cantidad de trabajo extra al construir cada parte en comparación al enfoque de Data Warehouse Corporativo.

Una vez establecido el análisis sobre cada una de las metodologías estudiadas se han sentado las bases para la construcción de la guía objeto de estudio.

D. Síntesis

Las filosofías de Inmon y Kimball difieren principalmente en escala, Data Warehouse Corporativo versus Data Warehouse por incrementos de Data Marts. Devlin es coincidente con Inmon en colocar los Data Marts como una capa aparte del Data Warehouse central, pero introduce otra capa intermedia, la capa de datos reconciliados en su arquitectura del Data Warehouse.

Sin embargo, la diferencia real entre ellos se encuentra en el enfoque con el que se modela la Vista Corporativa. Todos coinciden en que es necesario al iniciar el proyecto analizar el modelo de datos empresarial con el que cuenta la empresa. A partir de aquí:

- Kimball propone un esquema de denormalización del ERD para identificar procesos discretos de negocios (áreas de interés) con sus posibles tablas de hechos y dimensiones. Luego, selecciona un subconjunto de datos para modelarlo utilizando el esquema estrella y continuar el desarrollo del Data Warehouse de forma iterativa, modelando un nuevo subconjunto cada vez.
- Los otros autores realizan modelado Entidad Relación sobre todo el conjunto de datos empresariales durante la primera iteración del proyecto, de modo que se reconozca que datos son meramente operativos y cuales son de soporte a la toma de decisiones, se identifican las áreas de interés y se procede a una reorganización de los datos para acomodarlos a las exigencias del Data Warehouse. En cada etapa se implementa solo un área de interés. En las siguientes iteraciones, de ser necesario, solo ajustan el modelo inicial.

El fracaso que sufren algunas organizaciones al utilizar la metodología de Kimball se deriva de que obvian la visión corporativa del Data Warehouse. La ventaja del enfoque corporativo es que con un buen diseño para el Data Warehouse empresarial se puede prevenir de los esfuerzos de duplicaciones y sincronización, que suelen aparecer en el enfoque de incrementos por Data Marts propuesto por Kimball.

III. CONTEXTO NICARAGÜENSE

La tecnología de Data Warehousing no es nueva en el Mercado, pues se remonta a más de una década atrás, pero su crecimiento en el mundo a un ritmo acelerado, denota que la última palabra no está dicha, sino más bien que apenas se comienza a vislumbrar la verdadera magnitud que ésta puede alcanzar.

El fundamento para la construcción de un Data Warehouse es algo ampliamente discutido y reconocido, por lo que cualquier empresa que desee ser competitiva y obtener valor a partir de los datos que posee, indudablemente se involucrará en un proyecto de Data Warehouse, en mayor o menor escala. Nicaragua no debe ser la excepción a la regla. Las empresas para sobrevivir y crecer en el país y más allá de sus fronteras deben estar en la capacidad de obtener el mayor beneficio posible de cada uno de sus recursos y la información es un recurso que, sin lugar a dudas, es invaluable.

Para la obtención de información confiable, consistente y oportuna, no es posible basarse en un entorno tecnológico que esté diseñado y dirigido a automatizar transacciones (OLTP), sino hacer un cambio de paradigma y moverse a otro entorno, el analítico (OLAP) orientado a la toma de decisiones, que gestione la información necesaria para dirigir un negocio, ésta es la premisa del Data Warehousing.

Según el Directorio Económico Urbano INEC/MIFIC/GTZ, en su edición de Febrero de 1998, en Nicaragua la cantidad de empresas que caben en la categoría de grande y mediana con un rango de trabajadores de [21- 100 +] se encuentran clasificadas de la siguiente manera:

Tabla III-1 Empresas en Nicaragua

TIPO DE EMPRESA	No. DE EMPRESAS	No. EMPLEADOS
Grande	160	100 o más
Mediana	712	[21-100]
Pequeña	4,526	[6 - 20]

Y estas constituyen aproximadamente el 10 % de participación en el mercado laboral de la nación. Las, pequeñas y mediana empresa constituyen la inmensa mayoría de las empresas en el país [La PYME en Nicaragua/MIFIC], y están en crecimiento elevando

su clasificación a una categoría superior, las grandes empresas se consolidan en el mercado nacional y buscan su expansión.

Nicaragua, se encuentra en una posición tecnológica distinta de la mayoría de los países de la región, ya que debido a circunstancias que se dieron en su historia, el boom empresarial y tecnológico no tuvo lugar sino hasta hace algunos años, motivo por el cual el mayor porcentaje de empresas y organizaciones que radican en Nicaragua y que cuentan con sistemas de cómputo, son medianas y grandes empresas que exhiben una plataforma tecnológica relativamente reciente, donde la existencia de sistemas de información, no supera la década de antigüedad. Además, la presencia de proyectos de Data Warehouse es insipiente, pero definitivamente esta tecnología se desarrollará prontamente, pues las organizaciones en crecimiento necesitan ser cada vez más competitivas, disminuir la brecha tecnológica e incluirse cada vez más en un mundo globalizado.

En Nicaragua, debido a que hasta el momento no se ha desarrollado grandemente la industria Data Warehousing, no existen cursos de titulación o especialización en la materia y el personal capacitado y con experiencia en el ramo es realmente escaso, por lo cual se considera que la mayoría de tecnólogos de la información presentes en el país son noveles en la materia.

El contexto tecnológico de Nicaragua, representa una situación muy especial que muestra la necesidad de una metodología adaptada a la situación presentada, que permita sacarle provecho a la realidad nicaragüense y sus características propias. Es posible, para esto, adaptar las metodologías de Data Warehouse más exitosas a las circunstancias particulares del país unificando, reconciliando, condensando y adecuando las diferentes visiones, enfoques y procedimientos, para obtener una guía metodológica que incluya los aspectos que más se ajusten al contexto tecnológico de Nicaragua y al hecho de que la mayoría de los profesionales de tecnología de información del país son noveles en este campo. Para este caso se abordan las metodologías de Barry Devlin, William Inmon y Ralph Kimball, por su claro éxito y reconocimiento y que previamente se discutió.

Es un hecho real que casi ningún equipo en un proyecto de Data Warehouse construye un Data Warehouse utilizando una metodología pura. Básicamente cada equipo esta construyendo su propio enfoque y utilizando algunas de sus propias técnicas, una vez que han asimilado lo mejor de las metodologías que elijan. Por ello, ésta guía significa una

ayuda en la reducción de la curva de aprendizaje de la tecnología de Data Warehouse para el público nicaragüense. Dicha guía servirá de directriz al personal que realice por primera vez un proyecto de Data Warehouse y les presentará un ejemplo real de su aplicación en una institución financiera, también permitirá que los usuarios de negocios comprendan en que consiste un Data Warehouse y los beneficios que se pueden obtener, para que reconozcan sus ventajas y se impulsen a implementarla en sus organizaciones.

La guía constituye una base teórica sólida que le permite servir de referencia y material de apoyo a todos aquellos que deseen construir por primera vez un Data Warehouse, y les habilita a luego profundizar en la metodología de alguno de los autores mencionados que mejor se adapte a sus requerimientos, al brindarles una visión precisa y compacta de las propuestas por las que cada uno ha destacado y se han merecido el reconocimiento mundial en este ámbito.

IV. ESTUDIO DE FACTIBILIDAD

El estudio de Factibilidad consta de cuatro apartados de factibilidad: Técnica, Operativa, Económica-Financiera y Legal, los cuales permitirán concluir y tomar la decisión de poner en práctica un proyecto, en este caso: El Data Warehouse.

El apartado de la **Factibilidad Técnica** es un estudio de la función, rendimiento y restricciones que puedan afectar la consecución de un sistema aceptable para la implementación y uso del Data Warehouse. Esto implica: adquirir y configurar los servidores, alojar el espacio de disco, construir las conexiones de red y seguridad; básicamente la arquitectura técnica del Data Warehouse.

La **Factibilidad Operativa** permite verificar si la organización está en condiciones de absorber el cambio que significa la implementación y uso del Data Warehouse para la cultura organizacional, entre estos podemos mencionar:

- El personal no tiene experiencia en el manejo de este tipo de sistemas.
- Los gerentes verán modificados los procedimientos de obtención, análisis y uso de información y además deberán estar a la cabeza del cambio.
- Se debe agregar el factor de la resistencia natural al cambio en todos los niveles.
- Se requerirá la capacitación del personal.
- Además de todo esto se deberá seguir cumpliendo con las actividades diarias para que el Negocio siga funcionando.

Como vemos no se puede trivializar el estudio de factibilidad operativa, pues este permite poder reducir el margen de riesgo del proyecto y asegurar su éxito dentro de la organización.

La **Factibilidad Económico-Financiera** del proyecto es una evaluación del costo de desarrollo sopesado con los ingresos netos o beneficios obtenidos del producto desarrollado, en este caso del Data Warehouse. La justificación económica incluye una amplia gama de aspectos a tener en cuenta como son el análisis de costo-beneficios, las estrategias de ingresos de la empresa a largo plazo, costo de recursos necesarios para el desarrollo y crecimiento potencial del mercado.

La **Factibilidad Legal** permite determinar y delimitar cualquier infracción, violación o responsabilidad legal en que se podría incurrir por el desarrollo del Data Warehouse.

Por último, se confeccionará el informe final para la alta dirección, con el resumen del Estudio de Factibilidad del Proyecto de inversión en un entorno analítico para la toma de decisiones.

A continuación el estudio de Factibilidad para la empresa donde se aplica la guía desarrollada.

A. Factibilidad Técnica

En primer lugar se procederá a calcular la capacidad y a establecer el perfil técnico de los dispositivos y periféricos en función de los datos y aplicaciones que serán procesados por la futura instalación. También se necesita determinar que aplicaciones son necesarias para el desarrollo del proyecto de Data Warehouse.

1. Requerimientos de la instalación y aplicaciones

La tabla IV-1 presenta el dimensionamiento de los elementos técnicos necesarios para llevar a cabo un proyecto de Data Warehouse.

Tabla IV-1 Requerimientos de Data Warehouse

Categoría	Elemento	Cnt	Consideraciones	Requerimientos
Hardware	Servidor	1	Tamaño y organización de la memoria	1.0 GB memoria o mayor
			Escalabilidad	Escalable en memoria, procesadores y disco duro
			Tecnología Procesador	RISC(*) o CISC
			Cantidad de procesadores	2 o más
			Velocidad de procesador	550 MHz o superior
			Cantidad de Discos	6 discos en RAID 0+1 (*) 3 discos en RAID 0
			Tiempo medio acceso a disco	SCSI – 3 (*) 10krpm
			Capacidad de almacenamiento en discos	100 GB en conjunto
			Tipo (Fijos o removibles)	Hot Swap Plug
			CD's y Unidades de CD's	Unidad de CD-ROM
			Unidades de Tape Backup	Presente(*)

Categoría	Elemento	Cnt	Consideraciones	Requerimientos
	Impresora	1	Tipo (tinta o láser)	Láser
			Velocidad	12 ppm o superior
	UPS	2	Capacidad	2000 watts
Software	Catálogo de Metadatos		Servicios de tipo general	Control administrativo Información personal Información administrativa Calificación de contenido
	Sistema Operativo		Procesamiento en lotes de programas de aplicación	Indispensable
			Multiprocesador	Indispensable
			Multiusuario	Presente
			Multitareas	Indispensable
	DBMS		Tecnología (RDBMS, MDBMS)	RDBMS (*), MDBMS
			Open System	Preferido pero no indispensable
			Lenguaje de Consulta	Compatible con estándar SQL92
			Características de Data Warehouse	Juego de Instrucciones de DW Soporte para particiones que contengan datos fuertemente relacionados
			Soporte de esquema estrella	Indispensable
			Uso de metadatos	Siempre que sea posible
	Modelador de Datos		Tipo de modelado (Dimensional o ERD)	ERD DM(*)
	Herramienta ETL		Funcionalidades	Extracción, deduplicación, duplicación
			Soporte de plataformas	Microsoft Windows, Linux
			Estandarización	
	Ofimática		Procesamiento de textos	Indispensable
			Hojas de Cálculo	Indispensable
			Software para modelar	Opcional
Herramienta de Acceso a Datos		OLAP	HOLAP (*), MOLAP, ROLAP	
		Reportes	Acceso a base de datos	
		Data Mining	No requerido	
		EIS	No requerido	
		Productos hábiles para Web	Opcional	

(*) Indica la opción preferente

2. Opciones Disponibles

Una vez determinados los requerimientos técnicos se procederá a evaluar opciones disponibles en el mercado que cumplen los requisitos establecidos y determinar su viabilidad. La tabla IV-2 presenta un resumen de éstas opciones disponibles en el mercado nacional y que llenan los requisitos establecidos anteriormente.

Sobre los elementos detallados en la tabla IV-2 se analizará la viabilidad de adquisición para determinar cual será la propuesta de adquisición óptima para la empresa en la que se realice el proyecto ejemplo.

Tabla IV-2 Alternativas de Adquisición

Elemento	Opciones
Servidores	Dell Power Edge 2500
	Compaq Proliant ML-350
	Hewlett-Packard Netserver TC4100
Impresora	Hewlett-Packard LaserJet 1200N
	Epson LA LQ-2180
Catálogo de Metadatos	Computer Associates Platinum Repository
	Metadata Management Corporation Design Bank
DBMS	IBM DB2
	Microsoft SQL Server 7.0
	Oracle9i Enterprise Edition
Herramientas ETL	Informatica Power Mart Designer
	IBM Warehouse Manager
	Oracle Warehouse Builder
Modelador de Datos	Oracle Designer/2000
	Logic Works ERWin
Herramientas de Acceso a Datos	Microsoft Access, Excel
	Crystal Decisions Crystal Reports
Ofimática Suites	Microsoft Office 2000
	StarOffice 6.0
Sistema Operativo (en dependencia del DBMS)	Microsoft Windows 2000 Server
	Red Hat Linux 8.1

En conclusión se ha establecido que la alternativa que representa una mejor adquisición y desempeño desde el punto de vista técnico, pues cumple plenamente con los requerimientos del sistema, en cuanto a programas y equipos esta dada por las variables que aparecen en la opción A de la tabla siguiente. El software exclusivo de Data Warehouse (catálogo de metadatos, herramienta de ETL) no tiene disponible un representante en el país, a diferencia de los restantes elementos para los que si se cuenta con un distribuidor y departamento de soporte técnico. Los elementos fueron seleccionados producto de un análisis de sus ventajas y desventajas, así para el caso del servidor se tomó en cuenta la estabilidad de la firma que lo respalda, escalabilidad, precio, características de los componentes, soporte técnico disponible.

Una segunda alternativa a considerar en el estudio de factibilidad Económica Financiera, introduce una variante en la alternativa del servidor sustituyendo el *Hewlett Packard Netserver TC4100* por un *Compaq Proliant ML-350*, el cual consta con más de un distribuidor autorizado en el país, mientras el resto de opciones se conserva.

Tabla IV-3.- Opciones recomendadas

Elemento	Opción A	Opción B
Servidores	<i>Hewlett Packard</i> Netserver TC4100	<i>Compaq</i> Proliant ML-350
Impresora	<i>Hewlett-Packard</i> LaserJet 1200N	<i>Hewlett-Packard</i> LaserJet 1200N
Catálogo de Metadatos	<i>Computer Associates</i> Platinum Repository	<i>Computer Associates</i> Platinum Repository
DBMS	<i>Microsoft</i> SQL Server 2000	<i>Microsoft</i> SQL Server 2000
Modelador de Datos	<i>Computer Associates</i> ERWin	<i>Computer Associates</i> ERWin
Herramientas de Acceso a Datos	<i>Crystal Decisions</i> Crystal Reports	<i>Crystal Decisions</i> Crystal Reports
Ofimática Suites	<i>Microsoft</i> Office 2000	<i>Microsoft</i> Office 2000
Sistema Operativo (depende del DBMS)	<i>Microsoft</i> Windows 2000 Server	<i>Microsoft</i> Windows 2000 Server

Debido a que los costos del proyecto de data Warehouse podrían resultar ser muy elevados para esta institución en su etapa inicial y dado que ésta se encuentra en un momento de renovación de su plataforma tecnológica, se ha suprimido la Herramienta de ETL (Extraction, Transformation, Load) del cuadro anterior y compensar su funcionalidad con algunas características propias del DBMS escogido. Es posible además, prescindir del Catálogo de Metadatos y utilizar la característica de metadatos propia de SQL Server 2000 con objetivos de la disminución de costos.

Por lo tanto, no se consideraran ambas herramientas en el apartado de Factibilidad Económica- Financiera.

B. Factibilidad Operativa

El Data Warehouse tendrá un profundo impacto en el funcionamiento de la organización, debido a que una vez implementado, es un elemento estratégico de la empresa pues disminuye el grado de incertidumbre con la que se toman las decisiones.

Algunas de las ventajas obtenidas del uso de un Data Warehouse que modifican, de manera positiva, el desempeño de la organización son:

- Contar con una fuente confiable de información que presenta una visión consistente de los datos e información de la empresa y su entorno.
- Elimina la necesidad de calcular información rutinaria de manera manual.
- Agiliza la gestión de datos e información relevante.

- Pone a la disposición de los niveles gerenciales la habilidad de hacer reportes de manera improvisada para satisfacer sus necesidades de información.
- Permite monitorear oportunamente los indicadores del negocio.
- Habilita el análisis de tendencias, lo que permite descubrir oportunidades de utilidades para la empresa y debilidades que se deben corregir.

Alcance y Oportunidad

Todo esto implica cambios organizacionales importantes en el esquema de trabajo. Estos cambios ocurrirán principalmente en los niveles donde se efectúa la toma de decisiones: la media y alta gerencia. Sin embargo, a pesar de las ventajas que presenta la introducción de un entorno analítico para quienes hagan uso del mismo, la curva de aprendizaje y la resistencia al cambio son factores a considerar.

La empresa objetivo se encuentra en un proceso de cambio tecnológico en la que se incluye poner a disposición del personal tecnología que le facilite el desarrollo de sus funciones, como la migración de sistemas transaccionales a una plataforma más adecuada y la introducción de ordenadores del tipo Handheld para personal operativo y gerentes de todas las escalas y a todo el personal de tecnología de información. En este marco el Data Warehouse se convierte en la secuencia ideal, al brindar soporte a la toma de decisiones. Las condiciones presentadas por el personal en las circunstancias mencionadas, permite afirmar que el personal de la organización presenta poca resistencia al cambio y que asimilarán de forma natural el nuevo entorno, si se establece un esquema de capacitación adecuado como se ha hecho en las otras ocasiones, lo cual trae consigo además una drástica disminución en la curva de aprendizaje.

El alcance del Data Warehouse será organizacional. En cuanto al personal que conforma la media y alta gerencia, quienes harán mayor uso del Data Warehouse, el nivel de preparación medio es muy alto. Ellos constituyen una base de recursos humanos que tiene interés en explotar las ventajas que brinda el Data Warehouse. La estructura organizativa de la empresa no se verá afectada, exceptuando que la gestión de información será rápida y efectiva y el personal dispondrá de mayor tiempo para realizar otras funciones, lo que aumentará el nivel de rendimiento y productividad del personal así como la calidad del trabajo realizado.

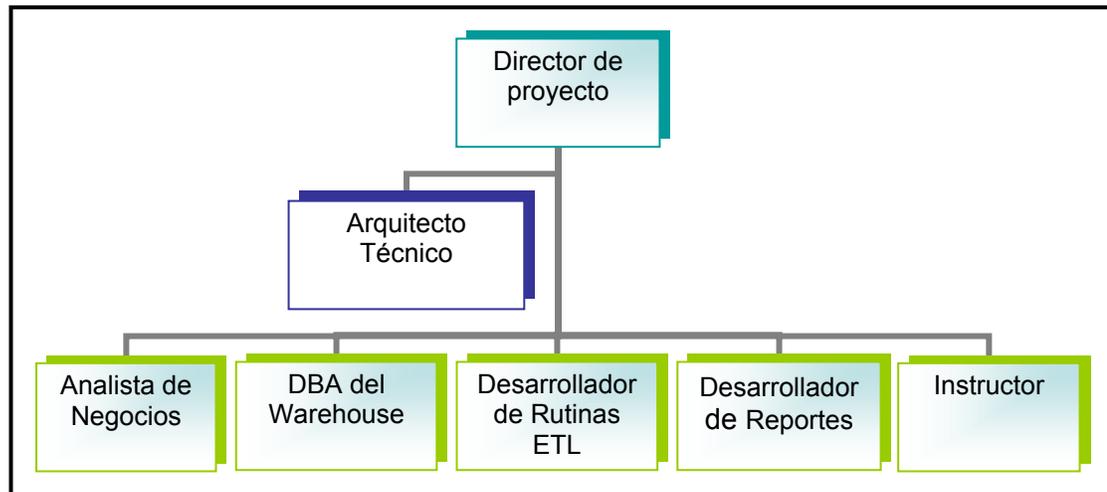
Dotación

El personal necesario para construir un Data Warehouse, no se puede establecer con exactitud y antelación al desarrollo del proyecto, pues existen muchas circunstancias que no es posible anticipar y que pueden ocurrir y necesitar de la intervención de personal ajeno al equipo base del proyecto que se encargará de desarrollar y dar mantenimiento al Data Warehouse. No obstante, es posible hacer una buena estimación del grupo de recursos humanos que intervendrán y conformarán el equipo base, y las funciones que cumplirán. Los datos proporcionados a continuación son una descripción de las funciones básicas que se deben desempeñar, para construir un Data Warehouse.

Tabla IV-4.- Dotación de Personal

Cargo	Funciones
Analista de Negocios	Establece la definición de términos y operaciones del negocio. Análisis de los procesos de negocios. Obtiene los requerimientos de información de los usuarios
Arquitecto Técnico	Análisis de los sistemas fuentes de datos, construcción de los modelos de datos del Data Warehouse, definición de la apariencia de la Base de Datos
Instructor	Capacitación a los usuarios en el uso correcto de las funciones y habilidades del Data Warehouse
DBA - Administrador de la Base de Datos	Implementa los modelos de datos del Data Warehouse, mejora el desempeño de la Base de Datos, controla y monitorea el acceso y uso del Data Warehouse por los usuarios
Director de Proyecto	Coordina todas las actividades del personal. Establece el límite del proyecto y el alcance de cada iteración. Gestiona el plan del proyecto. Decide sobre las herramientas y metodología a utilizar.
Desarrollador de Reportes	Diseño e implementación de los modelos de reportes a disposición de los usuarios del Data Warehouse
Desarrollador de rutinas ETL	Desarrollo de los diseños para poblar la Base de Datos con los datos necesarios cuando se realice la carga del Data Warehouse. Asegurar la calidad de los datos
Soporte Técnico	Soporte con las plataformas de sistemas operativos, aplicaciones de escritorio, infraestructura de telecomunicación.

El organigrama del grupo base del proyecto Data Warehouse es como sigue:

**Ilustración IV-1 Organigrama**

Luego de la emisión del proyecto de Data Warehouse y de que se inicie su explotación es necesario alojar personal que se encargue de su mantenimiento. Es el caso de que las funciones anteriormente expuestas, es muy aconsejable que sean desempeñadas casi todas por personal perteneciente a la organización (ver apartado de factibilidad económica-financiera). Se deberá reestructurar el departamento para que personal que haya desempeñado funciones de desarrollador de rutinas ETL y desarrollador de reportes puedan en adelante darle mantenimiento y solucionar posibles problemas en el Data Warehouse, cuando se requiera. El DBA debe estar en la posición de implementar herramientas de monitoreo que le permitan ajustar mejor el desempeño del Data Warehouse a las necesidades de la organización.

La dotación de personal descrita, esta compuesto por personal de tiempo completo, de los que se exceptúan el DBA y Soporte Técnico los cuales son de carácter ocasional. Del personal de tiempo completo el único cargo que debe ser desempeñado durante todo el proyecto es el de Director de proyecto. Los roles de Analista de Negocios y Arquitecto Técnico son desempeñados en las primeras etapas del proyecto, Desarrollador de rutinas ETL y Desarrollador de Reportes aparecen después cuando ya se tiene listo el diseño del Data Warehouse. El DBA aparece en la implementación del Data Warehouse, mientras que Instructor es un rol que se desempeña cuando el Data Warehouse esta a punto de entrar en producción. Soporte Técnico se requiere esporádicamente para preparar y mantener la plataforma tecnológica que soporta el Data Warehouse.

Una sola persona puede desempeñar más de una función. En este sentido es posible que el personal que ha trabajado en los primeras etapas luego se mueva a otro puesto según avance el proyecto como pasar de Arquitecto Técnico a Desarrollador de rutinas ETL o que ya se desliguen del grupo base en modo de tiempo completo en el proyecto y pase a ser de carácter consultivo ocasional, lo cual implica un costo menor en el presupuesto.

El Data Warehouse es un proyecto incremental, dividido en iteraciones, el resultado de cada etapa se presenta para la explotación por los usuarios según se concluye cada iteración. La capacitación del personal es también en etapas. Para cada iteración la capacitación del personal afectado, será en sesiones de grupo de cinco a diez (5-10) usuarios por grupo con una duración máxima de 3 horas y el Instructor realizara 2 sesiones por día. Cada grupo de usuario asistirá a la sesión de capacitación una vez, con lo que en un plazo de una semana se concluirá de capacitar a todo el personal involucrado.

Finalmente se pueden resumir los resultados:

1. La implementación de un Data Warehouse, en la institución analizada, tiene una alta probabilidad de éxito.
2. El impacto del uso del Data Warehouse seguirá una trayectoria de arriba abajo iniciando en los niveles de los tomadores de decisiones y se reflejara en un cambio en la gestión de la empresa.
3. El personal necesario para construir el Data Warehouse es un grupo base reducido con ayuda de las demás áreas del departamento de tecnología de la información. Una persona puede realizar más de una función.
4. Los costos de la dotación de personal no constituyen el mayor rubro de costos del proyecto.

C. Factibilidad Económico-Financiera

Para efectuar el estudio de Factibilidad económico-financiera resulta imprescindible conocer la totalidad de los costos en que incurrirá la empresa para incorporar el nuevo sistema analítico: el Data Warehouse. También se debe contemplar el incremento de los costos debido a un aumento en las cargas de trabajo sobre la estructura, debido al funcionamiento del sistema luego de su puesta en marcha.

1. Cuantificación de la inversión

Todos los costos que aparecen en el presente apartado son expresados en Dólares Estadounidenses (US, \$), por motivos del deslizamiento de la moneda nacional con respecto al mismo. Además de los costos de hardware y software que representan el grueso de la inversión en el proyecto, será necesario conocer los siguientes costos complementarios:

- **Dotación:** Son los recursos humanos que intervendrán como grupo base en el desarrollo del proyecto de Data Warehouse, es preciso detallar el sueldo que percibirán, para la totalidad de puestos y vacantes, se puede establecer por proyecto o por periodo de tiempo.

Tabla IV-5.- Costos de Personal

Cargo	Cant (unit)	Tiempo (mes)	Sueldo (mes)	Costo Total
Analista de Negocios	1	1.50	\$600.00	\$900.00
Arquitecto Técnico	1	2.50	\$600.00	\$1,500.00
Instructor	1	0.50	\$450.00	\$225.00
DBA	1	1.50	\$800.00	\$1,200.00
Desarrollador de Reportes	2	1.50	\$500.00	\$1,500.00
Desarrollador de rutinas ETL	2	2.00	\$600.00	\$2,400.00
Director de proyecto	1	4.75	\$900.00	\$4,275.00
Soporte Técnico	1	0.50	\$450.00	\$225.00
			TOTAL	\$12,225.00

Cronograma del proyecto

Las tareas que conforman el proyecto y que tendrán que ser desempeñadas por el personal tienen establecidas una duración estimada para cada tarea, con los que se calculan los costos en que se incurrirá en concepto de personal para el proyecto de Data Warehouse. La realización de dichas tareas tiene un carácter secuencial y en otras de simultaneidad, ya sean estas dependientes o independientes y según el orden de las tareas predecesoras ya finalizadas. En la Tabla.IV-6 se enumeran las tareas a realizar en el proyecto y su duración mientras que en los anexos se puede encontrar un Diagrama de Gant del Cronograma del proyecto.

Tabla IV-6-. Cronograma

ID	Nombre Tarea	Duración
1	Definición del Proyecto	5 días
2	Planeación y Administración del proyecto	140 días
3	Definición de los Requerimientos del usuario	6 días
4	Modelado Dimensional	22 días
5	Análisis de las fuentes de Datos	10 días
6	Diseño de la Arquitectura Técnica	18 días
7	Selección de Productos	10 días
8	Instalación de Productos	12 días
9	Diseño físico de la Base de Datos	21 días
10	Implementación física de la Base de Datos	30 días
11	Diseño y Desarrollo de la preparación de datos	40 días
12	Población y Validación de Datos	18 días
13	Ajustes de Desempeño	16 días
14	Especificación de las aplicaciones de usuario final	5 días
15	Desarrollo de las aplicaciones de usuario final	33 días
16	Prueba del sistema completo	3 días
Total días		140 días

El total de días se estableció utilizando como período laborable base un día hábil que consta de nueve horas diarias de trabajo de Lunes a Viernes, la semana hábil comprende un total de cuarenta y cinco horas hombre para un trabajador cualquiera.

- **Elementos Físicos:** Existen requerimientos para el desarrollo del proyecto que son de carácter físico y forman parte de la infraestructura que servirá de soporte al Data Warehouse, su consideración en la factibilidad económica-financiera ayuda a precisar los costos totales del desarrollo del proyecto. Los elementos en la categoría disponible significa que existen a la disposición dentro de la empresa y no deben ser adquiridos.

Tabla IV-7.- Costos de Infraestructura

Concepto	Elemento	Cant (unit)	Disponible (S/N)	Costo Unitario	Costo Total
Mobiliario	Rack para servidor	1	S	\$180.00	\$180.00
	Mesas para impresoras	1	N	\$20.00	\$20.00
Instalación eléctrica	Regulador de tensión	1	N	\$60.00	\$60.00
	Ramal trifásico	1	S	\$15.00	\$15.00
	Puesta a tierra (Jabalinas de cobre)	1	S	\$8.00	\$8.00
				TOTAL	\$283.00

A estos costos se deberán agregar los que tiene el estudio técnico: Costo de Hardware, Software, más Cargos por instalación y Servicio Técnico.

Tabla IV-8.- Costos Hardware y Software

Elemento	Cant	Disponible (S/N)	Valor Unit	Total
DBMS	5	N	\$188.48	\$942.40
Modelador de Datos	1	S	\$2,900.00	\$2,900.00
Herramienta de Acceso a Datos	1	S	\$468.00	\$468.00
Sistema Operativo	5	N	\$33.00	\$165.00
Ofimática	1	S	\$130.00	\$130.00
Cargos por instalación	1	N	\$100.00	\$100.00
Servidor Hewlett Packard	1	N	\$9,923.00	\$9,923.00
Servidor Compaq	1	N	\$7,931.55	\$7,931.55
Impresoras	2	N	\$747.00	\$747.00

Para establecer estos costos se consultaron a tres proveedores en los cuales el tiempo de vigencia de los presupuestos es de 15 o 30 días, la forma de pago puede ser en efectivo, con cheque o con tarjeta de crédito, todos entregan e instalan los equipos en un período menor a 20 días laborables y su disponibilidad es inmediata, las aplicaciones tienen soporte técnico por el fabricante una vez adquirida su licencia a como lo estipula la ley de derechos de autor vigente, mayores detalles en el apartado de factibilidad legal.

Con todos estos datos se procederá a cuantificar la inversión, evaluar las opciones y determinar la modalidad de contratación más ventajosa.

2. Análisis Comparativo

En primer lugar se establecerá el monto de inversión inicial, que incluye gastos por única vez, y luego, el cargo mensual del nuevo centro de costos.

El cargo inicial del proyecto para que este pueda ser puesto en marcha, está compuesto por los conceptos de Dotación de Personal, Hardware, Software, Mobiliario, Instalación Eléctrica y Sistema de Seguridad, los costos de elementos físicos y hardware considerados para los elementos que están disponibles en la empresa se sustituyen por valor cero (\$0.00) para el análisis comparativo, en el caso del software se establece valor nulo para el software ya licenciado, así el costo total de adquisición al que ascienden las opciones A y B aparecen a continuación:

Tabla IV-9.- Costos Comparativos de opciones de compra

Concepto	Opción A	Opción B
Dotación de Personal	\$11,725.00	\$11,725.00
Hardware	\$10,670.00	\$8,678.55
Software	\$1,107.40	\$1,107.40
Mobiliario	\$40.00	\$40.00
Instalación Eléctrica	\$60.00	\$60.00
TOTAL	\$23,602.40	\$21,610.95

Los costos mensuales estimados, están dados por los conceptos de suministros para los equipos y del personal de mantenimiento y estos están en relación directa con el uso que se haga del Data Warehouse, un estimado de dichos costos aparece en la siguiente tabla:

Tabla IV-10.- Costos mensuales del proyecto

Concepto	Valor total
Suministros	\$30.00
Personal de Mantenimiento	\$400.00
TOTAL	\$430.00

3. Análisis Costo Beneficio

Los beneficios que se obtienen del data Warehouse no son cuantificables en términos monetarios, pues su valor reside en el soporte a la toma de decisiones.

Cuando el data Warehouse entre en producción existirá una variedad de maneras en que los usuarios serán capaces de extraer información del sistema, ya que los usuarios tendrán la capacidad de efectuar reportes para sus fines específicos o darle tratamiento a información que ciertamente es confiable y veraz, para utilizarla de base en la toma de decisiones.

Los principales beneficios derivados de la construcción de un Data Warehouse son:

- La obtención de una plataforma que consta de una mejor calidad de datos y de más fácil acceso, que solucione la problemática de los niveles de decisión de la empresa, de no poder acceder a la información de su interés al no estar ésta disponible.
- Hacer que el proceso de toma de decisiones sea más acertado al estar basado en hechos y no en la intuición o en datos de dudosa calidad.

- Facilitar un acercamiento a los clientes al descubrir aspectos desconocidos de ellos como su patrón de comportamiento, preferencias y quejas.
- La construcción de un ambiente donde no es necesario un basto conocimiento técnico para realizar consultas y elaborar reportes que satisfagan las necesidades de información, relevándose así la utilización de parte del personal técnico de la organización para satisfacer necesidades específicas de los usuarios.
- Proveer un repositorio de datos “limpios” de los sistemas de procesamiento transaccional contra el que se pueden realizar consultas sin necesariamente requerir un cambio en los sistemas transaccionales. Estos datos pertenecen a un intervalo de tiempo mucho mayor que el que pueden manejar con eficiencia los sistemas transaccionales haciendo posible análisis de tendencias y de otros tipos.
- Ganar competitividad y una mejor posición en el mercado con la implementación de un sistema estratégico que le dé ventajas competitivas reales a la organización.

El valor monetario de una decisión bien tomada es difícil de conocer en avance, pero experiencias similares alrededor del mundo aseguran que si un data Warehouse es bien implementado y se encuentra en explotación el retorno a la inversión es halagador.

El costo total estimado para la primera fase de este proyecto es de **\$23,602.40** para la primera opción, y de **\$21,610.95** para la segunda, pero los beneficios obtenidos indudablemente justifican tal inversión dado que la organización funcionará de una manera más eficaz y ágil, produciendo un alto retorno a la inversión.

En este estudio se han ahorrado las consideraciones sobre la seguridad física de los equipos de cómputo que albergarán al Data Warehouse, debido a que se encuentra fuera de los objetivos y alcance de este estudio. Se da por sentado que al ser el Data Warehouse un activo estratégico de una organización las medidas para su protección serán igual o incluso más meticulosas y cuidadosas que las que se tomen para proteger el entorno transaccional. Las medidas que apliquen en cada empresa estarán en dependencia directa de su situación organizacional.

Para el caso de la empresa objetivo, la localización física del Data Warehouse será en las instalaciones del centro de cómputo de la empresa, el cual fue previamente acondicionado con las condiciones y políticas de seguridad que amerita el entorno

tecnológico operativo de la financiera, motivo por el que no se considera como parte de los costos del proyecto los costos de seguridad física.

Consideraciones sobre la Contratación de un Servicio Externo de Computación

La opción de outsourcing para la elaboración de un proyecto de DW para la empresa en cuestión no presenta una ventaja en costo beneficio, debido al alto costo de los profesionales expertos en la tecnología de Data Warehousing, los cuales tendrían que ser reclutados en el exterior. Esto trae consigo además de costos de desarrollo, problemas con el mantenimiento.

En cambio es posible contratar al profesional que desempeñará el rol de Director de Proyecto para que guíe y coordine al grupo base conformado de recursos humanos pertenecientes a la empresa, los cuales tienen la ventaja de conocer ya la situación y contexto de la organización. Además, se evita la problemática con el mantenimiento.

D. Factibilidad Legal

Se deben considerar conceptos legales externos a la institución, como parte de la factibilidad de un proyecto, para ello se toman en cuenta las leyes vigentes en este momento en el país que pueden afectar el desarrollo de un proyecto de Data Warehouse, por lo que es necesaria su consideración. Las leyes a discutir son:

La ley de Derechos de Autor

Ley de Derechos de Autor aprobada por la Asamblea Nacional de la República de Nicaragua y publicada por el diario oficial “La Gaceta” en su edición del treinta y uno de Agosto de mil novecientos noventa y nueve.

El marco legal de esta ley afecta directamente al proyecto ya implica la utilización de recursos de informáticos de Software como parte integrante del proceso de desarrollo del Data Warehouse, los cuales están protegidos por la ley de derechos de autor y derechos de propiedad.

El corolario, de lo expuesto en la ley se traduce en disposiciones para la adquisición y legalización de todo producto informático que deba utilizarse y que se encuentre protegido por la Ley de Derechos de Autor para no ser ilegal o sujeto de duda.

Según la Ley de Derechos de Autor en su arto.2 párrafo 2.26 se define un programa de cómputo como un “conjunto de instrucciones expresadas mediante palabras, código, gráficas, diseño o en cualquier otra forma, que al ser incorporadas en un dispositivo de lectura autorizada, es capaz de hacer que un ordenador, aparato electrónico o similar capaz de elaborar informaciones y que ejercite determinado resultado. También forma parte del programa su documento técnico y sus manuales de uso”.

Luego en el arto.13 de esta ley se establece que “están protegidas por la ley todas las creaciones originales y derivados literarios, artísticos, científicos independientemente de su género, mérito, forma actual o futura tales como [...], programas de cómputo, sean éstos programas fuentes o programas objetos y cualquiera que sea su modo o forma de expresión”.

Capitulo VI se expresa otra regulación importante de ser tomada en cuenta: La Transmisión de Derechos Patrimoniales, en el arto.52 expresa que “cuando se trate de una obra realizada por un actor por cuenta de persona natural o jurídica en el marco de un contrato de trabajo y de su empleo, salvo disposición en contrario del contrato, el primer titular de los derechos morales y patrimoniales será el autor, pero los derechos patrimoniales sobre dicha obra se considerarán transmitidos al empleador en la medida justificada por las actividades habituales del empleador en el momento de creación de la obra”.

La Ley de Derechos de Autor obliga a todo aquel que haga uso de programas de cómputo a adquirir las debidas licencias, para no infringir dicha ley, ni ser objeto de multas por su desobediencia.

El presente proyecto requiere de la adquisición de diversos programas de cómputo para su desarrollo. En la mayoría de los casos ya se cuenta con la licencia requerida para dichos programas de cómputo, pues fueron previamente adquiridos por la institución, pero en algunos otros se deberán adquirir durante el desarrollo.

Conclusión:

- La Ley de Derechos de Autor involucra que todas las aplicaciones a utilizarse en el proyecto deben de ser debidamente licenciadas y registradas. En caso contrario la

empresa se verá expuesta a la aplicación de multas sobre el ámbito del cumplimiento de tal normativa.

- Los derechos de autor sobre el Data Warehouse desarrollado son propiedad intelectual exclusiva de la empresa que lo desarrolle.

Otras Consideraciones:

Sigilo Bancario

- La información personal de clientes que se pueda alojar en el Data Warehouse, es para uso exclusivo de la empresa y no puede ser divulgada o vendida a terceros en carácter detallado e individual según lo establece la regulación de sigilo bancario sobre las instituciones financieras.

Daños a Terceros

- La empresa no se hace responsable por los inconvenientes que pueda ocasionar a sus clientes si estos resultaran seleccionados como público de campañas publicitarias basándose en la información que la empresa tiene sobre sus clientes.
- La empresa no se hace responsable de la información errada que el cliente pueda haber proporcionado, ni de las consecuencias que de este hecho se deriven.

E. Sugerencia para el proyecto

Es recomendable la selección de la opción A para la inversión ya que es muy atractiva desde el punto de vista de desempeño versus monto invertido, sin embargo la opción B es una alternativa un poco más económica, sin que ocurra una disminución muy significativa en el desempeño.

La opción de outsourcing para este proyecto debe ser estudiada con cuidado, para no elevar significativamente el monto total a invertir en el proyecto. Es importante, para asegurar el éxito del proyecto definir un adecuado programa de capacitación para el personal de la institución que intervendrá en el mismo.

V. GUÍA METODOLÓGICA PARA LA CONSTRUCCIÓN DE UN DATA WAREHOUSE

La esencia del Data Warehousing en sí, es una metodología no una tecnología, es una metodología para alcanzar soluciones a los requerimientos de información del negocio.

La guía metodológica no pretende ahondar o explicar los conceptos abordados con mayor o menor amplitud en el marco teórico sino más bien componer con los conceptos destilados un nuevo enfoque reconciliado que dirija a los noveles en su camino para determinar la estrategia de Data Warehouse que mejor se adapte a sus necesidades.

Para determinar que metodología o enunciado de metodología conviene incluirse en la guía, además de ser acorde al objetivo del desarrollo, debe de tomar en cuenta los siguientes propósitos que sirven de base para mantener viable un proyecto de desarrollo Data Warehouse:

1. El retorno a la inversión debe ser lo más pronto posible. Es esencial que al invertir en un entorno analítico la empresa vea resultados del esfuerzo y los utilice en el menor período de tiempo.
2. El Data Warehouse debe pasar a ser un activo estratégico de la organización rápidamente, los beneficios obtenidos del Data Warehouse deben poner en ventaja estratégica a la empresa.
3. Los costos de mantenimiento deben reducirse al mínimo. Si los costos de mantenimiento son muy elevados, producto de una mala concepción y planeación del proyecto de data Warehouse, este no será rentable, ni seguirá siendo viable y fracasará.

En este punto es importante destacar lo siguiente:

- Al iniciar un proyecto de Data Warehousing no se debe olvidar establecer una base sólida para el desarrollo, un marco de referencia para la construcción del Data Warehouse: la arquitectura de la implementación, como la llama Kimball.
- En el proceso de construcción se puede distinguir dos etapas principales: además de la definición de una arquitectura Data Warehouse, la construcción de los incrementos de Data Warehouse.

La guía entonces, está estructurada en etapas más que en pasos, pues los pasos pueden variar en cada proyecto, pero no así las etapas que este involucra.

1. Etapa 1: Análisis y Comprensión del Entorno Analítico
2. Etapa 2: Definición de una Arquitectura
3. Etapa 3: Enriquecimiento de Datos
4. Etapa 4: Construcción en Incrementos, Data Marts

El desarrollo de la presente guía, ha sido acompañado de su puesta en práctica en una institución financiera de tamaño mediano que opera en el ámbito nacional, bajo la modalidad de un proyecto piloto para la introducción de una plataforma analítica que brinde ventajas competitivas a la organización. El desarrollo mencionado servirá entonces para propósitos ilustrativos de cada etapa de la guía.

En el entorno operativo de dicha institución financiera, en la cual se desarrolla el proyecto piloto que servirá de ejemplificación, se presentan ciertas condiciones organizacionales que influye y determinan en cierta medida los ejemplos que se detallan para ilustrar la guía. Con miras a lograr la adecuada comprensión de los ejemplos presentados se debe considerar los siguientes elementos:

1. La institución financiera consta de más de una sucursal. Todas las sucursales son totalmente operativas.
2. El principal rubro de operación de la financiera lo constituyen los créditos otorgados a los clientes. Los clientes disponen de una variedad de productos financieros entre los cuales pueden escoger el que mejor les convenga a sus intereses.
3. Todo crédito gestionado está a cargo de un analista de crédito que realiza los trámites que la empresa requiera e interactúa directamente con el cliente.
4. La plataforma tecnológica operativa está compuesta por Microsoft Windows 2000 Server como Sistema Operativo de Servidores y Microsoft Windows XP en las estaciones de trabajo, Microsoft SQL Server 2000 es el Sistema para Gestión de Bases de Datos, una aplicación que consta de siete módulos desarrollada con Visual Fox Pro 7.0 denominada LFS como sistema transaccional y una aplicación en Clipper encargada de administrar la planilla.

Una vez establecido el marco de trabajo del presente estudio se abordan a continuación las etapas de la guía con ejemplificación de su aplicación en la institución financiera.

A. Etapa 1: Análisis y Comprensión del Entorno Analítico

Como en cualquier proyecto y mucho más en aquellos de gran envergadura como un Data Warehouse, al iniciarlo hay que establecer sólidas bases teóricas que permitan vislumbrar un plan de acción efectivo y que minimice el riesgo de fracaso y dominar, además los elementos y herramientas que intervendrán en el desarrollo de dicho proyecto.

El desarrollo de un Data Warehouse no debe ser iniciado mientras no se tenga suficiente dominio sobre la filosofía detrás del proyecto, la forma en la que se puede realizar y cuales son los objetivos que se pretende alcanzar, este es el mapa que dirige el camino del proyecto. Para un Data Warehouse es prioritario *reconocer las necesidades analíticas que tiene una organización y establecer los objetivos que se persiguen al desarrollar un Data Warehouse en la organización en la que se desarrolle*. Este es el primer paso a seguir, el análisis y la comprensión del nuevo entorno, el entorno analítico.

Esta etapa se debe cubrir con miras a cumplir los siguientes enunciados:

1. Entender las principales diferencias entre un Data Warehouse y un sistema transaccional.
2. Establecer y delimitar la diferencia entre un Data Warehouse Corporativo y los Data Marts, asimilar el objetivo de los distintos enfoques y su aplicación a la situación de la empresa.
3. Reconocer que patrones han sido exitosos en la industria y los errores que se han cometido.
4. Identificar y comprender las distintas herramientas a utilizar en las distintas etapas del desarrollo, asimilar las técnicas de modelado utilizadas en el Data Warehouse o en los Data Marts y otros instrumentos.
5. Identificar y establecer qué es lo que los directores de la organización esperan obtener con la implementación del proyecto.
6. Determinar la viabilidad del proyecto mediante la realización de un Estudio de Factibilidad.

Al iniciar el proyecto en la institución financiera se tomó en consideración que requisitos expuestos en los primeros cuatro enunciados se cumplen con la comprensión y el dominio logrado sobre la teoría y metodologías de Data Warehousing expuestas en el marco teórico, mediante el cual se analizaron y comprendieron los distintos enfoques de un proyecto Data Warehouse, así como las características y beneficios que un entorno analítico brinda a una empresa.

Luego, para determinar la situación de la empresa e identificar las necesidades de información de los encargados de administrar el negocio, se utilizó como herramienta de recopilación de datos, entrevistas aplicadas a nivel gerencial dentro de la empresa [Ver detalles de entrevistas en Anexo D].

La realización de entrevistas tiene un doble propósito para fines del proyecto, no solo permite obtener información de primera fuente y asegurar que se obtienen impresiones y necesidades que no están plasmadas en ningún documento, sino que además introduce el concepto de Data Warehouse en aquellos que tienen poder de decidir y que sería muy útil tener como apoyo. Así se logró averiguar sobre los siguientes aspectos medulares:

1. Objetivos de las distintas áreas involucradas en el proyecto.
2. Prioridades de cada una de ellas.
3. Factores críticos de éxito.
4. Decisiones más relevantes y frecuentes.
5. Información Clave para dirigir el negocio.
6. Indicadores del estado del Negocio.
7. Reportes de uso más frecuente.
8. Información crítica faltante.
9. Frecuencia de uso de información.

La alineación de la estrategia de negocios con la estrategia del Data Warehouse y la elección de una metodología sólida con los correctos entregables son aspectos críticos de una estrategia de Data Warehouse, por eso es importante que la mayor cantidad de personal que explotará al Data Warehouse sea involucrada en este proceso de recopilación de información.

En esta etapa del proyecto aún no se puede conocer cual será el objetivo de la primera entrega del proyecto, por lo que es importante tomar en cuenta todas las áreas de la organización. En la institución financiera donde se aplicó esta guía, las entrevistas fueron efectuadas en las áreas de Crédito, Finanzas, Mercadeo, Sistemas y la Gerencia General.

Es en esta etapa inicial cuando es adecuado realizar un Estudio de Factibilidad para estudiar la viabilidad del proyecto, dicho estudio para este proyecto fue presentado previamente, en el apartado anterior. Vale la pena aclarar que dicho estudio refleja un proyecto completo de Data Warehouse con sus costos reales y ejecutado por profesionales del área de Data Warehousing, y servirá de marco de referencia para determinar cuales serían los costos en los que incurriría un proyecto de esta envergadura en una empresa de tamaño mediano en el contexto nicaragüense.

B. Etapa 2: Definición de una Arquitectura.

Arquitectura Data Warehouse

Una Arquitectura Data Warehouse establece el marco de trabajo, estándares y procedimientos a seguir para la construcción de un Data Warehouse a nivel empresarial. El objetivo de las actividades de la arquitectura es simple, *integrar al Data Warehouse las necesidades de información empresarial*. Ningún proyecto de Data Warehouse a ninguna escala puede ser exitoso si la base sobre la que se erige no es sólida, sin que represente las necesidades de la organización y sin que permita el mejor aprovechamiento de los recursos disponibles.

Los tres autores abordados brindan amplias descripciones sobre como debe ser una arquitectura de Data Warehouse, pero es Ralph Kimball quien en este aspecto propone la arquitectura estructurada y de más sencilla comprensión, que previamente se introdujo en el Marco Teórico.

Para quienes nunca han participado en una proyecto de las características de un Data Warehouse, es difícil conceptualizar claramente el objetivo que se persigue cuando se indica que se debe establecer una arquitectura sólida, contando solamente con las referencias teóricas y abstractas que detallan los otros dos autores. Así que se escoge la arquitectura propuesta por Kimball debido a su pragmatismo, aplicabilidad y más fácil

concepción lo que lo convierte en un recurso valioso para detallar los distintos elementos que componen la arquitectura de un Data Warehouse.

Esta arquitectura y sus resultados pueden, sin embargo, ser alcanzados utilizando algunas técnicas muy apropiadas apuntadas por los otros autores.

Resultados de la Arquitectura

De los elementos podemos destacar como principales resultados de la aplicación de la arquitectura de Data Warehouse los siguientes entregables, previamente abordados con mayor detalle:

1. El modelo de datos fuente.
2. El modelo de datos conceptual Data Warehouse.
3. Arquitectura tecnológica Data Warehouse.
4. Estándares y procedimientos para desarrollo y mantenimiento del Data Warehouse.
5. El plan de implementación incremental para el Data Warehouse.
6. Definición a distintos niveles de los componentes del Data Warehouse.

Los modelos de datos proveen una estructura para identificar, nombrar, describir y asociar los componentes de una base de datos. En general se necesitan modelos de datos para los datos fuente como para los datos seleccionados para existir en el Data Warehouse y la estructura que estos presentarán en el nuevo entorno.

El primer entregable de esta etapa consiste en el establecimiento del *modelo de datos fuente*, en este aspecto Inmon introduce el análisis al Diagrama Entidad Relación de la empresa y a los distintos elementos del desarrollo de sistemas transaccionales para identificar elementos de interés del proyecto de Data Warehouse.

Para los propósitos de desarrollo del proyecto en mención, esto es muy adecuado y como resultado de su ejecución se pudo comprobar que el estudio inicial de modelo de datos de la organización permite que se obtenga el dominio necesario del mismo para poder desenvolverse con soltura en las etapas sucesivas del mismo.

El modelo de datos de una organización puede estar más o menos normalizado y en ocasiones presentar cierto grado de redundancia o derivación de datos con el objetivo de suplir ciertas necesidades de información. Es aconsejable identificar estos aspectos para

seleccionar solamente los datos originales ya que estos son más fiables que los derivados y son una buena pista para reconocer información importante para la empresa.

En la Ilustración V-1 se puede apreciar una muestra del Modelo de Datos del Sistema Transaccional de la financiera sobre la que se aplicó la guía y en la cual se distingue claramente que el modelo no está normalizado. Aquí cabe hacer un señalamiento sobre la presencia de llaves inteligentes (el campo ccodcta es compuesto) y sobre los nombres de campos nada descriptivos (csececo significa sector económico), propios de los sistemas transaccionales, pues facilitan la programación de aplicaciones, pero que impide deducir el significado de los datos contenidos en el campo.

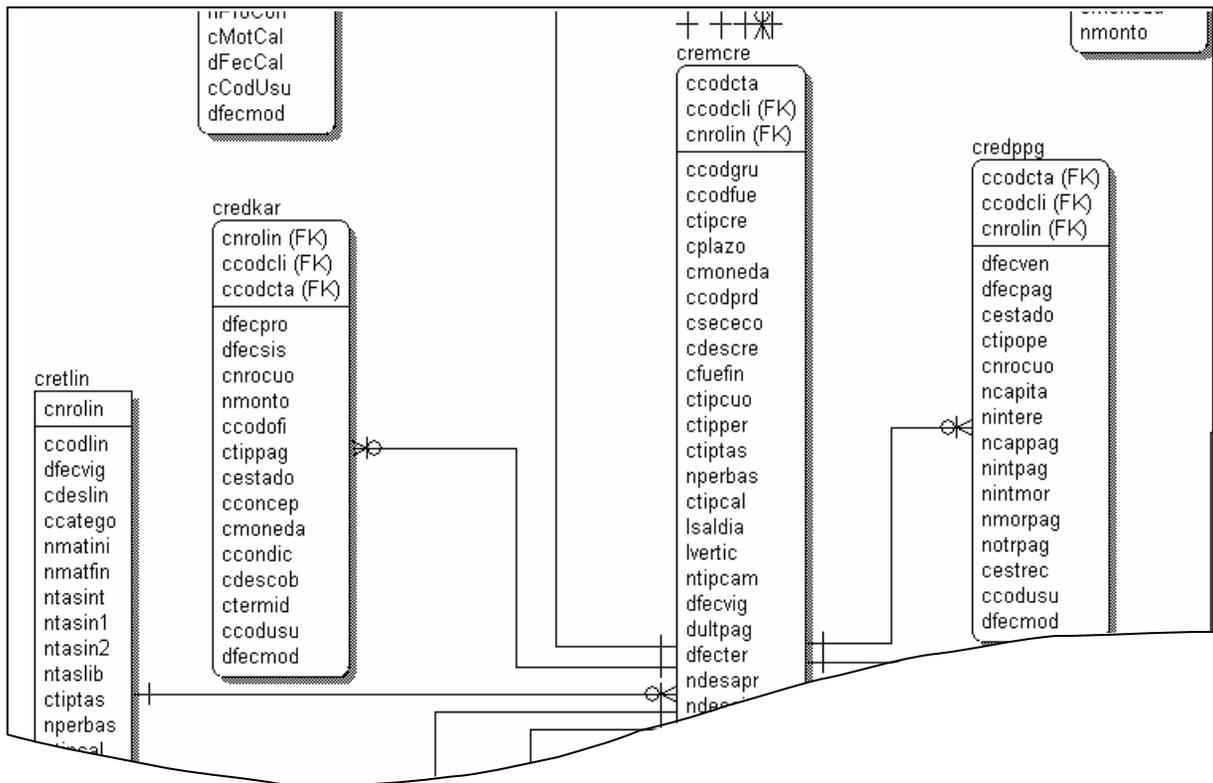


Ilustración V-1 Modelo de Datos de una Financiera

El modelo de Datos presentado en la Ilustración V-1, se modeló utilizando la herramienta CASE Erwin/ERX 3.5.2 debido a que entre sus funcionalidades permite

efectuar Ingeniería en Reversa sobre la base de datos objetivo y determinar su modelo de datos, también influyó su simple, pero clara forma de presentar el diagrama del modelo de datos, para una interpretación más fácil.

El análisis realizado en esta etapa permitió además, identificar algunas problemáticas del modelo de Datos inicial de la empresa que deben ser solucionadas como por ejemplo, presencia de tablas temporales para realizar consultas y redundancia que podría afectar la integridad entre otras. Esto es originado porque un único entorno está supliendo las necesidades transaccionales y analíticas, situación que cambiará con la presencia de un Data Warehouse. El análisis también permitió que se establezcan las bases de cambios futuros para adaptar los sistemas transaccionales al nuevo entorno que se introduce.

En cuanto al modelo de datos del Data Warehouse se cuenta con dos posibles caminos a seguir. A continuación algunas consideraciones sobre ambas opciones:

- **Modelado Relacional:** uso de Diagramas Entidad Relación (ERD), el ERD está asociado con las leyes de normalización, lo que no es adecuado para el desempeño de las consultas. Otro riesgo asociado con el uso de ERD es que se vuelve un poco difusa la diferencia entre la estructura de un Data Warehouse y un sistema transaccional. También se puede mencionar que un Data Warehouse relacional plantea la necesidad de contar con una plataforma extra para los data marts que sea capaz de utilizar tecnología relacional y multidimensional. No obstante, siempre se debe recordar que en situaciones donde las cantidades de datos a manipular sean muy grandes esta solución permite manipular cantidades ingentes de datos sin disminuir dramáticamente el desempeño, en comparación a si se utiliza el modelado dimensional.
- **Modelado Dimensional:** uso de Esquemas Estrellas o de Esquemas Copo de Nieve. Una debilidad de este modelado es que la cantidad requerida de almacenamiento crece exponencialmente, conforme se agregan nuevas dimensiones. El Data Warehouse se concibe como una composición de Data Marts unidos por dimensiones conformadas por lo que no introduce una nueva plataforma, pero requiere de distintos procesos de población de Datos para distintas estrellas, lo que conlleva a dificultad en mantener concordantes los datos cuando estos aparecen en Data Marts que cumplen distintas funciones.

Para decidir que tipo de modelado es más adaptado al contexto nacional se tomó en cuenta además que la mayor parte de las empresas objetivo de la guía no generan cantidades masivas de datos; son de tamaño mediano a grande, operan sólo dentro del país e implementaron sus sistemas operacionales hace menos de una década, esto hace que los volúmenes de datos que albergue un Data Warehouse construido en este ambiente, sean modestos en escala comparativa. Lo anterior también es responsable de una disminución en la consideración de la duración del proyecto y de las etapas de cada uno.

Por lo tanto es adecuado aprovechar las ventajas del modelado dimensional y es está la opción que se aplica en el desarrollo del proyecto de la institución objetivo de la aplicación de la guía.

En síntesis, en lo correspondiente al tipo de modelado del Data Warehouse la solución apropiada para el propósito de esta guía es modelar los datos dimensionalmente utilizando el esquema estrella siempre y cuando las cantidades de datos a manipular no impliquen un volumen de datos tan grande que vuelva más complicada la solución que el problema, al evitar se alcance uno de los principios del Data Warehouse su disponibilidad rápida de información para soporte a la toma de decisiones y se vuelva lento el acceso a la información, en el caso de que la cantidad de datos se prevea crezca en demasía se puede contemplar el modelado copo de nieve o según el caso el modelado relacional.

Una vez elegido como método el modelado dimensional se debe elegir que Data Mart será el primero en desarrollarse, para esto se utiliza la matriz de arquitectura del Data Warehouse, que se describió en el Marco Teórico. La matriz de Data Marts muestra la relación entre los posibles Data Marts y las dimensiones. Cualquier columna con más de una “X” implica que ésta dimensión debe ser conformada a través de múltiples Data Marts para que pueda caber en el Data Warehouse.

Basándose en los resultados obtenidos de las entrevistas realizadas, se pudo concluir que los tópicos de mayor interés para la financiera objeto son: Colocación, Recuperación y Mora, los cuales corresponde cada uno a un Data Mart y fueron marcados con “**” en la matriz de la Tabla V-1, ya que fueron seleccionados para ser implementados en el primer incremento debido a la naturaleza crítica de estos tópicos para la empresa. Estos luego se pueden consolidar como un primer incremento de un único Data Mart al que se denominará “Gestión de Créditos”.

Tabla V-1 Matriz de Data Marts

Data Mart	Fecha	Fecha_Vencimiento	Fecha_Final	Hora	Analista	Cliente	Producto	Sucursal	Tasas	Convenio	Fuente Finan	Garantía	Cuenta LM	Modo Pago	Circunstancia
**Recuperaciones	X	X	X	X	X	X	X	X	X	X		X		X	X
**Colocaciones	X	X	X		X	X	X	X	X	X	X	X		X	
**Mora	X	X				X	X	X	X		X				X
Rentabilidad	X					X	X	X	X	X	X		X		
Productividad	X				X			X							
Promociones de Mercadeo	X					X	X								
Cobranzas	X					X		X				X			
Embargos	X					X		X				X			
Quejas	X				X			X							

Un aspecto importante de utilizar la matriz de la arquitectura es que permite realizar el modelado de las dimensiones que se incluyan en el primer incremento, de manera que sea conforme con todos los data Marts que la utilizarán. En la matriz de la Tabla V-1 es obvio que los temas elegidos a implementar en un único Data Mart comparten muchas dimensiones, lo cual se debe a que son temas complementarios y por eso se consideró adecuado y beneficioso para la empresa que su desarrollo se consolide en un único incremento. En adición se hace patente que dimensiones como *Fecha* deben ser diseñadas con cuidado, pues de ello dependa que los siguientes incrementos las puedan utilizar y se relacionen sin problemas para juntos consolidar el Data Warehouse Corporativo de la organización.

A continuación, una vez elegido el Data Mart, el modelado puede iniciarse con una abstracción para visualizar las dimensiones y los hechos y las estructura de cada estrella y luego dar paso a las definiciones formales del modelo lógico y físico que presentará la base de datos del Data Warehouse. El modelo físico, contrasta con el modelo lógico en que este es dependiente de la plataforma tecnológica seleccionada y por lo tanto se podrá disponer o no de ciertas capacidades dependientes de la implementación.

En el desarrollo del proyecto ejemplo, la abstracción mencionada previamente, tiene el aspecto mostrado en la Ilustración V-2, en la que se muestra un esquema sencillo que permite visualizar para una dimensión su elemento clave y también el resto de información que la acompaña. La dimensión mostrada es la dimensión fecha siendo su elemento clave o de mayor detalle del día y los demás atributos de la dimensión el mes, año y otros más.

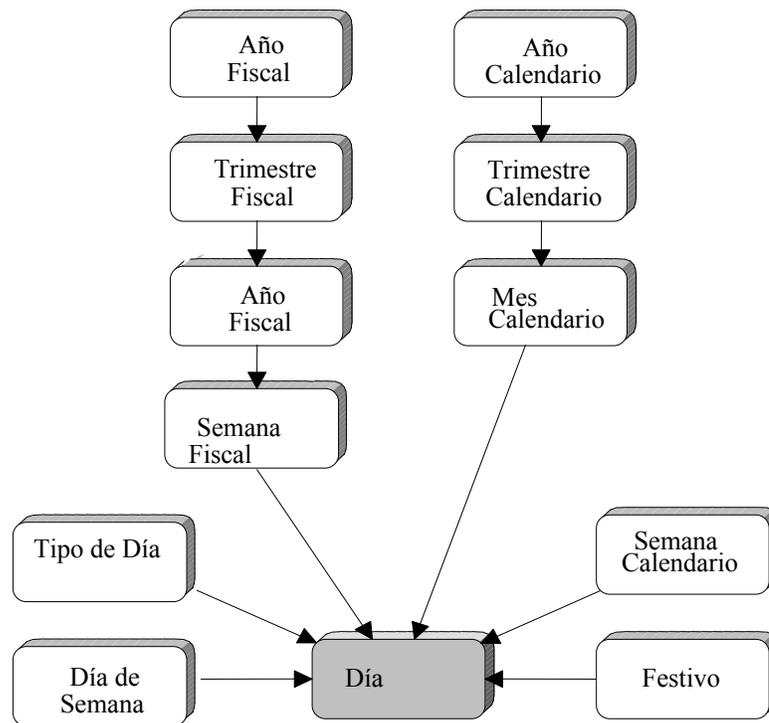


Ilustración V-2- Dimensión Fecha

Este esquema es sencillo y de fácil interpretación, sin embargo se puede utilizar otro tipo de herramientas gráficas o de esquemas para la representación siempre y cuando se logre el objetivo de tal práctica que es determinar la composición de las dimensiones, de las tablas de hechos y sus relaciones, siguiendo el esquema de cuatro pasos detallados por Kimball para el modelado dimensional y que se puede encontrar en el marco teórico.

En el desarrollo del proyecto piloto, luego que se tuvo clara la forma que debían de diseñarse el Data Warehouse para representar las necesidades de información de la compañía, era necesario pasar al siguiente nivel y definir de manera formal el diseño lógico

en un diagrama de entidades que presentará cada uno de los campos que contiene cada tabla de hecho o dimensión, el nombre lógico que recibirán y las llaves primarias de cada tabla.

El siguiente paso es transformar el Diagrama de Entidades al diseño físico que le corresponde, este diagrama además mostrará las características de los campos en dependencia del tipo de dato que debe albergar (tipos de datos y longitud, por ejemplo).

En la ejecución de este nuevo procedimiento, se hizo uso de la herramienta Oracle / Designer 2000. Esta elección obedece a las bondades de la aplicación de permitir pasar de una definición de alta abstracción como un Diagrama de Entidades, hasta el script de definición para la generación de la Base de Datos, según la plataforma destino que se haya elegido, sin más entrada de datos que la necesaria para realizar el diagrama de entidad, en el que también se indica que tipo de datos y longitud contendrá el campo. Designer también permite el trabajo en equipo, donde dos o más diseñadores pueden estar trabajando sobre las mismas entidades y elementos sin corromper el diseño. Su salida gráfica es sencilla y muy descriptiva, lo que facilita la interpretación de los diagramas.

La Ilustración V-3 muestra una parte del diagrama de entidades para la financiera ejemplo, se pueden observar algunas dimensiones importantes como Cliente, Analista, Producto y una tabla de hechos, la tabla de Abonos de Créditos que se encuentra en el extremo izquierdo.

Como se muestra claramente en la Ilustración V-3, el diseño del Data Warehouse dista mucho del diseño del sistema transaccional y como ejemplo puedo nombrar las tablas de *Cliente_Dim* y *Abono_Credito_Fact*, que corresponden a una dimensión y una tabla de hechos respectivamente.

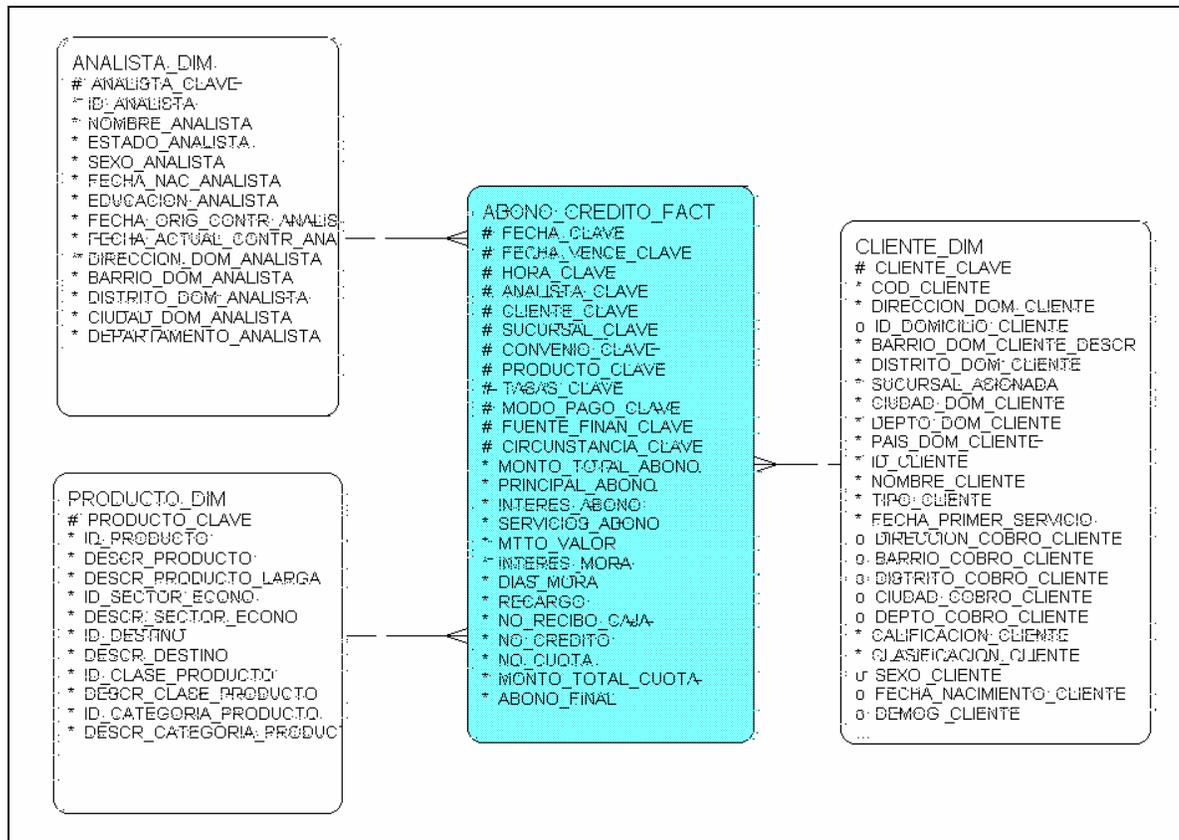


Ilustración V-3 Diagrama de Entidades

En el primer caso, siguiendo las etapas de la guía y para obtener la tabla clientes, se tuvo que tomar en cuenta además, la situación encontrada en la institución objetivo, en la cual existen muchas inconsistencias entre los datos, como problemas de duplicación de códigos para clientes, por lo que un mismo cliente puede presentar en algunos casos dos o más códigos asociados, lo que provoca se genere información errada sobre algunos indicadores de la empresa como la cartera de clientes de la institución y el promedio de créditos por cliente; para solucionar esta complicación se introdujo un nuevo código que permita que cada cliente tenga un identificador único lo que habilita que la información de la empresa sea correcta.

Para el caso de la tabla de hechos, se consideró importante por ejemplo incluir una referencia a la dimensión hora, lo cual se hace introduciendo una llave externa a dicha dimensión, por que esto permite luego obtener información como a cuales horas ocurre el mayor movimiento de caja y esto puede servir para habilitar cajas temporales en ese período o redoblar la vigilancia en la oficina y brindarle mejor servicio al cliente. También

se observó que dada la situación particular del país en donde la moneda nacional se desliza de manera diaria en comparación con el dólar, era necesario indicar de forma separada el monto abonado en este concepto y que no se incluya en ninguno de los otros conceptos, ya que de lo contrario no se podría conocer el impacto, que para la empresa tiene una política monetaria de país como ésta y también se podría dar la impresión que el total recuperado es mucho mayor que las cantidades desembolsadas a los clientes originalmente.

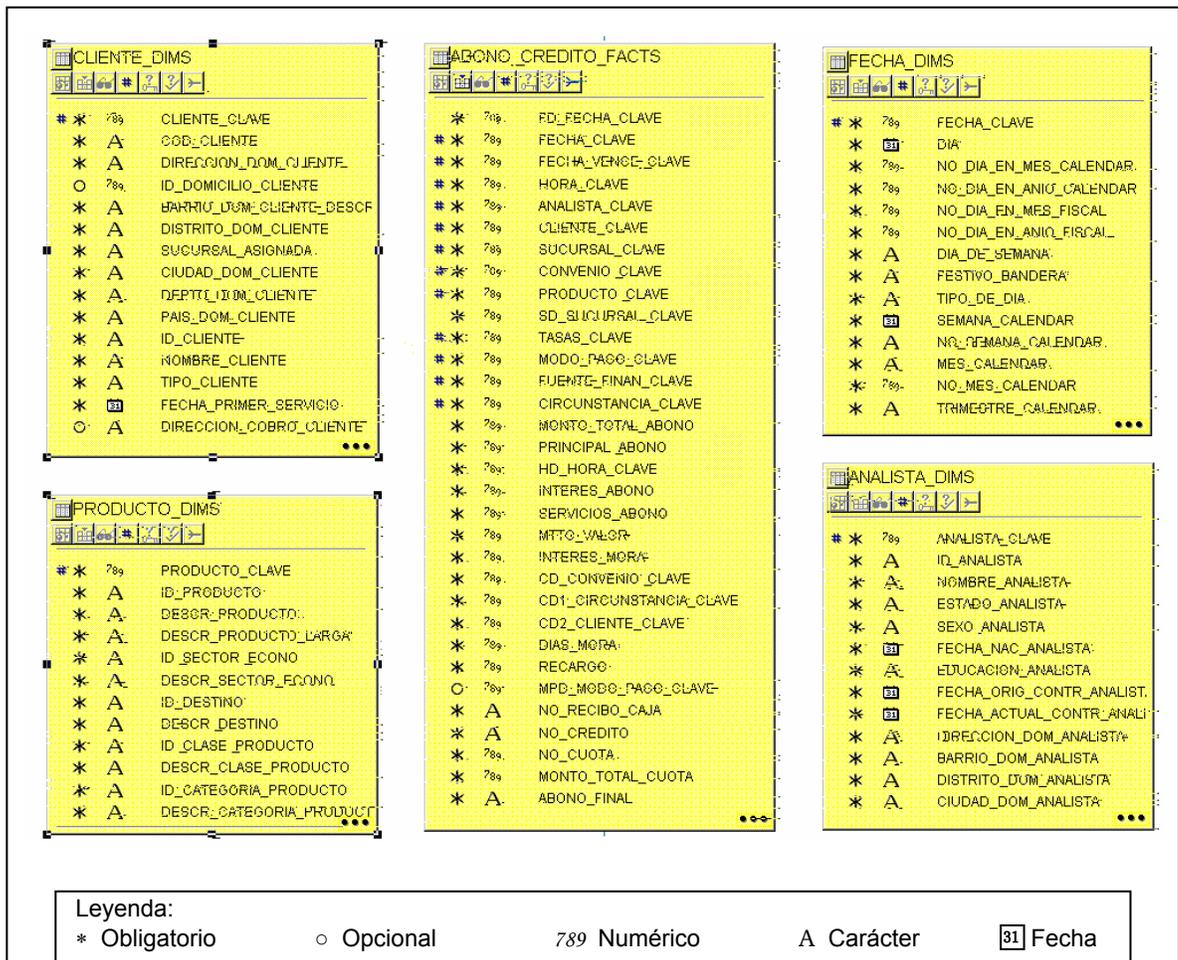


Ilustración V-4 Diseño físico de la base de datos del Data Warehouse

En el diseño físico de la financiera, como se puede apreciar en el diagrama de la Ilustración V-4, no se indican las relaciones que existen entre las tablas, esto es por una consideración al desempeño del Data Warehouse. La presencia de relaciones propia de las bases de datos normalizadas trae consigo la verificación de integridad referencial, esta verificación tiene como costo la disminución de desempeño cuando se realiza la carga del

Data Warehouse. Se ha decidido que el mecanismo de identificar relaciones para asegurar la integridad referencial no es necesario puesto que los datos que llegan al Data Warehouse son originados en sistemas transaccionales que fueron los encargados de realizar la verificación.

Para información completa acerca del modelo de datos y también del diseño lógico y físico de la base de Datos, referirse al anexo de Modelado donde aparecen todos los documentos producidos para el Modelo de Datos del Data Warehouse.

Las consideraciones de requerimientos de la arquitectura tecnológica necesaria para implementar un Data Warehouse en el contexto de las empresas nicaragüense ya descrito son expuestas en el estudio de factibilidad detallado en el anexo respectivo.

Los estándares para el desarrollo del Data Warehouse son una parte importante de su arquitectura. Sin estándares, oportunidades para re-usar no son posibles y hay riesgos de que partes del desarrollo no funcionen trabajando juntos. Los estándares aseguran que todos los involucrados en el desarrollo de proyecto cuenten con un marco que les ayude a interactuar positivamente.

Para esta implementación en la financiera, los estándares utilizados son con respecto a nomenclatura de campos, de paquetes de transformaciones de datos, y en cuanto al diseño de dimensiones conformadas. Los estándares para el desarrollo de rutinas en lenguajes de programación no se lograron concretar por carecer de este concepto en la financiera, al no contar con estándares los sistemas transaccionales, la implementación organizacional de estándares de codificación escapaba al objetivo de ejemplificación de la guía del proyecto dentro de la financiera. Solamente se hizo uso de las buenas prácticas de programación estructurada y de comentar el código fuente. Ver anexo G con definiciones sobre las rutinas ETL.

El plan de implementación Data Warehouse es la parte de la arquitectura de Data Warehouse que identifica los incrementos del Data Warehouse y describe la secuencia de desarrollo de estos incrementos, así se aprovechan los esfuerzos que se realicen para los incrementos posteriores.

La definición de los componentes indica las técnicas a utilizar para soportar, presentar y explotar la funcionalidad del Data Warehouse. Aquí se deben de tomar en cuenta políticas de la empresa y las restricciones propias del proyecto, como ejemplo: el presupuesto asignado, herramientas con que se cuentan, plataforma OLTP, cultura organizacional.

C. Etapa 3: Enriquecimiento de Datos

Enriquecimiento de Datos

Los datos del Data Warehouse son integrados atravesando los márgenes de los sistemas de aplicación, dichos datos son extraídos de las restricciones y límites impuestos por las aplicaciones que los originaron, luego deben ser estandarizados e integrados para proveer una perspectiva de información que traspasa los procesos la cual es esencial para el análisis de utilidades.

Una vez que se tiene en su lugar el modelo de datos del Data Warehouse es necesario poblar su estructura con sus datos respectivos, pero para lograr que los datos del sistema transaccional se acomoden y quepan en el entorno analítico y se conviertan en información, es necesario que dichos datos sufran procesos de transformación que los reconcilien, limpien y enriquezcan ya sea dividiéndolos en algunos casos, fusionándolos en otros, adicionándolos, o derivándolos y que posteriormente estén listos para que cumplan apropiadamente su nuevo rol como base para la toma de decisiones.

Con estos requerimientos plenamente identificados se toma en cuenta además la situación empresarial nicaragüense, cabe destacar que las condiciones más comunes, de los sistemas organizacionales, consiste en implementaciones que abarcan distintos DBMS y/o distintos lenguajes de programación, otro aspecto es que la tecnología de información predominante es moderna, con apenas un poco más de una década de antigüedad o menos en la mayoría de los casos.

La etapa del Enriquecimiento de Datos como se le ha designado en la guía debe proveer mecanismos para realizar la migración de datos desde los sistemas fuente, que pueden residir en distintas plataformas y transformarlos hasta que estén en condiciones de poblar el Data Warehouse, para servir de ayuda en la toma de decisiones. Esta es la etapa más tardada y que más esfuerzos necesita, en lo relativo a la cantidad de trabajo por

realizar, dentro de un proyecto de Data Warehouse y depende directamente de la calidad de los datos de los sistemas transaccionales, de la diferencia tecnológica entre la plataforma fuente versus la Plataforma Destino, fundamentalmente entre otros aspectos.

Resultados del Enriquecimiento

Para que la implementación de esta etapa no salga fuera de control ponga en riesgo el margen de éxito del Data Warehouse, es necesario determinar detalladamente el plan a seguir. Las resultas correspondientes a esta etapa son:

- Documento de Correspondencia Fuente Destino.
- Diseño del Área de Preparación de Datos.
- Diseño de las rutinas de Extracción, Transformación y Carga de Datos.

No importa cual metodología se elija seguir en el desarrollo de un Data Warehouse, siempre se está refiriendo a la construcción de un nuevo entorno, el cual es poblado con los datos, provenientes de los sistemas fuentes, luego de ser extraídos y transformados para enriquecerlos.

Para facilitar el inicio del proceso de enriquecimiento de datos, es muy conveniente hacer una correspondencia entre los datos para identificar el sitio en el cual estos se originan y su campo destino en el Data Warehouse, seleccionando aquellos datos que sean más confiables y que mejor se acomoden al nuevo entorno, un “mapeo fuente destino” como le llama Kimball.

Para el caso de ejemplo en la financiera, la situación es de las más sencillas que se puede encontrar por cuanto, solo existe un sistema fuente que utiliza un DBMS relacional para gestionar la Base de datos, aunque mucha información se encuentra en tablas auxiliares creadas mediante la implementación propia del sistema y que no son del dominio del DBMS sino del lenguaje de programación con el que se desarrolló la aplicación.

En cuanto a la forma en la que se puede formalizar el documento de correspondencia, este puede ser simplemente una tabla en la que se indican los campos fuente (origen) de datos, la narración de las transformaciones que se deben aplicar hasta llegar a los campos destino y el tipo de datos del Data Warehouse, como se observa en la Tabla V-2, también se deben indicar el sistema origen en el caso de ser más de uno y las

fuentes de datos externas que sean necesarias incluir para desarrollar la información pertinente en el Data Warehouse. La tabla de correspondencia completa aparece también en el Anexo G.

Tabla V-2 Correspondencia Fuente Destino

Tabla Destino	Columna Destino	Tipo de Datos	Tabla Fuente	Campo Fuente	Notas transformación de Datos
Abono_Credito_Facts	Fecha_Clave	Integer	Credkar	Dfecsis	reemplazar fecha con la clave en fecha_dim
Abono_Credito_Facts	Fecha_Vence_Clave	Integer	Credkar	Dfecpro	reemplazar fecha con la clave en fecha_vence_dim
Abono_Credito_Facts	Hora_Clave	Integer	Cjgdlog	Dfectrx	Obtener la hora con respecto a la fecha
Abono_Credito_Facts	Analista_Clave	Integer	Credkar	Ccodcta	asociar la cuenta con el analista en cremcre
Abono_Credito_Facts	Cliente_Clave	Integer	Credkar cremcre	ccodcta , ccodcli	asociar la cuenta con el cliente en cremcre
Abono_Credito_Facts	Sucursal_Clave	Integer	Credkar	Ccodcta	left(ccodcta,2) para el código oficina
Abono_Credito_Facts	Convenio_Clave	Integer	Credkar cremcre	ccodcta ccodgru	asociar la cuenta con el convenio en cremcre
Abono_Credito_Facts	Producto_Clave	Integer	Credkar cremcre	ccodcta ccodpro, cdescre csececo	asociar la cuenta con ccodpro, cdescre y csececo en cremcre
Abono_Credito_Facts	Circunstancia_Clave	Integer	Credkar cremcre	Ccodcta	Asociar la cuenta con datos en cremcre
Abono_Credito_Facts	Monto_Total_Abono	Numeric(82)	Credkar	Nmonto	Elegir el cconcep con valor cj(caja)
Abono_Credito_Facts	Principal_Abono	Numeric(82)	Credkar	Nmonto	Elegir el cconcep con valor kp capital
Abono_Credito_Facts	Interes_Abono	Numeric(72)	Credkar	Nmonto	Elegir el cconcep con valor in interes
Abono_Credito_Facts	Servicios_Abono	Numeric(72)	Credkar	Nmonto	Elegir el cconcep con valor '02' servicios
Abono_Credito_Facts	Interes_Mora	Numeric(72)	Credkar	Nmonto	Elegir el cconcep con valor mo mora
Abono_Credito_Facts	Dias_Mora	Integer	Credkar	dfecpro – dfecsis	efectuar calculo de dfecpro dfecsis
Abono_Credito_Facts	No_Recibo_Caja	Varchar(12)	Credkar	Cnrodoc	
Abono_Credito_Facts	No_Credito	Varchar(12)	Credkar	Ccodcta	
Abono_Credito_Facts	No_Cuota	Integer	Credkar	Ccodcta	
Abono_Credito_Facts	Monto_Total_Cuota	Numeric(72)	Credkar	Ccodcta	

En la Tabla V-2 aparece parte de la correspondencia de datos para la tabla Abonos_Credito_Facts que guarda los hechos concernientes a la forma en que los clientes realizan abonos a los créditos que se le otorgaron. Esta tabla se corresponde principalmente con dos tablas observadas en el Ilustración V-1, credkar y cremcre, la información de ambas tablas es necesaria para completar la estructura de la mayor parte de los datos que requiere la tabla Abono_Credito_Facts.

La correspondencia de la tabla de Abono_Credito_Facts, a la cual se le ha dado seguimiento, muestra la aplicación de algunas consideraciones como por ejemplo en el campo de hora, a pesar de que la tabla credkar es el origen de datos de la fecha y de que el campo correspondiente es un datetime, este campo se almacena con la hora vacía, por lo que es mejor fuente de datos lo registrado en la tabla cjdlog para el campo hora pues almacena la hora en la que se realizó el abono, que es el dato que se pretende guardar en la tabla de Abonos_Credito_Facts del Data Warehouse.

Esto también se aplica para la clave de sucursal, que a pesar de contar con un campo ccodofi, que guarda el código de la sucursal, del estudio del modelo de datos se conoce que esta información presenta inconsistencias, lo que motivo se decidiera mejor utilizar como fuente del dato un subconjunto de la llave inteligente del numero de crédito.

Otro caso de ejemplo se puede observar que en la situación de los campos Monto_Total_Abono, Principal_Abono, Interes_Abono, Servicios_Abono, Interes_Mora, todos tienen como origen el mismo campo Nmonto, pero ellos realmente solo toman el valor de Nmonto en función del valor que tome el campo cconcep, conocimiento que queda registrado en la tabla de correspondencia.

El objetivo de tal documento de correspondencia es servir de guía para conocer de forma inequívoca el origen de los datos, lo cual es indispensable para el mantenimiento y durante la siguiente iteración del Data Warehouse, para asegurar que se conoce el significado y origen de los datos y facultar la compatibilidad de ambas iteraciones. La tabla de correspondencia presenta las derivaciones y todos los tipos de transformaciones que debe sufrir un dato hasta llegar a su destino final.

Este documento es fuente de metadatos, habilita a que los demás integrantes del equipo de desarrollo y en algunos casos los usuarios que lo requieran puedan comprender que significa la información almacenada en el Data Warehouse. Otro de los beneficios que se obtiene de elaborar este documento, es poder homogenizar el proceso de extracción, transformación y carga, ya que se pueden detectar fácilmente incongruencias en el procesamiento de algunos datos que deberían ser coincidentes.

Finalmente, es importante tener detallada la correspondencia fuente destino ya que esta documentación sirve de guía para la definición de las rutinas ETL y para la concepción del Área de Preparación de Datos. Una vez determinada la correspondencia y con el conocimiento de las transformaciones necesarias es posible diseñar el espacio intermedio, llámese Área de Preparación de Datos, con el objetivo de minimizar este procesamiento y que las rutinas ETL lleven a cabo su trabajo en el menor tiempo posible, optimizando el desempeño del sistema.

El Diseño del Área de Preparación de Datos es abordado por todos los autores analizados anteriormente no obstante, a pesar de que todos coinciden en su conceptualización como una instancia intermedia que albergará los datos crudos o en procesamiento antes de que éstos sean depositados en la estructura definitiva del Data Warehouse, las características, temporalidad y otros aspectos son disímiles para ellos. Uno la concibe como inmersa en una capa del Data Warehouse, en la capa de datos reconciliados, que es donde se solucionan las diferencias de formato, tipo de dato y organización de los datos provenientes de distintos sistemas fuentes, los cuales pueden ser o no compatibles, para luego convertirse en información. La capa de Datos Derivados representa el modelo de Datos de la organización y en ella es donde se almacenan todos los datos ya integrados para el Data Warehouse. Kimball no detalla mucho en ella, pero la menciona como parte de la arquitectura técnica de la adquisición de datos. Y finalmente Inmon, quien presenta como el área de preparación de Datos al Operational Data Store (ODS), un entorno en el que se integran y reconcilian los Datos de los sistemas transaccionales, pero que es volátil pues sólo presenta el instante actual de los datos de negocio, cambia a la misma velocidad que los sistemas transaccionales y no cuenta con la presencia de Datos derivados, ni el contexto histórico como en un Data Warehouse, pues su objetivo es el monitoreo operacional.

En esta etapa de la guía se estableció la necesidad de diseñar el área de preparación de datos, porque en ella es que se integran y reconcilian los datos fuentes hasta convertirse en datos destino, pero este diseño depende en cierta medida del enfoque se le elija.

En la institución financiera ejemplo, para decidir que enfoque utilizar se consideraron además, los objetivos del proyecto y los costos que conlleva cada propuesta.

Cuando el Área de Preparación de Datos está inmersa en una capa del Data Warehouse, como propone Devlin, el área de almacenamiento permanente requerida puede disminuir, si los datos son desechados luego de la carga. Sin embargo, hay que apreciar que sobre esta capa existe la capa de datos derivados, análoga a los Data Marts dependientes de Inmon, lo que efectivamente aumenta de forma considerable los requerimientos de almacenamiento para albergar simultáneamente ambas capas, se aumenta el tiempo de desarrollo del proyecto, lo que también provoca una elevación en los costos y entra en contradicción con el objetivo de la guía de adaptarse al contexto nicaragüense en el que las empresas medianas y grandes se verían en dificultad para disponer de recursos suficiente para el mantenimiento de ambas capas paralelamente.

El ODS de Inmon requiere un esfuerzo similar al de desarrollar un Data Warehouse para su construcción, pero no es una plataforma para el análisis sino de monitoreo, valga hacer la salvedad que luego que se desarrolla un ODS, la construcción de un Data Warehouse, es mucho más sencilla pues ya se cuenta con una plataforma de datos de calidad. Si se utiliza un razonamiento análogo al aplicado a la propuesta de Devlin, se aprecia que la introducción de una nueva plataforma para el ODS, sin embargo, podría tener un costo restrictivo para el proyecto.

Inmon aparte de la proposición del ODS deja abierto el criterio sobre la forma en la que se deben preparar o reconciliar los datos al igual que Kimball.

En la presente guía de ambas propuestas se consolidó una proposición atractiva en la que se disminuye el costo de desarrollo, la redundancia y la necesidad de almacenamiento, considerando siempre el objetivo de la guía de mantenerse en el marco empresarial nicaragüense. El área de preparación de Datos, será una Base de Datos de transición que albergará los datos que están en proceso de transformación mientras las rutinas de ETL hacen su trabajo, hasta que éstos no sean depositados en su destino final dentro del Data Warehouse, y que una vez completada la carga de datos, es limpiada de contenido, dejando solamente su estructura, a menos que alguna consideración propia de cada proyecto indique que éstos deben permanecer. Siempre se reservará cierto espacio dispuesto para contener los datos en el momento en que se estén llevando a cabo las transformaciones, para posteriormente ser liberado. No se considera una capa más del Data Warehouse sino un espacio pre-formateado que sirve de apoyo para que los datos se

depositen mientras siguen su camino al Data Warehouse, que podría estar conformado de Data Marts y no en dos capas.

En el Anexo F aparece la definición del Área de Preparación de Datos del proyecto que se utiliza de ejemplo.

Pasando ahora a la siguiente resulta de esta etapa, las rutinas de Extracción de Datos, deben ser diseñadas siguiendo los estándares definidos en la arquitectura, con el objetivo de minimizar el esfuerzo necesario para implementar los distintos pasos de la ETL al Data Warehouse y que la ETL sea lo más eficiente posible. Este aspecto es fundamentalmente importante cuando existen varias personas trabajando en el desarrollo de las rutinas de ETL para que los distintos componentes se unan y funcionen de manera apropiada. Aunque ésta no es la parte más difícil si consume mucho tiempo y recursos para definir las rutinas siguiendo los patrones ya establecido y el nivel de detalle suficiente.

D. Etapa 4: Construcción en Incrementos, Data Marts.

Construcción en Incrementos:

Cualquier proyecto de Data Warehouse, en cualquier contexto, tiene mayor probabilidad de éxito si se construye mediante incrementos, que permitan que los usuarios tengan prontamente a su disposición un subconjunto del Data Warehouse completo y se inicie la retroalimentación y ajustes al desarrollo. Por supuesto, que no se debe menospreciar los beneficios que esta modularidad implica para las finanzas de un proyecto.

Las tres metodologías abordadas, presentan aproximaciones distintas de construcción en incrementos.

Inmon, por ejemplo, coloca primero todo el diseño del Data Warehouse en su lugar y luego indica la elección de subconjuntos del Data Warehouse Corporativo para implementar, sin embargo los usuarios solo comienzan a explotar el desarrollo del Data Warehouse, hasta que aparecen los Data Marts dependientes y eso será hasta que se complete la construcción del Data Warehouse corporativo. Devlin presenta una propuesta parecida.

Kimball es radicalmente distinto desde su concepción inicial, el Data Warehouse Corporativo surge luego de los Data Marts, cada Data Mart es un subconjunto del Data Warehouse que puede ser desarrollado de forma independiente y puesto a la disposición de los usuarios antes de iniciar otro Data Mart.

A esta altura del desarrollo de la guía, se debe considerar además las alternativas previamente seleccionadas y puestas en práctica. Al haber escogido en la segunda etapa el modelado dimensional, se eligió también una aproximación a un Data Warehouse mediante la construcción de Data Marts que utilizan dimensiones conformadas para unirse y poder funcionar como un verdadero Data Warehouse.

La implementación en incrementos del Warehouse desarrolla y genera un subconjunto del Data Warehouse total. La implementación incremental es una aproximación pragmática para construir un Warehouse a un nivel empresarial en forma evolutiva. Esta es la etapa en que toda la preparación y diseños anteriores son llevados a la realidad y puestos en funcionamiento.

En esta etapa ya se vincula estrechamente el diseño de las etapas anteriores con la plataforma Hardware y Software que se decidió conformar a tomando en cuenta todo lo reflejado en el estudio de factibilidad.

Resultados de Implementación:

Los resultados más significativos obtenidos de la implementación de un incremento Data Warehouse, incluyen:

- Definiciones de las bases de datos del Data Warehouse.
- Programas y procedimientos para extraer, transformar y cargar datos.
- Población del Data Warehouse con los datos necesarios.
- Población del catálogo de metadatos.
- Selección, instalación y configuración de herramientas de acceso a los datos.
- Técnicas de uso y soporte del Warehouse.

Las definiciones de las bases de datos del Data Warehouse, consisten en el procedimiento automatizado que genere la estructura de la base de datos a como se definió en su diseño en la plataforma elegida para sustentar el Data Warehouse.

El DBMS elegido en la financiera ejemplo es Microsoft SQL Server 2000, por lo que las definiciones de las bases de datos, usan un dialecto denominado Transact SQL, que se base en el estándar SQL92. El script que genera la Base de Datos del Data Warehouse aparece en el Anexo F.

Lo que se enuncia como programas y procedimientos para extraer, transformar y cargar datos, no es más que el código fuente y los ejecutables del software necesario para realizar la ETL. Existen dos tipos de programas y procedimientos los correspondientes a la carga inicial que ocurre cuando se realiza la población del Data Warehouse por primera vez y los procedimientos de la carga periódica que se encargan de mantener actualizada la información del Data Warehouse.

De estos dos tipos de carga, la inicial y la periódica, para el propósito de esta guía se construyó únicamente una parte del conjunto de carga inicial, pues la ETL es una de las partes que más recursos se necesitan y el objetivo de ejemplificación de la guía se cumple con seleccionar ciertas dimensiones y hechos para ser poblados y construir programas y procedimientos necesarios.

Como ya se mencionó, SQL Server 2000 será el gestor de la base de datos que albergará el Data Mart de prueba que se implementa en la financiera, una de las razones por las cuales se haya tomado dicha decisión es por los Servicios de Transformación de Datos que SQL Server 2000 presenta incluidas con la licencia del gestor y que facilita y automatiza grandemente el proceso de ETL.

Los Servicios de Transformación de Datos o DTS (como se les llama por sus siglas en Inglés), se presentan con una interfaz gráfica que permite visualizar las tareas que se deben ejecutar en el proceso de carga y su secuencia de ejecución, cada tarea sin embargo, puede corresponder a la llamada a un programa previamente desarrollado, ejecución de sentencias SQL o ejecución de tareas nativas de carga y transformación de datos que realizan transformaciones de registros definidas en Visual Basic Script o Java Script. Una muestra del paquete que realiza la carga de la tabla de Hechos de los Abonos a Créditos (Abonos_Credito_Facts), se puede observar en la Ilustración V-5

Cada icono representa una conexión con una Base de Datos o una tarea del DTS. Las tareas pueden ser tan sencillas como la ejecución de una sentencia SQL o más complejas como transformaciones de registros con base a criterios especificados con los lenguajes de scripting que soporta.

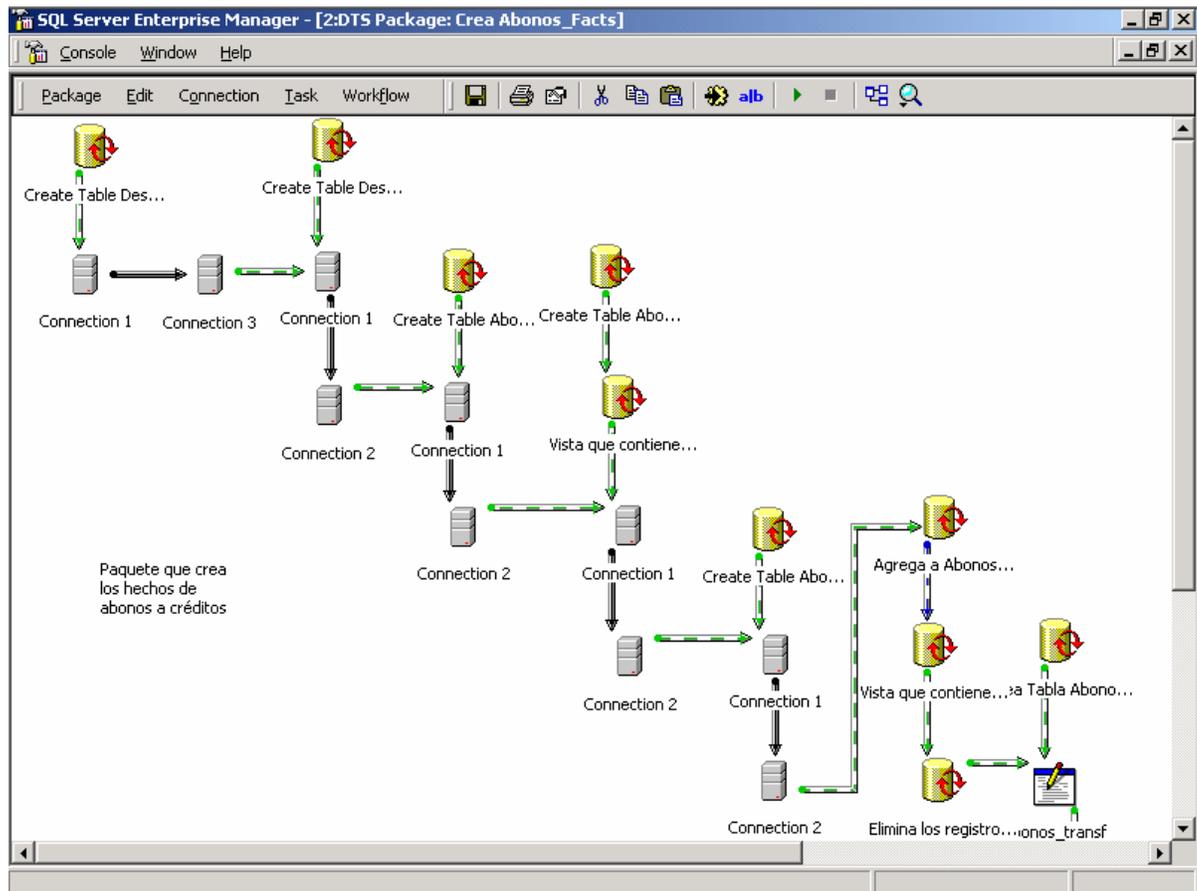


Ilustración V-5 Paquete DTS de carga a una tabla de hechos

Para que los usuarios inicien la explotación del Data Warehouse, éste debe contener los datos de la compañía. A pesar de todo el cuidado y dedicación que se haya puesto en las etapas anteriores siempre habrá que hacer algunos ajustes cuando llegue el momento final en que se realiza la población completa por primera vez y todas las piezas pasen a ocupar su lugar respectivo y a interactuar con las demás. Luego habrá que monitorear persistentemente la carga del Data Warehouse y sus componentes para que esta se esté realizando libre de errores, además de verificar su desempeño con el objetivo de hacer ajustes de mantenimiento y de optimización de procesos y de la base de datos del Data Warehouse.

Esta resulta de la construcción en incrementos, queda pendiente de realizar en la empresa financiera ejemplo, pues aún no se ha terminado la población completa con todas las rutinas necesarias, y mientras no se realice la carga periódica y se puedan identificar partes rutinarias que puedan ser optimizadas o mejoradas si están presentando errores.

Un elemento importante y que no debe ser dejado de un lado pues es indispensable para la adecuada gestión del Data Warehouse es la gestión de los metadatos, ya existen para esto herramientas especializadas que se encargan de ellos y les administran en un catálogo. Para que los metadatos estén a disposición de los usuarios y entre en funcionamiento debe realizarse la población del catálogo de metadatos. Una vez en línea esta herramienta, se logra la automatización de muchas rutinas y un correcto seguimiento e interpretación de los datos contenidos en el Data Warehouse. Aquí también es importante destacar, como claramente expresa Devlin, que los usuarios harán uso del catálogo de metadatos para poder entender e interpretar los datos que son puestos a su alcance.

Por restricciones de la naturaleza del proyecto piloto en la financiera, y algunas consideraciones de costo, el componente del catálogo de metadatos y el repositorio de metadatos, no pudo ser llevado a la práctica, esta limitante sin embargo, sirvió para verificar, que si se descuida la gestión de los metadatos o se ignora la importancia de ellos, el resultado es que el proceso de seguimiento del desarrollo del proyecto puede llegar a ser bastante complejo.

En lo respectivo al mantenimiento del Data Warehouse, la ausencia de metadatos adecuados se manifiesta como la falta de una fuente de datos en formato digital extensiva y formal que pueda ser utilizada como apoyo por otras herramientas para dirigir los distintos procesos que se necesitan en la rutina diaria de un Data Warehouse, como por ejemplo la carga de actualización o la redefinición de algún dato que varíe con el tiempo como la tasas de interés aplicada.

El proceso de selección, instalación y configuración de herramientas de acceso a los datos, determina que el front room del Data Warehouse sea concretado. Este proceso es delicado puesto que se debe tener siempre en cuenta que el front room es la fachada del Data Warehouse ante la mayor parte de los usuarios y por lo tanto mucho del éxito del

proyecto depende de que tan bien y útil sean las herramientas de acceso a datos, para que se obtenga el mayor beneficio posible de la información contenida en el Data Warehouse. Es necesario mencionar que las herramientas a elegir deben ser configurables hasta el punto que se puedan adaptar adecuadamente al ambiente de la organización y tratar de disminuir al mínimo posible la curva de aprendizaje y sin perder de vista un adecuado balance costo - beneficio. En esta parte se incluye la instalación de las herramientas en el ambiente del usuario y su preparación para el uso.

Para poner la carga de prueba en manos de los usuarios y determinar el éxito del proyecto piloto desarrollado en la financiera, se utilizó como herramienta de acceso a datos un servicio propio de SQL Server 2000, cuyo nombre es OLAP Server, el cual brinda la posibilidad de gestionar multidimensionalmente cubos de datos de n-dimensiones a partir de la información obtenida del Data Mart. La definición de los cubos, indica se realicen y almacenen los cálculos de agregación para cada posible combinación de hechos–dimensión que conforman el cubo.

Una vez creados los cubos, estos pueden proporcionar su información para ser analizada a las aplicaciones cliente que así lo requieran, como los cubos fueron construidos con el Servidor OLAP de SQL Server 2000 pueden consultarse utilizando una simple hoja de cálculo de la aplicación Excel que viene incluida en la Suite Office de Microsoft. Excel utiliza para hacer consultas al cubo fuente, una tabla dinámica que tenga como origen de datos el cubo previamente definido y procesado. En la Ilustración V-6 se puede apreciar un reporte elaborado de la forma anteriormente descrita.

INFORME DE GESTION DE CREDITOS

3.- Sucursal y Moneda por Periodo y Estado del Credito

Fecha Semana		NOVIEMBRE					
Estado Credito		Vigente					
Datos		Moneda Credit					
Sucursal	Creditos		Principal_cor		Total Creditos	Total Principal_cor	
	CORDOBAS	DOLARES	CORDOBAS	DOLARES			
Managua	211	31	C\$ 2,227,683.88	C\$ 1,700,882.32	242	C\$ 3,928,566.20	
Leon	144	30	C\$ 2,004,343.33	C\$ 1,233,015.93	174	C\$ 3,237,359.26	
Matagalpa	68	14	C\$ 884,291.32	C\$ 1,949,497.23	82	C\$ 2,833,788.55	
Rio blanco	170	9	C\$ 4,298,271.70	C\$ 801,063.75	179	C\$ 5,099,335.45	
Masaya	51	5	C\$ 807,802.79	C\$ 607,274.17	56	C\$ 1,415,076.96	
Esteli	52	24	C\$ 466,392.20	C\$ 1,459,904.03	76	C\$ 1,926,296.23	
Rivas	85	16	C\$ 1,054,448.47	C\$ 841,639.27	101	C\$ 1,896,087.74	
Sebaco	36	9	C\$ 441,336.17	C\$ 1,509,398.55	45	C\$ 1,950,734.72	
Chinandega	73	35	C\$ 930,142.01	C\$ 1,554,909.59	108	C\$ 2,485,051.60	
Ivan Montenegro	96	19	C\$ 1,435,589.08	C\$ 1,509,487.18	115	C\$ 2,945,076.26	
Total general	986	192	C\$ 14,550,300.95	C\$ 13,167,072.02	1178	C\$ 27,717,372.97	

Ilustración V-6 Informe de Colocaciones

En la ilustración anterior se muestra claramente que es posible obtener de forma sencilla y precisa indicadores de desempeño de una empresa, en este caso de una financiera, en tan solo una hoja, los campos que son cuadros desplegables permiten seleccionar los elementos que se quieren visualizar con solo elegirlos; por ejemplo el cuadro desplegable de estado_credito permite elegir entre créditos vigentes o cancelados en el período especificado que puede llegar hasta el detalle de semana o totalizar para un año.

Las técnicas de uso y soporte del Warehouse comprenden en parte a los metadatos, pues significa plasmar en un documento técnico la forma en que se debe dar mantenimiento al Data Warehouse y la forma en la que los usuarios deben interactuar con él para aprovechar mejor sus bondades.

Al ser el Data Warehouse parte de los sistemas de información de una compañía es importante se realice un documento técnico que indique la forma en la que debe mantener en producción el Data Warehouse y otro en el que se le indique a los usuarios como deberán utilizar las herramientas que se están poniendo a su disposición y las bases para interpretar los resultados obtenidos, para que ellos se familiaricen con la filosofía del entorno analítico y comiencen la verdadera explotación del Data Warehouse recién desarrollado.

Un Data Warehouse es exitoso en la medida en que más usuarios lo utilicen y obtengan beneficio de él, así se mejorará la situación estratégica de la empresa en el competitivo mundo actual de los negocios.

CONCLUSIONES

Con esta guía se demuestran los elementos esenciales del desarrollo de un Data Warehouse con el propósito de que sirva de partida a los noveles en la materia y darle sentido a las numerosas metodologías disponibles.

Existen muchas estrategias de Data Warehouse contradictorias y similares a la vez que han tenido éxitos en diferentes ambientes empresariales que dificultan el aprendizaje y dominio de esta tecnología a los noveles en la materia. Adaptar y conciliar las metodologías de Data Warehousing más difundidas en el mercado tiene un alto sentido, pues permite la creación de un marco de trabajo que engloba elementos cuya aplicación adecuada ha demostrado asegurar la realización y utilidad de un proyecto de Data Warehouse.

El desarrollo de un proyecto de Data Warehouse introduce un nuevo entorno en el cual es posible obtener información vital para la empresa casi instantáneamente lo que habilita la toma de decisiones que mejoren el desempeño de la compañía y los réditos que ésta obtiene. Además, pone al descubierto problemáticas en los sistemas transaccionales que son origen de debilidades y que deben ser solucionados. Un Data Warehouse es parte de los sistemas de información corporativos.

La forma que cada organización y equipo elija para implementar las etapas de la guía y la utilización o no de las herramientas presentadas es decisión de los responsables de cada proyecto de Data Warehouse y estará en dependencia de lo que más le convenga a los intereses de la organización.

Se debe propiciar el desarrollo de proyectos del tipo de Data Warehouse como una herramienta para hacer más competitivas las empresas nicaragüenses en crecimiento, sin embargo la implementación de un entorno analítico debe ser cuidadosamente analizado para no correr el riesgo de realizar una alta inversión y no generar beneficios para la organización. No hay que olvidar que la justificación de una empresa para la inversión en la construcción de un Data Warehouse, es que el Data Warehouse suplirá la necesidad de información estratégica que tiene la empresa y no precisamente del tamaño que ésta presenta.

Si existen porciones de una metodología que son inapropiados para el proyecto de Data Warehouse que se este desarrollando, éstas deben ser excluidas; no se debe sentir la obligación de completar todas las partes de la metodología que se elige implementar. La mejor opción es consolidar un enfoque híbrido que tome en cuenta el contexto del proyecto y que permita ser lo más efectivo en cuanto a la obtención de una plataforma analítica que satisfaga las necesidades de la empresa, para la que se está llevando a cabo el proyecto, como se hizo con los metadatos, que quedaron fuera en el primer incremento por motivos presupuestarios.

RECOMENDACIONES

Las recomendaciones sirven de base para que futuros trabajos que tomen como referencia el presente estudio monográfico, estén en mejores condiciones de poder realizarse con éxito. Esto deriva del hecho de que se cuenta con la ventaja de aprovechar recomendaciones basadas en la experiencia, que sirvan para evitar cometer errores graves como menospreciar tópicos importantes que afecten considerablemente el proyecto, lo cual sin duda es valioso y se encarga de complementar las conclusiones a las que se llegó con el desarrollo.

De la realización del presente estudio monográfico se desprende las siguientes recomendaciones:

1. Se recomienda la utilización de herramientas que permitan el seguimiento de las actividades realizadas en el desarrollo del proyecto de Data Warehouse, para facilitar su posterior mantenimiento y éxito ya que si la definición de los metadatos se deja para el final su definición será muy pobre y esto le resta utilidad a un Data Warehouse.
2. Se debe incluir en todo equipo de desarrollo un integrante con amplio dominio de los sistemas transaccionales y de la forma en que opera la organización para disminuir la duplicación de esfuerzos aplicada a determinar la utilidad de los datos, su origen, forma y significado.
3. Se debe involucrar de la mayor manera posible al personal importante de la organización que hará uso del Data Warehouse, para que la implementación tenga éxito en suplir las necesidades de información de ellos y con esto se asegura en gran parte el uso y éxito del Data Warehouse.
4. No se debe menospreciar la importancia de un buen diseño de la ETL, pues aun con un buen diseño, en la implementación surgen imprevistos y elementos no tomados en cuenta originalmente, lo que puede alargar el proceso más de lo estimado poniendo en peligro la continuidad del proyecto.

5. Se recomienda aplicar la metodología de Ralph Kimball a proyectos de Data Warehouse que tengan un tamaño moderado, por motivos de costo, tiempo y menor riesgo de fracaso, además de permite priorizar las áreas de mayor necesidad de la compañía para ser desarrolladas en el Data Warehouse.

6. Es recomendable utilizar la metodología de Inmon para modelar los datos de una organización, cuando el proyecto desde su primera entrega, implique la inclusión de más de tres o cuatro departamentos o más de un sistema transaccional como fuente de información, porque en estas condiciones aumenta el trabajo necesario para integrar, reconciliar y enriquecer los datos, y la metodología de Inmon ayuda a disminuir el riesgo en este tipo de ambientes delicados de construir incrementos que posteriormente no se puedan integrar.

ANEXO A: INDICES DE ILUSTRACIONES Y TABLAS

INDICE DE ILUSTRACIONES

<i>Ilustración I-1 Diferencias entre Sistemas Operacionales y Data Warehouse</i>	4
<i>Ilustración I-2. Arquitectura de tres capas</i>	8
<i>Ilustración I-3. Estructura BII</i>	12
<i>Ilustración I-4. Estructura BIG</i>	14
<i>Ilustración I-5. Metodología Data Warehouse de Inmon</i>	23
<i>Ilustración I-6 Ciclo de Vida Dimensional para implementar un Data Warehouse</i>	27
<i>Ilustración IV-1 Organigrama</i>	60
<i>Ilustración V-1 Modelo de Datos de una Financiera</i>	76
<i>Ilustración V-2- Dimensión Fecha</i>	80
<i>Ilustración V-3 Diagrama de Entidades</i>	82
<i>Ilustración V-4 Diseño físico de la base de datos del Data Warehouse</i>	83
<i>Ilustración V-5 Paquete DTS de carga a una tabla de hechos</i>	94
<i>Ilustración V-6 Informe de Colocaciones</i>	97

INDICE DE TABLAS

<i>Tabla I-1.- Diferencias entre Data Warehouse y las Bases de Datos Estándar</i>	5
<i>Tabla I-2. Marco de Trabajo de la Arquitectura</i>	33
<i>Tabla III-1 Empresas en Nicaragua</i>	50
<i>Tabla IV-1 Requerimientos de Data Warehouse</i>	54
<i>Tabla IV-2 Alternativas de Adquisición</i>	56
<i>Tabla IV-3.- Opciones recomendadas</i>	57
<i>Tabla IV-4.- Dotación de Personal</i>	59
<i>Tabla IV-5.- Costos de Personal</i>	62
<i>Tabla IV-6-. Cronograma</i>	63
<i>Tabla IV-7.- Costos de Infraestructura</i>	63
<i>Tabla IV-8.- Costos Hardware y Software</i>	64
<i>Tabla IV-9.- Costos Comparativos de opciones de compra</i>	65
<i>Tabla IV-10.- Costos mensuales del proyecto</i>	65
<i>Tabla V-1 Matriz de Data Marts</i>	79
<i>Tabla V-2 Correspondencia Fuente Destino</i>	87

ANEXO B: BIBLIOGRAFÍA

Bibliografía

- [Devlin88] Devlin, Barry. An Architecture for a Business and Information System, IBM Systems Journal, 1988
- [Devlin97] Devlin, Barry. Data Warehouse: from Architecture to Implementation, Addison Wesley Longman, 1997
- [Geiger00] Geiger, Jonathan G. The Data Warehouse Model. EC Media Group, 2000
- [Humphries99] Humphries, Mark; Hawkins, Michael and Dy, Michelle. Data Warehousing, Architecture and Implementation, Prentice Hall, 1999
- [INEI97] INEI, Instituto Nacional de Estadística e Informática de Perú, Manual de Construcción de un Data Warehouse, 1997
- [Inmon92] Inmon, William H. Building the Data Warehouse, Technical Publishing Group, 1992
- [Inmon296] Inmon, William H., Imhoff, C. and Battas, G. Building the Operational Data Store, John Wiley & Sons, 1996
- [Kimball98] Kimball, Ralph, et. al. The Data Warehouse Lifecycle Toolkit, John Wiley & Sons, 1998
- [Mattison96] Mattison, Rob. Data Warehousing, 1996
- [Porter85] Porter, M. E, Competitive Advantage: Creating and Sustaining Superior Performance, Free Press, New York, 1985
- [SAS01] SAS Institute, Inc. SAS Rapid Warehousing Methodology, SAS Institute White Paper, 2001
- [Simon01] Simon, Alan and Shaffer, Steven L. Data warehousing and Business Intelligence for e-commerce. Morgan Kauffmann Publisher. 2001.

ANEXO C: COTIZACIONES DE EQUIPOS Y CRONOGRAMA

**COTIZACION No. 220-056**

FINDESA
ATN: ERIKA SEVILLA

Fecha: 30/08/2002
Ciudad: Managua
País: Nicaragua

NOTA 1: Garantía Estándar de Hardware de: 3 Años

NOTA 2: Esta cotización es válida por 15 días.

NOTA 3: Este producto lo tenemos de entrega inmediata hasta agotar existencias

CODIGO: 354

La garantía solo por desperfectos de fabrica

CODIGO	DESCRIPCION	Cant	Precio Unit	Total
P3549A	HEWLETT PACKARD NETSERVER TC4100 Hewlett Packard Netserver TC4100 Procesador Pentium III DE 1.4 Ghz con capacidad de dos (2) procesadores Memoria Cache L2: 512 kb en cada procesador 256 MB memoria RAM PC-133MHZ ECC SDRAM DIMM EXPANDIBLE A 4 GB	1	\$2,338.28	\$2,338.28
P3559A	HP PROCESADOR 1.4GHZ 133MHZ FRONT-SIDE BUS*	1	\$1,350.00	\$1,350.00
D8266A	HP MEMORIA 256 MB PC-133 MHZ ECC SDRAM DIMM*	3	\$180.00	\$540.00
P4620A	Discos Wide Ultra-3 SCSI de 36 GB HOT-SWAP EXPANDIBLE A 584 GB Capacidad de 8 discos instalados De 10,000 rpm MARCA: HP, 10k Ultra 3 Wide SCSI-3 hot Swap HDD Unidad de CD-Rom, 48X Tolerancia a Fallas, RAID (1,3 y 5)	3	\$570.00	\$1,710.00
P3410A	HP NetRAID 1M Controller Tarjeta Controladora de Red Inteligente WAKE ON LAN, FAST Ethernet de 10/100 Mbps Tarjeta de Red Incluida Interfases: 2 serial, 1 paralelo, 1 Grafico, 1 puerto de Mouse PS/2 3.5" 1.44MB Unidad de Disco Flexible 6 Slots de Expansión PCI (4- 32 Bits, 2- 64 Bits) HP NetServer Navigator, Symantec PC ANYWHERE Conforme con SNMP, DMI e IPMI, incluye Top Tool de HP administración HP Mini DIN Mouse HP NetServer Keyboard Fuentes de Poder de 350Watts	1	\$736.17	\$736.17
P2535A	HP fuente de Poder redundante de 350 Watts for TC4100	1	\$619.00	\$619.00
D8899A	Monitor HP Color de 15" HP Top Tools para Servidores HP Openview Managex Event Manager Integrated Remote Assistant Sistemas soportados Microsoft Windows NT4, Windows 2000, Novell NetWare 4.2,5.0, Red Hat Linux, SCO Open Server and UNIXWARE 7.x	1	\$180.00	\$180.00
C1556D	HP SureStore DAT24E Externo de 24GB	1	\$1,272.00	\$1,272.00
C7474A	HP SureStore Ultra3 Wide PCI HBA Kit	1	\$335.00	\$335.00
C7434A	68 Pin HD to 68 Pin VHD SCSI Adapter	1	\$64.00	\$64.00
INST	Instalación y Configuración de Software	1	\$100.00	
			SUB TOTAL	\$9,144.45
			IGV	\$1,371.67
			TOTAL	\$10,516.12

Plaza de compras, colonia C.A, P.O. BOX #2820, Managua, 3505-026, Nic.

Tel.(505) 278-8000 Ext: 150 Fax: (505) 278-0041

e-mail:eddy.perez@corporación.teran.com.ni



Net Solutions

Vicky 1-1/2 C al Sur # 251
Altamira, Managua
Telefax : 277 3185

COTIZACION No 1321

CLIENTE: Ing. Ericka Sevilla
EMPRESA: _____
DIRECCION: _____
TELEFONO: _____ **No FAX:** _____
E-MAIL: _____
FECHA: 16 de Septiembre de 2002 _____
EJECUTIVO DE VENTAS: Lic. Denis Ríos

EQUIPOS	CANT.	PRECIO US \$	TOTAL US \$
Equipos de Cómputo Server Compaq PROLIANT ML350 TWR G2 Procesador Pentium III de 1266MHz Memoria RAM de 1024MB expandible a 4GB CD-ROM de 48x Disk Floppy de 1.44MB	1	3,609.00	3,609.00
Disco duro LA 36.4GB U160 SCSI HD	3	745.00	2,235.00
Procesador P3/1266 512K PROC OPT Pentium III de 1266MHz	1	1,053.00	1,053.00
Para más especificaciones ver archivo PDF adjunto o visite: http://www.compaq.com/products/servers/proliantml350/index.html			
Condiciones : Forma de Pago : 100% al aprobar cotización Tiempo de Entrega : 10 días laborales Oferta válida por 15 días.		Sub Total US \$ IGV 15% Total	6,897.00 1,034.55 7,931.55



ANEXO D: ENTREVISTAS

Diseño de las Entrevistas

Gerencia General:

1. ¿Cuáles son los objetivos de su organización? ¿Cuáles son sus metas de negocio de mayor prioridad? ¿Qué está tratando de alcanzar?
2. ¿Cuáles son sus métricas de éxito (indicadores)? ¿Qué tan frecuente mide sus factores clave de éxito? ¿Qué tipo de decisiones son las que usted toma frecuentemente? ¿Cómo sabe que está haciendo lo correcto?
3. ¿Qué funciones y departamentos dentro de la organización son los cruciales para asegurar que estos factores clave de éxito sean alcanzados? ¿Qué rol juegan ellos? ¿Cómo trabajan en conjunto para asegurar el éxito?
4. ¿Cuáles son las cuestiones (elementos) claves del negocio que usted debe enfrentar en este momento? ¿Qué impide que usted alcance dichos objetivos? ¿Cuál es el impacto en la organización?
5. ¿Cómo identifica problemas/excepciones o reconoce que va en dirección hacia los problemas?
6. ¿Qué visualiza como oportunidades adicionales de utilidades que no son aprovechadas en este momento?
7. ¿Dónde se sitúan en el uso de tecnología de información en comparación con la competencia? ¿Se encuentra en la capacidad de responder rápidamente a las condiciones del mercado y asegurar la productividad de su staff?
8. ¿Qué papel juegan los análisis de datos en decisiones que deben hacer los gerentes para dirigir el negocio?
9. ¿Qué información necesita responder para mantener su puesto? ¿Cuál es la 1er, 2da, 3ra... pregunta relativa a créditos/ganancias que usted debe responder en su área de trabajo?
10. ¿Qué información clave se requiere para hacer o soportar las decisiones que usted toma en el proceso de alcanzar las metas y sobreponerse a los obstáculos? ¿Cómo obtiene esta información en este momento?
11. ¿Cuáles son los reportes que más le ayudan a hacer su trabajo? ¿Qué datos en el reporte son importantes? ¿Cómo usa esta información? ¿Si el reporte fuera dinámico, que sería distinto en él?

12. ¿Qué preguntas le han quedado sin responder simplemente porque usted sabe que no hay respuesta disponible? ¿Existe alguna otra información faltante que crea puede tener un impacto significativo en ayudarlo a alcanzar sus metas?
13. ¿Qué tan difícil es en este momento conseguir esa información? ¿Cómo debería presentarse, de que forma? ¿Qué tan frecuente la necesita: Diario, Semanal, Mensual, Anual?
14. ¿Qué capacidades analíticas le gustaría tener?
15. ¿Qué oportunidades existen para mejorar dramáticamente el negocio, basado en un mejor acceso a la información? ¿Cuál es el impacto financiero?
16. ¿Qué debe alcanzar este proyecto para ser considerado exitoso? El criterio debe ser medible.

Gerencias de Crédito, Finanzas y Mercadeo:

1. ¿Qué tipo de decisiones son las que usted toma frecuentemente?
2. ¿Qué información necesita responder para mantener su puesto?
3. ¿Cuáles son los indicadores que debe monitorear constantemente?
4. ¿Si tuviera abundancia de información sobre su área, que haría con ella?
5. ¿Qué tan difícil es en este momento conseguir esa información?
6. ¿Qué preguntas le han quedado sin responder simplemente porque usted sabe que no hay respuesta disponible?
7. ¿Qué tan frecuente necesita esta información: Diario, Semanal, Mensual, Anual? ¿Cómo debería presentarse esta información, de que forma?
8. ¿Cuál es la 1er, 2da, 3ra... pregunta relativa a créditos/ganancias que usted debe responder en su área de trabajo?
9. ¿Cuáles son los reportes que más le ayudan a hacer su trabajo?
10. ¿Con quienes discute y analiza la información de su área?
11. ¿Cómo se organiza y funciona su departamento? ¿Cómo fluye la información en su depto?
12. ¿Cómo le ayuda el sistema actual a desempeñar su trabajo y a la toma de decisiones?
13. ¿Cuáles son las metas de su área? ¿Cuáles son las prioridades de su área?
14. ¿A que le llamaría factor crítico de éxito en su departamento?
15. ¿Cómo ayuda la información de su área al desempeño de la empresa?

Gerencia de Sistemas:

1. ¿Cuáles son las características de los sistemas actuales?
2. Haga una breve Descripción de la plataforma tecnológica con la que cuentan
3. ¿Cuánto tiempo de información se tiene en línea?
4. ¿Qué tan estable es el modelo de datos de la organización?
5. ¿Quiénes son los responsables de los módulos de los sistemas transaccionales?
6. ¿Quiénes están a cargo de los cambios en el sistema transacciones?
7. ¿Existe algún tipo de documentación sobre los sistemas que utiliza la organización? Diccionario de Datos, Manuales Técnico y de Usuario, por ejemplo
8. ¿Cuáles son los sistemas heredados o antiguos con los que se cuenta?
9. ¿Existen datos históricos almacenados fuera de línea? ¿Cuánto tiempo de datos?
10. ¿Dónde guardan el histórico?
11. ¿Quién conoce la estructura de los datos del sistema transaccional? ¿Están limpios?
12. ¿Cuáles herramientas de usuario final para reportes y análisis tiene la empresa?

Resultados de las Entrevistas

Gerencia General	
Objetivos	<p>Proveer servicios financieros innovativos, ágiles y ajustados a las necesidades del mercado</p> <p>Participar del veinticinco por ciento del mercado financiero en los próximos cinco años</p> <p>Contar con una cartera de estipulada a mediano plazo.</p> <p>Capitalizar cierto monto sobre las acciones.</p>
Prioridades	<p>Correcta selección de Clientes</p> <p>Disminuir riesgos en la operación financiera</p> <p>Disponer de Recursos Humanos adecuados</p> <p>Identificación de riesgos y dificultades</p> <p>Análisis de problemas de la institución</p> <p>Coordinación de las áreas y sucursales de la empresa</p>
Factor crítico de éxito	<p>Rendimiento de Capital</p> <p>Rendimiento de Equidad</p>
Decisiones más frecuentes	<p>Determinar soluciones posibles a los problemas de la empresa</p> <p>Aprobar estrategias y solicitudes presentadas por los directivos de departamentos y sucursales</p>
Información Clave	<p>Análisis del rendimiento sobre desembolsos</p> <p>Análisis de disponibilidad: Capacidad de pago de cuota, cobertura promedio de las garantías</p> <p>Líquidez</p> <p>Rendimiento financiero</p> <p>Costos financieros</p> <p>Gastos Administrativos</p> <p>Calidad de los activos y Comprobación de activos</p> <p>Productividad por analista</p> <p>Recursos líquidos</p> <p>Relación de cantidad total por pasivo total</p> <p>Estructura de la cartera por moneda contra el pasivo por moneda</p> <p>Calidad de Cartera</p>
Indicadores	<p>Mora</p> <p>Calidad de Cartera</p> <p>Gastos Administrativos</p> <p>Rentabilidad sobre capital</p> <p>Productividad de Clientes por analistas</p>
Disponibilidad de información	<p>La información no esta accesible</p> <p>Existen datos básicos pero deben realizar cálculos sobre ellos si se quiere disponer de agregados.</p> <p>Los datos tienen poco significado pues no están relacionados para la toma de decisiones.</p>
Reportes de uso más frecuente	<p>Informes de Mora</p> <p>Para cobranza</p> <p>Resumen de contaminada</p> <p>Informes de Colocaciones:</p>

	<ul style="list-style-type: none"> ▪ Por fecha ▪ Por actividad económica ▪ Por analista <p>Informes de Recuperaciones</p> <ul style="list-style-type: none"> ▪ Para cobranza ▪ Por analista ▪ Resumen de recuperaciones ▪ Pagos recibidos <p>Informe de Saldos de Cartera Calidad de Cartera Informe de Liquidez Contabilidad</p>
Información faltante	<p>Clasificación de mora por:</p> <ul style="list-style-type: none"> ▪ Edad ▪ Sexo ▪ Área geográfica ▪ Garantías ▪ Actividad Económica ▪ Inventario ▪ Volumen de Venta ▪ Sucursal -Analista <p>Clasificación de la Cartera Activa por:</p> <ul style="list-style-type: none"> ▪ Área geográfica ▪ Producto ▪ Monto ▪ Calificación del Sistema ▪ Sector Económico ▪ Sexo ▪ Sucursal – Analista <p>Cruce entre las Clasificaciones anteriores</p>
Frecuencia de uso de información	<p>Semanal Mensual</p>
Forma de presentación preferida	<p>Reportes Gráficos</p>
Interacción con demás dependencias	<p>Gerencia Financiera Gerencia de Crédito Gerencia de Mercadeo Gerencia de Sistemas Recursos Humanos</p>
Desempeño del sistema actual en la toma de decisiones	<p>Los reportes brindan datos crudos, que luego deben ser procesados para poder realizar los informes gerenciales y otra documentación para monitorear el desempeño y tomar decisiones.</p>

	Gerencia Financiera
Objetivos	<p>Obtener fondos necesarios para incrementar la Cartera y mantener liquidez.</p> <p>Liquidez mayor o igual al un porcentaje establecido en relación con la cartera bruta.</p> <p>Registro contable de las operaciones.</p> <p>Control Interno enfocado al control del efectivo</p>
Prioridades	<p>Mantener el flujo de recursos necesarios para las operaciones de la institución.</p> <p>Garantizar procedimientos que regulen y controlen las operaciones de caja, tesorería y contabilidad.</p> <p>Cumplimiento de metas establecidas en presupuestos</p>
Factor crítico de éxito	<p>Gestión de Recursos</p> <p>Liquidez</p> <p>Rentabilidad de la institución</p>
Decisiones más frecuentes	<p>Autorización de traslados de fondos a sucursales</p> <p>Gestión de desembolsos a entidades fuentes de financiamiento.</p> <p>Venta o compra de divisas extranjeras</p>
Información Clave	<p>Estado de Liquidez</p> <p>Estados Financieros</p> <p>Estado de la Gestión de Recursos</p> <p>Rentabilidad de Colocación</p> <p>Costo de Pasivos</p> <p>Calce de plazos y calce de monedas</p> <p>Gastos Administrativos</p>
Indicadores	<p>Saldos de Caja por cada Oficina</p> <p>Nivel de Recuperación</p> <p>Nivel de Colocación</p> <p>Pagos de obligaciones</p> <p>Gastos Administrativos reales</p> <p>Comparación ejecución real contra el presupuesto</p>
Disponibilidad de información	<p>Es obtenida de las sucursales existentes.</p> <p>También por gestión de la Gerencia de Sistemas en donde se encuentra la Base de Datos Central.</p> <p>Datos de algunos créditos que se perdieron o corrompieron en la migración al sistema actual.</p>
Reportes de uso más frecuente	<p>Estados Financieros</p> <p>Plan de Negocios:</p> <p>Dirección del volumen de colocación / Recuperación</p> <p>Presupuesto de colocación / Recuperación</p> <p>Gastos administrativos</p> <p>Estados de Resultados Anual</p> <p>Disponibilidades</p> <p>Obligaciones</p>
Información faltante	<p>Reportes de cartera por fuentes de fondos :</p> <p>Colocaciones</p> <p>Recuperaciones</p> <p>Mora</p>

	<p>Indices de captación de otros ingresos a través del sistema</p> <p>No se puede registrar los pagos de clientes por honorarios, avalúos, inscripciones, tarjetas inteligentes, pagarés, fondos en custodia.</p> <p>No hay distinción entre el Estado de Resultado de los ingresos financieros que sólo están devengados (registrados como ingresos pero no pagados) de los que están devengados y efectivamente pagados.</p> <p>Reporte de Estado de cumplimiento de proyección de colocación por montos</p>
Frecuencia de uso de información	Semanal
Forma de presentación preferida	Reportes
Interacción con demás dependencias	<p>Gerencia General</p> <p>Gerencia de Crédito</p> <p>Mercadeo</p> <p>Gerencias de Sucursales</p> <p>Contabilidad General</p> <p>Programación Financiera</p>
Desempeño del sistema actual en la toma de decisiones	Ayuda de alguna forma, pero es muy pobre
Información del departamento que ayuda al mejor desempeño de la empresa	<p>Índice de Ganancia y Pérdida</p> <p>Índice de Rentabilidad</p> <p>Necesidad de modificación al plan de negocio</p> <p>Apertura de nuevos mercados</p> <p>Tasa</p> <p>Comisión</p> <p>Costos</p>

Gerencia de Crédito:	
Objetivos	Asegurar una cartera de Calidad Garantizar nivel de Calidad de los Clientes Desarrollar productos atractivos al mercado Desarrollar una metodología crediticia eficiente y consistente Cumplir con las metas de la empresa
Prioridades	Mantener niveles saludables de Saldos de Cartera Asegurar una Cartera Sana
Factor crítico de éxito	Mantener una Cartera Sana
Decisiones más frecuentes	Aprobación de Crédito: revisión de expedientes Medidas de contención para clientes Morosos Gestión del Manual de Créditos Gestión de Procedimientos de Crédito Contratación / Despidos
Información Clave	Nivel de Colocación Nivel de Recuperación Calidad de Cartera Concentración de mora Productividad de los analistas Productividad de las Sucursales Tiempo de Procesamiento de Solicitud
Indicadores	Mora Colocación Recuperación Productividad Margen de Ganancia por Productividad
Disponibilidad de información	Se dispone de información elemental. Otra información se puede conseguir a través de la Gerencia de Sistemas.
Reportes de uso más frecuente	Informe de Mora Informe de Colocaciones Informe de Recuperaciones Informe de Saldos de Cartera
Información faltante	Reporte de Mora por Calificación del Sistema Clasificación de Clientes por tamaño de su negocio Nivel de Ingreso de los Clientes Distribución Geográfica de los Créditos Consolidados Anuales Clasificación de la Cartera Activa por: <ul style="list-style-type: none"> ▪ Área geográfica ▪ Producto - Monto ▪ Calificación del Sistema ▪ Sector Económico ▪ Sucursal – Analista ▪ Sexo Cruce de Clasificaciones

Frecuencia de uso de información	Diario Mensual(consolidado)
Forma de presentación preferida	Reportes
Interacción con demás dependencias	Gerencia General Gerencia Financiera Gerencia de Mercadeo Gerencias de Sucursales
Desempeño del sistema actual en la toma de decisiones	Ayuda a resolver necesidades de información en un 70% Informe de consolidado se debe hacer manual
Información del departamento que ayuda al mejor desempeño de la empresa	Índice de Colocación Índice de Recuperación Saldos de Cartera

	Gerencia de Mercadeo
Objetivos	Gestionar el crecimiento de los Clientes Nuevos Asegurar el mantenimiento de la Cartera Vigente Identificar oportunidades de Mercado Determinar campañas de inversión publicitaria
Prioridades	Satisfacer las necesidades de los Clientes
Factor crítico de éxito	Ayudar a conseguir metas de colocación Mejorar el nivel de satisfacción de nuestros clientes
Decisiones más frecuentes	Gestión de Publicidad Gestión de Atención al Cliente Visitas a Clientes Potenciales Investigar Quejas
Información Clave	Tiempo de gestión de solicitud de Crédito Cantidad de dinero reembolsado Motivos de no renovación de créditos Clientes que cancelan en el período Revolvencias de los créditos Rentabilidad de la Publicidad
Indicadores	Crecimiento de nuevos créditos Crecimiento de clientes nuevos Bajas de clientes Nivel de satisfacción de los clientes actuales Mantenimiento de la Cartera Vigente
Disponibilidad de información	Información sobre algunos tópicos importantes existe pero se debe obtener manualmente a partir de varios reportes y sobre otros no existe ni un dato
Reportes de uso más frecuente	Clientes actuales Clientes inactivos Créditos Otorgados por actividad económica y por analistas, Créditos repetitivos y nuevos Proyección de Reembolso SalDOS de Cartera
Información faltante	Tiempo de gestión de solicitud de crédito de inicio a fin Control de quejas Índice de renovación de Clientes Factor de retorno a la inversión publicitaria Fichas de perfil personal de los clientes Clasificación de la Cartera Activa por: Sucursal – Analista Zona Producto – Monto Fechas de solicitud, evaluación, desembolso Cruce de Clasificaciones Revolvencias de Líneas de crédito: Cantidad de Revolvencias por cliente Resumen de Revolvencias

Frecuencia de uso de información	Semanal Mensual (consolidado)
Forma de presentación preferida	Reportes
Interacción con demás dependencias	Gerencia General Gerencia de Crédito Gerencia de Sistemas
Desempeño del sistema actual en la toma de decisiones	Ayuda a resolver necesidades de información en un 60%. Informe de renovación de clientes en un período se elabora manual. Contiene a los clientes que no han renovado en un período determinado con sus correspondientes niveles de mora de cada crédito. Actualmente existe un reporte de renovación pero es incorrecto, ya que para algunos clientes asume la mora promedio de todos los créditos Datos sobre publicidad no se reflejan en ningún informe
Información del departamento que ayuda al mejor desempeño de la empresa	Manifestación de factores y puntos críticos en la gestión de la empresa hacia los clientes Pautas de inversión en Publicidad Establecimiento de oportunidades de ampliación de mercado

Gerencia de Sistemas:	
Sistema Actual	Sistema de Gestión de Crédito (LFS) consta de los módulos: Crédito, Clientes, Reportes, Caja, Contabilidad Sistema de Planilla Bases de Datos Distribuida entre las sucursales y fuera de línea. Existe una Base de Datos central que se actualiza en modo BATCH
Cantidad de Información en línea	LFS: 2 años
Estado de la información fuera de línea	Informix: 3 años Almacenado en forma de respaldos en CD y Disco Duro. Los datos se deben montar en la base de Datos para accederlos Los datos presentan cierta inconsistencia. La calidad de los datos no es óptima
Evolución del modelo de Datos	El Modelo de Datos (MD) de la organización es dinámico, cambia adaptándolo a nuevos requerimientos de la empresa. El MD aún no se ha consolidado de forma estable.
Módulos del sistema transaccional	Clientes Reportes Caja Crédito Ahorro Contabilidad Presupuesto
Documentación	LFS: Manual Técnico (80%): ERD, Diccionario de Datos del módulo clientes Manual de usuario: No existe Informix: Manual Técnico Manual de Usuario
Recursos Disponibles	Sistemas Operativos: Windows 98, Windows 2000 Server, Windows XP, Linux Suse 7.0 DBMS: SQL Server 7.0 Programación: VFP 7.0, Visual Studio 6.0, Clipper(planilla) Ofimática: MS-Oficce 97, 2000, XP, Crystal Reports

ANEXO E: REQUERIMIENTOS DE HERRAMIENTAS SOFTWARE

SOFTWARE UTILIZADO

Categoría	Software	Funcionalidades Requeridas	Licencia
Automatizador de ETL	Data Transformation Services de Microsoft SQL Server 2000	Capacidad de: Extracción de datos almacenados en diversas plataformas. Deduplicación de datos redundantes. Duplicación de Datos. Estandarización. Transformaciones por tabla, columnas, registros, campos. Programación de la ejecución de tareas automatizadas. Registro de operaciones efectuadas (logs).	Adquirida de agente autorizado
CASE	Platinum Erwin/ERX 3.5	Realización de Ingeniería en Reversa a partir de bases de datos almacenadas en SQL Server 2000. Presentación de modelos relaciones o multidimensionales. Soporte de distintas metodologías de diagramación. Capacidad de exportar a otros repositorios.	Licencia de periodo de evaluación
DBMS	Microsoft SQL Server 2000 Service Pack 2	Tecnología Relacional. Capacidad de interactuar con Base de Datos multidimensionales. Lenguaje de Consulta Compatible con estándar SQL 92. Administración de integridad referencial . Registro de operaciones efectuadas (logs). Capacidad de realizar respaldos de la información. Juego de Instrucciones para Data Warehouse. Soporte para particiones que contengan datos fuertemente relacionados. Soporte de esquema estrella . Soporte de roles de usuarios. Uso de metadatos y repositorio de metadatos.	Adquirida de agente autorizado
Herramienta de Acceso a Datos	Microsoft Analysis Services 2000	Plataforma OLAP con soporte para HOLAP MOLAP, ROLAP Gestión de cubos de datos multidimensionales Servir de fuente de datos para consulta de otras aplicaciones. Capacidad de utilización de metadatos. Realización y almacenamiento de agregados para mejorar consultas. Capacidad de procesamiento en lotes Soporte para Data Mining y para Executive Information Systems. Soporte para productos con habilidades Web. Acceso a Base de Datos SQL Server 2000.	Adquirida de agente autorizado

Categoría	Software	Funcionalidades Requeridas	Licencia
Modelador de Datos	Oracle Designer / 2000 2.0	Entorno integrado para modelar Bases de Datos. Capacidad de modelado Dimensional o Relacional Posibilidad de generar diseño físico a partir de Diagrama Entidad Relación de alto nivel Permite trabajo concurrente de diseñadores de datos Generación de Scripts de Definición de Bases de Datos, para la implementación del diseño Presentación de Diagramas del modelo desarrollado a distinto nivel de detalle, desde ERD hasta el diagrama de especificación física. Posibilidad de exportar a distintos repositorios Independencia de la plataforma del sistema operativo	Oracle Technology Network, licencia de período de evaluación
Ofimática	Microsoft	Hojas de Cálculo: Capacidad de conexión a servidor OLAP, Analysis Services Flexibilidad para manipular grandes cantidad de datos Capacidad de realizar análisis estadísticos como desviación estándar, 80-20, etc Consulta dinámica de datos. Drill down y roll up	Adquirida de agente autorizado
Sistema Operativo	Microsoft Windows 2000 Server.	Procesamiento en lotes de programas de aplicación Multiprocesador Multiusuario Multitareas Soporte para SQL Server 2000, Analysis Services y Oracle Designer	Adquirida de agente autorizado

ANEXO F: MODELOS DE DATOS Y SCRIPTS

Tabla de Contenido

DIMENSIONES DE LOS DATA MART

DETALLE DE LAS DIMENSIONES

- DIMENSIÓN FECHA
- DIMENSIÓN CLIENTE
- DIMENSIÓN PRODUCTO
- DIMENSIÓN TASAS
- DIMENSIÓN SUCURSAL
- DIMENSIÓN ANALISTA
- DIMENSIÓN MODO DE PAGO

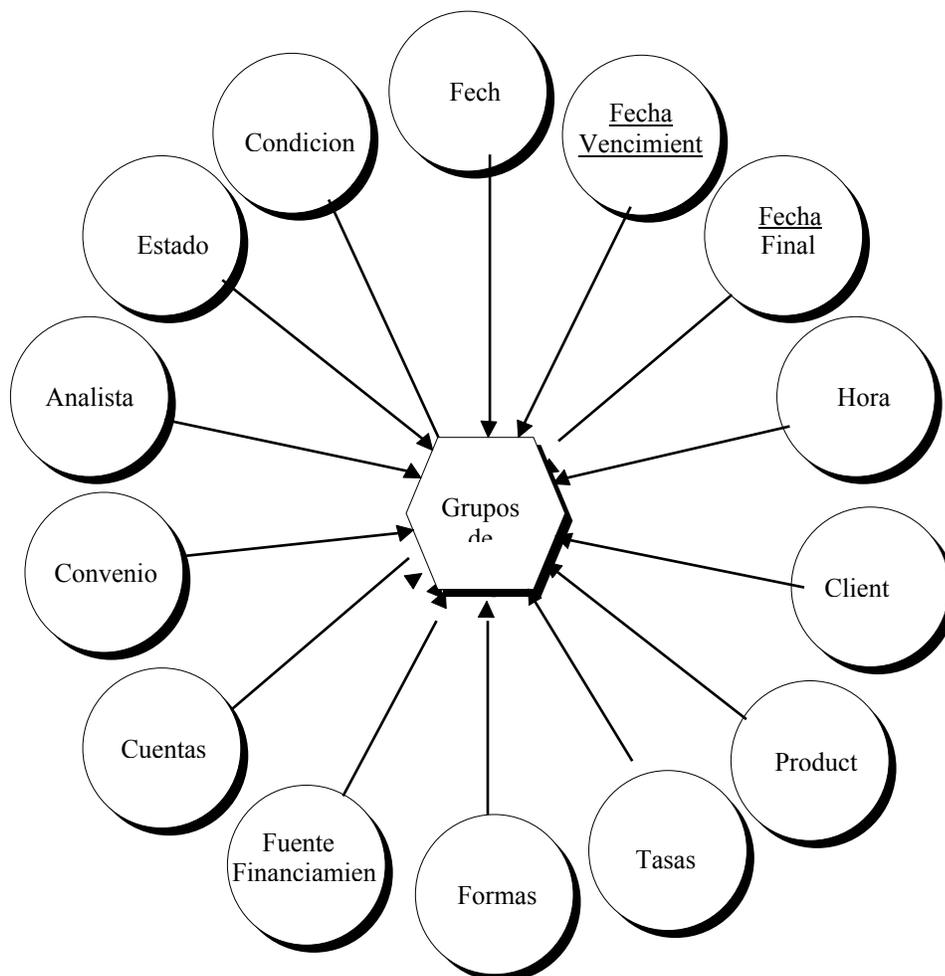
DETALLE DE LOS HECHOS

- CRÉDITOS _GESTIONADOS
- CRÉDITOS _ABONADOS HECHOS
- PROYECCIÓN _RECUPERACIÓN HECHOS

NIVEL 2: GRANO SEMANAL

- PRESUPUESTO _RECUPERACIÓN HECHOS
- CREDITOS _ABONADOS _SEMANALES HECHOS

Dimensiones de los Data Mart



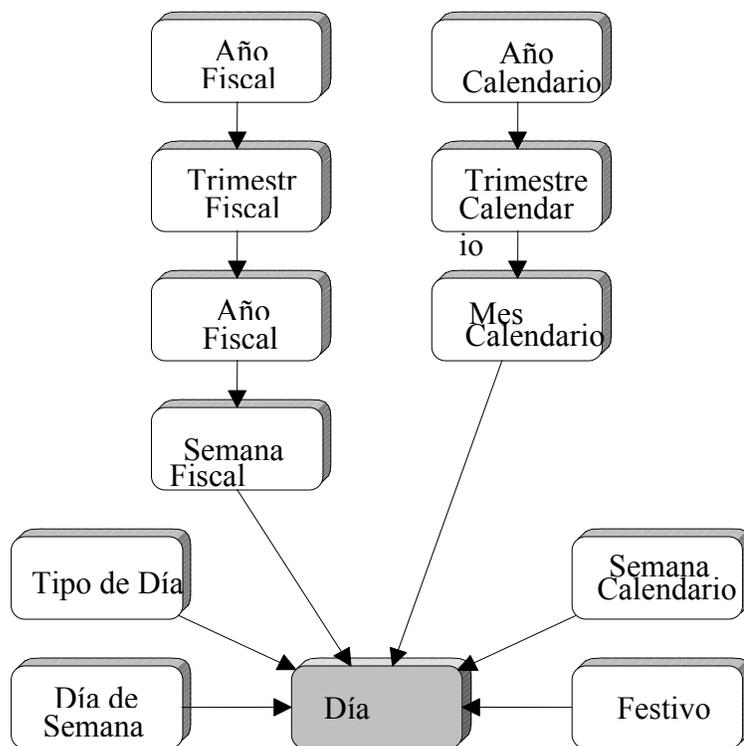
Descripciones de las Dimensiones

Nombre Dimensión	Descripción de la Dimensión
Fecha	Contiene todos los atributos asociados con la fecha en que la actividad ocurrió.
Fecha Vencimiento	Contiene todos los atributos asociados con la fecha en que vence el plazo de pago del monto otorgado o de la cuota actual
Fecha Final	Contiene todos los atributos asociados con un final o fecha de resolución (fin de contrato de crédito, fecha cancelación de deuda)
Hora	Contiene todos los atributos acerca de la hora en que una actividad ocurrió.
Cliente	Representa a ambas personas naturales o jurídicas y todos que son clientes y los atributos asociados a ellos.
Producto	Contiene información acerca de productos y servicios que se ofrecen.
Tasas	Describe las tasas de interés usado para los créditos de los clientes.
Modalidad de Pago	Describe las diferentes modalidades de frecuencia y moneda entre los que un

Nombre Dimensión	Descripción de la Dimensión
	cliente puede elegir para pagar
Fuentes Financiamiento	Características de compañías de las que se obtienen recursos liquidos para la institución
Cuentas LM	Numeros de cuentas del libro mayor utilizadas para seguimientos financieros y reportes.
Convenio	El Grupo de instituciones que han firmado convenios de pago con la empresa.
Analista	Describe a los analistas individuales de la compañía.
Condicion	Describe la condicion en la que se encuentra un credito según los pagos realizados.
Estado	Describe el estado del credito que gestiona un cliente.

Detalle de las Dimensiones

Dimensión Fecha

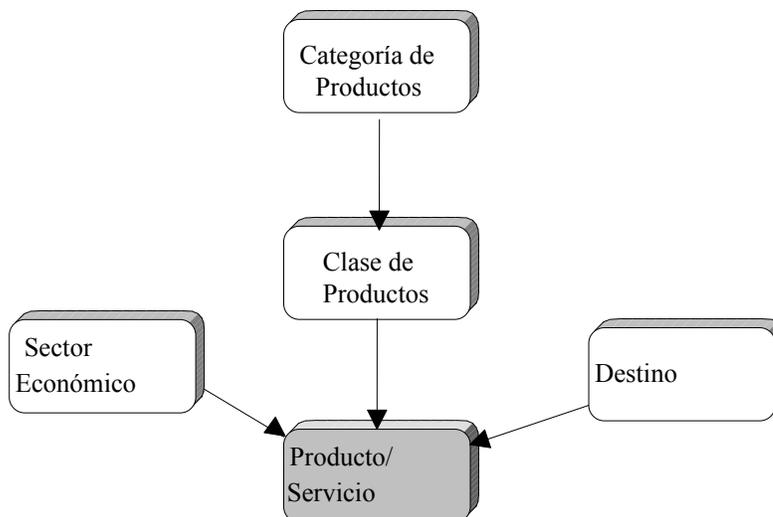


Descripciones de Atributos de la Dimensión Fecha

Nombre de Atributos	Descripción de Atributos	Valores de Muestra
Día	El Día específico que una actividad tuvo lugar.	06/04/2001;6/05/2001
Día de Semana	El nombre específico del Día.	Lunes; Martes
Festivo	Identifica que este día es Festivo.	Día Madre; Navidad
Tipo de Día	Indica si este Día es o no un Día Semana	Fin de Sem; Día de S
Semana Calendario	El día que finaliza la Semana, siempre es un Sabado. FS significa Fin de Semana.	FS 06/06/2001; FS 06/13/2001
Mes Calendario	El mes Calendario.	Enero,2001;
Trimestre Calendario	El Trimestre Calendario.	2001Q1; 2001Q4
Año Calendario	El Año Calendario.	2001
Semana Fiscal	La Semana que representa el Calendario de la empresa. Notar que la F en El valor del dato indica que este es un periodo de tiempo fiscal.	F Semana 1 2001; F Semana 46 2001
Mes Fiscal	El periodo fiscal comprendido de 4 o 5 Semanas. Notar que la F en El valor del dato indica que este es un periodo de tiempo fiscal.	F Enero, 2001; F Febrero, 2001
Trimestre Fiscal	El grupo de 3 meses fiscales.	F 2001Q1; F2001Q2
Año Fiscal	El grupo de 52 Semanas fiscales / 12 meses fiscales que comprende el año financiero.	F 2001; F 1999

Dimensión Cliente Descripciones de Atributos

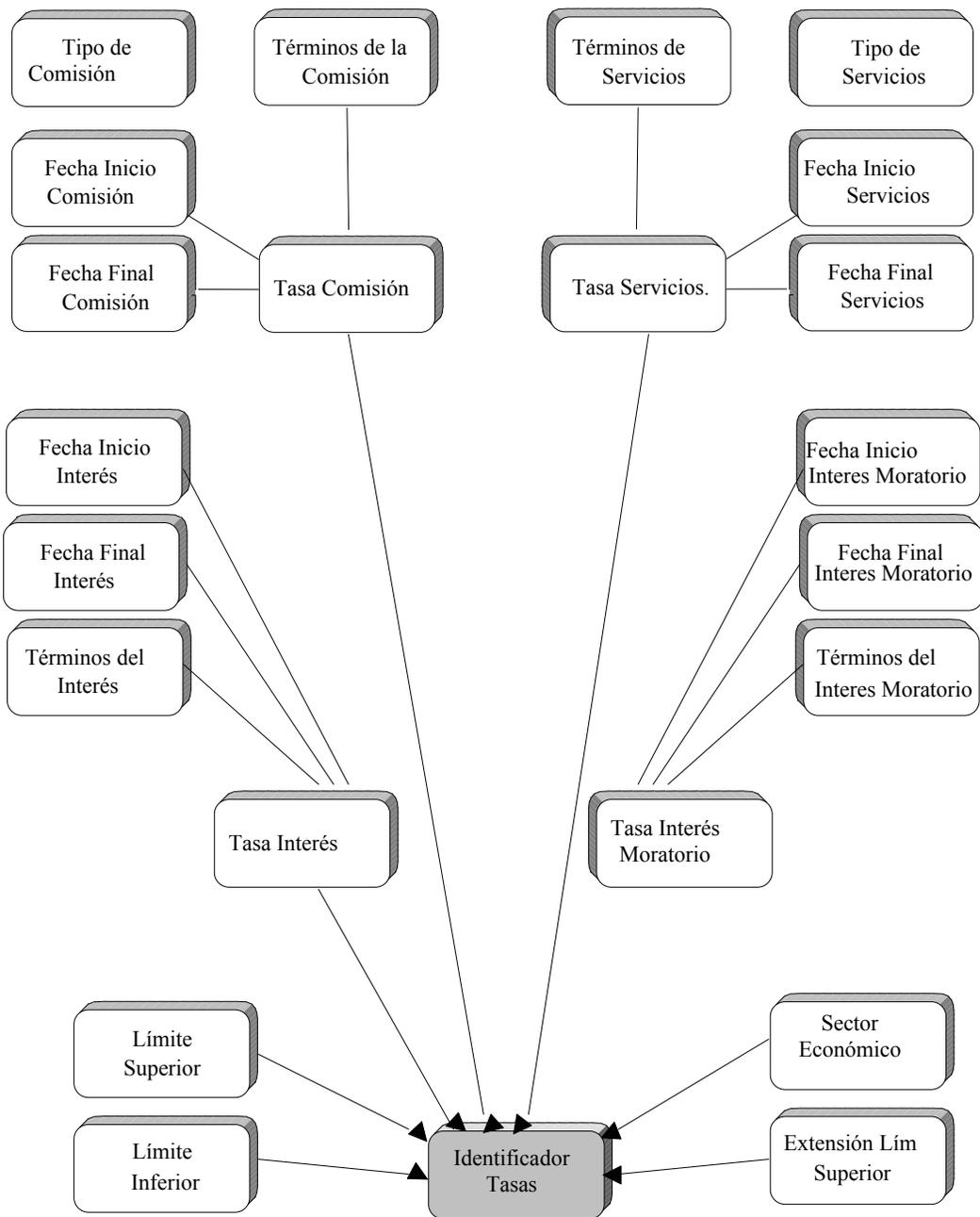
Nombre de los Atributos	Atributo Descripción	Valores de muestra
Identificador Cliente	Identificador o código actual para el cual se presta los servicios.	02-000405-1; 08-000010-3
Sucursal asignada al Cliente	Nombre de la sucursal específica que es responsable de manejar las cuentas del Cliente.	Matagalpa, Leon, Rio Blanco
Dirección de domicilio del Cliente	La dirección de domicilio para este Cliente. Nota: El barrio, distrito serán separados para una completa flexibilidad.	Casa Pellas Santa Ana 3c al Sur, Santa Ana, Distrito IV
Ciudad de domicilio del Cliente	El Nombre de la ciudad donde este Cliente tiene su domicilio.	Sébaco
Departamento de domicilio del Cliente	El Nombre del departamento donde este Cliente tiene su domicilio.	Matagalpa
Pais de domicilio del Cliente	El Nombre del país donde este Cliente tiene su domicilio.	Nicaragua
Cliente	Identificador único por cada Cliente. Este será representado de múltiples maneras incluyendo nombre e identificador de Cliente para los Clientes naturales	Talleres Rapidito.; Susana Báez
Tipo de Cliente	Indica si el Cliente es una personal jurídica o natural.	Natural; Jurídico
Sexo Cliente	Indica el sexo del Cliente natural.	Femenino, Masculino
Fecha de Nacimiento del Cliente	Indica la fecha de nacimiento del Cliente natural.	07/03/1960
Perfil Demográfico del Cliente	Cada Cliente natural es asignado al perfil demográfico que mejor se adapta.	Profesional Soltero, Padre de Familia
Actividad Económica	Cada Cliente natural es asignado al perfil de actividad económica que mejor se adapta a las labores que desempeña	Farmacia, Elaboración de Bebida, Ferretería
Casa	Cientes natural son agrupados en aquellos habitando en una unica residencia.	1000; 2000
Nivel de Educación del Cliente	Indica el más alto nivel de educación formal que ha completado el Cliente.	Secundaria; Universidad; Maestria
Dirección de Cobro	La dirección de cobro para este Cliente. Nota: El barrio, distrito serán separados para una flexibilidad completa.	Casa Pellas Santa Ana 3c al Sur, Santa Ana, Distrito IV
Cobro a Ciudad	El Nombre de la ciudad de la dirección de cobro de este Cliente.	Sébaco
Cobro a Departamento	El Nombre del departamento de la dirección de cobro de este Cliente.	Matagalpa
Industria	Para Clientes comerciales, esto indica el estándar industrial que mejor describe sus negocios.	Taller Mecánico.; Farmacia
Grupo Industrial	Un agrupamiento de industrias en categorías similares.	Salud, Hidráulica
Fecha de Primer Servicio	Indica la Fecha cuando se realize negocios con este Cliente por primera vez	1/04/1989
Clasificación Cliente	Indica si un cliente ha utilizado más una o más de una vez los servicios de la empresa	Nuevo, Repetitivo
Calificación Cliente	Agrupar a los clientes en categorías según su historial de pago, mora promedio, número de créditos	AAA; B

Dimensión Producto

Dimensión Producto Descripciones de Atributos

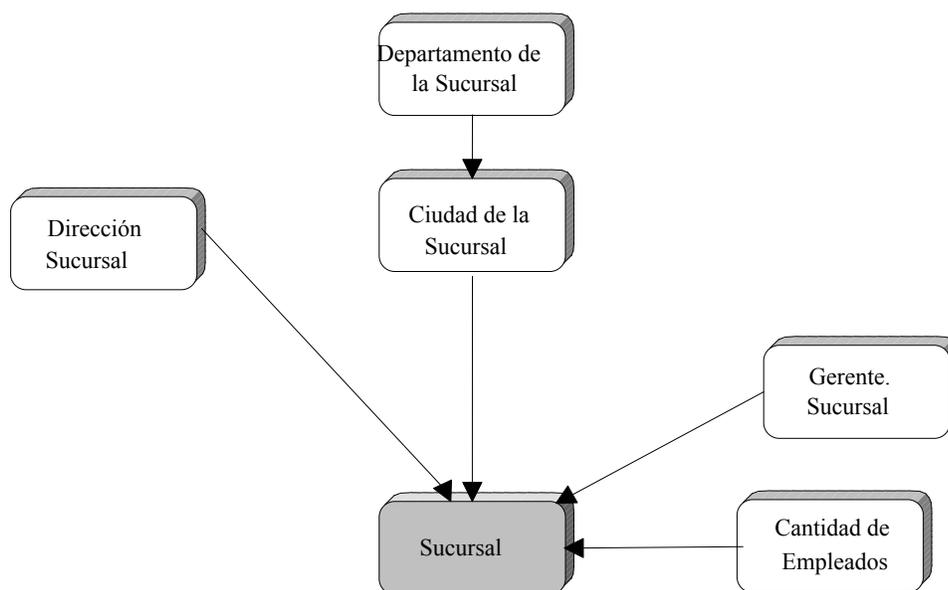
Nombre de Atributos	Descripción de Atributos	Valores de Muestra
Producto/Servicio	Describe el producto individual o servicio ofrecido, que un Cliente adquirió	Ordinario, Ra-Flá, Línea de Crédito, Personal
Sector Economico	Indica el sector económico que mejor se ajusta al perfil del producto que solicita el cliente	Industria, comercio, personal, servicio, agrícola, ganadero e hipotecario
Destino Crédito	Indica para que será utilizado el los fondos del crédito otorgado	Credicompu, Capital de Trabajo, Activo Fijo
Clase de producto	Una colección de productos	Productos para la micro empresa. Productos familiares
Categoría Producto	Una colección de clases de producto	Crédito, Ahorro

Dimensión Tasas



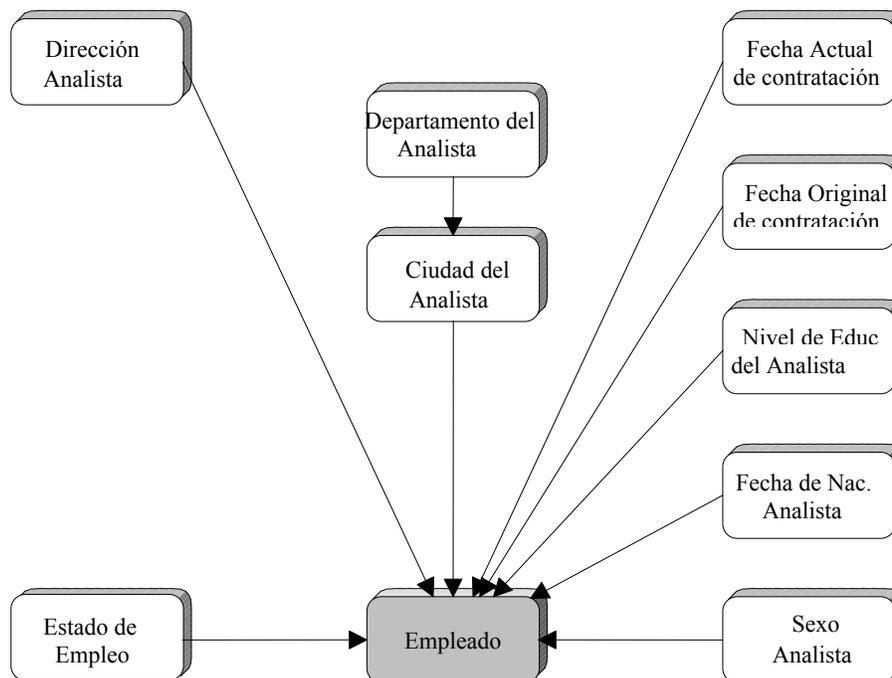
Dimensión Tasas Descripciones de Atributos

Nombre de Atributos	Descripción de Atributos	Valores de Muestra
Identificador de Tasa	Describe de manera unica una colección de información de las tasas que se le asignan al producto solicitado por el cliente.	001; 002
Limite inferior	Indica cual es el monto que se fija como límite inferior para aplicarle la tasa actual	10,000
Limite superior	Indica cual es el monto que se fija como límite superior para aplicarle la tasa actual	50,000
Extensión de limite superior	Describe la extensión en el monto del límite superior que es posible utilizar en algunos créditos especiales	10,000
Sector Económico	Indica el sector económico que se le asocia a una tasa individual	Ganadería; Comercio
Tasa de Interés	Indica el interés aplicado en la tasa actual definido de forma porcentual y en valores diarios	0.0833 %; 0.0667 %
Fecha Inicio Interés	La Fecha en que este tipo de interés se puso a la disposición de los clientes	04/15/1996
Fecha Final Interés	La Fecha en que este tipo de interés ya no estuvo disponible para los clientes.	3/26/1997
Términos del Interés	Indica los términos o condiciones especiales que también pueden aplicar a este interés.	Plazo mínimo 1 año
Tasa de comisión	Indica la comisión definida de forma porcentual que se aplica sobre cada desembolso al utilizar la tasa actual	3.5 %;
Tipo de Comisión	Forma en la que se determina será pagada la comisión	Deducida; Financiada
Fecha Inicio Comisión	La Fecha en que este tipo de comisión se puso a la disposición de los clientes	04/15/1996
Fecha Final Comisión	La Fecha en que este tipo de comisión ya no estuvo disponible para los clientes.	3/26/1997
Términos de la Comisión	Indica los términos o condiciones especiales que pueden aplicarse a esta comisión.	Nunca puede ser financiada
Tasa de Servicios	Indica la tasa de servicios expresada en valor porcentual que se aplica de forma anual en la tasa actual	1.5 % 6.0 %
Tipo de Servicios	Forma en la que se determina serán pagados los servicios	Financiados, Al vencimiento
Fecha Inicio Servicios	La Fecha en que este tipo de servicios se puso a la disposición de los clientes	04/15/1996
Fecha Final Interés	La Fecha en que este tipo de servicios ya no estuvo disponible para los clientes.	3/26/1997
Términos de los Servicios	Indica los términos o condiciones especiales que también pueden aplicar a este servicio.	Pagar al final del periodo
Tasa de Interés Moratorio	Indica el interés moratorio aplicado en la tasa actual definido de forma porcentual y en valores diarios	0.0833 %; 0.0667 %
Fecha Inicio Interés Moratorio	La fecha en que este tipo de interés moratorio se puso a la disposición de los clientes	04/15/1996
Fecha Final Interés Moratorio	La fecha en que este tipo de interés moratorio ya no estuvo disponible para los clientes.	3/26/1997

Dimensión Sucursal

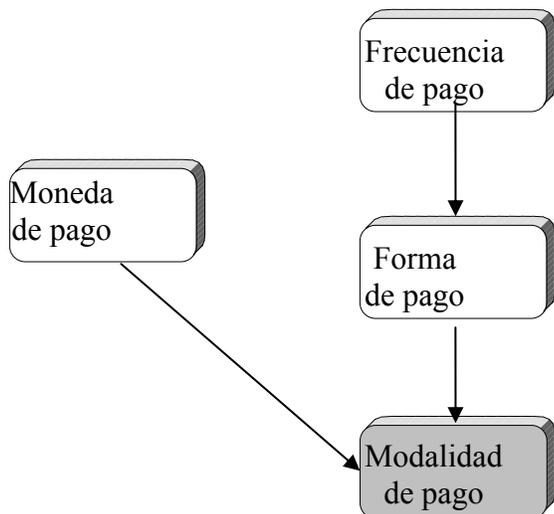
Dimensión Sucursal Descripciones de Atributos

Nombre de Atributos	Descripción de Atributos	Valores de Muestra
Sucursal	Representa las oficinas sucursales que la organización tiene en distintas localidades del país	Managua, Ivan Montenegro
Gerente Sucursal	Indica quien fungia en el cargo de gerente y dirigía la sucursal .	Pedro López, Aminta Solís
Dirección Localidad Sucursal	La dirección de localidad para esta sucursal.	
Ciudad Localidad Sucursal	El Nombre de la ciudad donde esta localizada la sucursal	Managua
Departamento Localidad Sucursal	El Nombre del departamento donde esta localizada la sucursal.	Managua
Cantidad de Empleados	Indica la cantidad de empleados que se encuentran operando en la sucursal.	10, 7, 12

Dimensión Analista

Dimensión Analista Descripción de Atributos

Nombre de Atributos	Descripción de Atributos	Valores de Muestra
Analista	Representa a los individuos contratados por Findesa. Analistas podrá ser visto por el número de Analista, Nombre, cédula y número de INSS.	Carla Chávez, Braudilio Urey
Estatus de empleo	El status de los individuos contratados.	Activo; Retirado; salio de la empresa
Sexo del Analista	Indica el sexo de este Analista.	Masculino; Femenino
Fecha de nacimiento del Analista	Indica la fecha de nacimiento de este Analista.	04/24/1974
Nivel de Educación del Analista	Indica el más alto nivel de educación formal que el Analista ha completado.	Secundaria, Licenciatura, Maestría
Fecha Original de Contratación	Indica la primera fecha de contratación del Analista.	09/15/1996
Fecha Actual de Contratación	Indica la fecha más reciente que el Analista fue contratado.	08/01/2001
Dirección de contacto del Analista	La dirección de contacto para este Analista. Nota: El barrio, distrito serán separados para una completa flexibilidad.	Casa Pellas Santa Ana 3c al Sur, Santa Ana, Distrito IV
Ciudad de contacto del Analista	El Nombre de la ciudad donde se puede establecer contacto con este Analista.	Managua
Departamento de contacto del Analista	El Nombre del departamento donde se puede establecer contacto con este Analista.	Managua

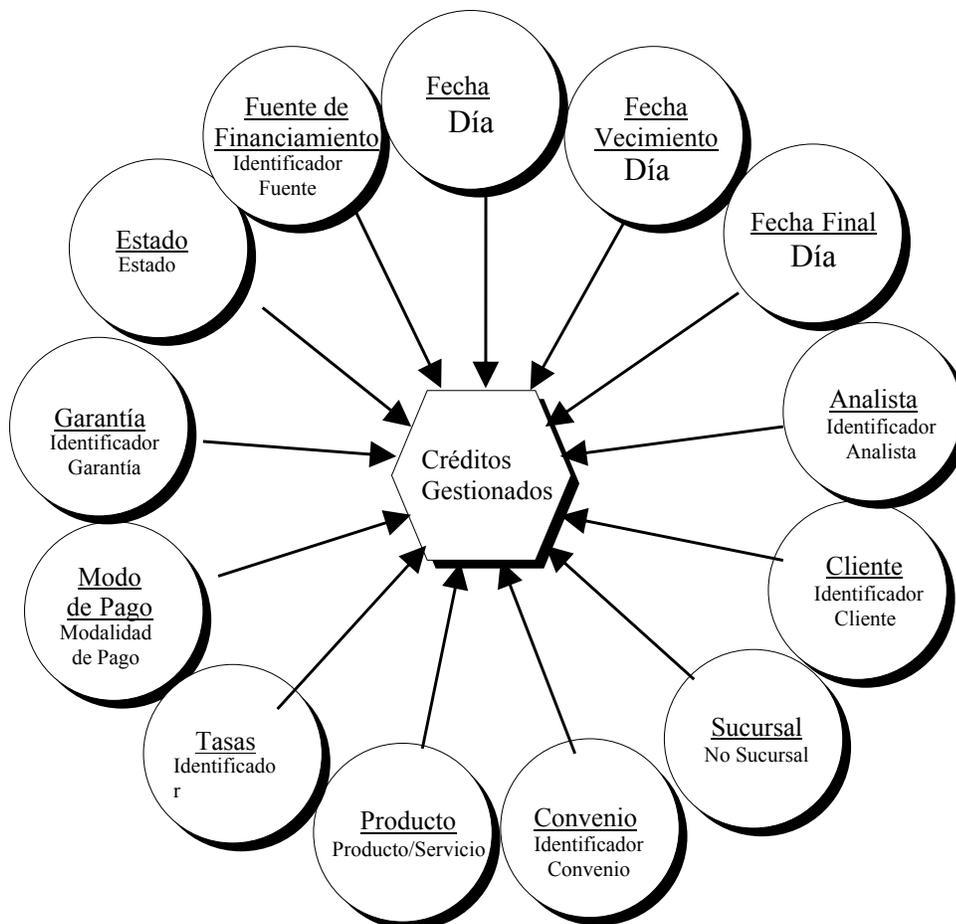
Dimensión Modo de Pago

Dimensión Modo de Pago Descripción de Atributos

Nombre de los Atributos	Descripción de Atributos	Valores de Muestra
Modalidad de Pago	Forma que identifica la modalidad de pago elegida por el cliente	Modo 1 Modo 2
Periodo de pago	Indica la forma que se elije los días de pago para pago de las cuotas en dependencia de la frecuencia de pago seleccionada	Fecha Fija Período Fijo
Frecuencia de pago	Representa la periodicidad que ha elegido un cliente para efectuar el pago de las cuotas del producto seleccionado.	Semanal, Catorcenal, Quincenal, Mensual, Bimensual, Trimestral, Semestral, Anual, Vencimiento
Cantidad de cuotas	Indica en cuantas cuotas un cliente desea cancelar su crédito	12, 24
Moneda de pago	Indica la moneda en la que se gestiona una transacción, pago o desembolso.	Córdobas, Dólares

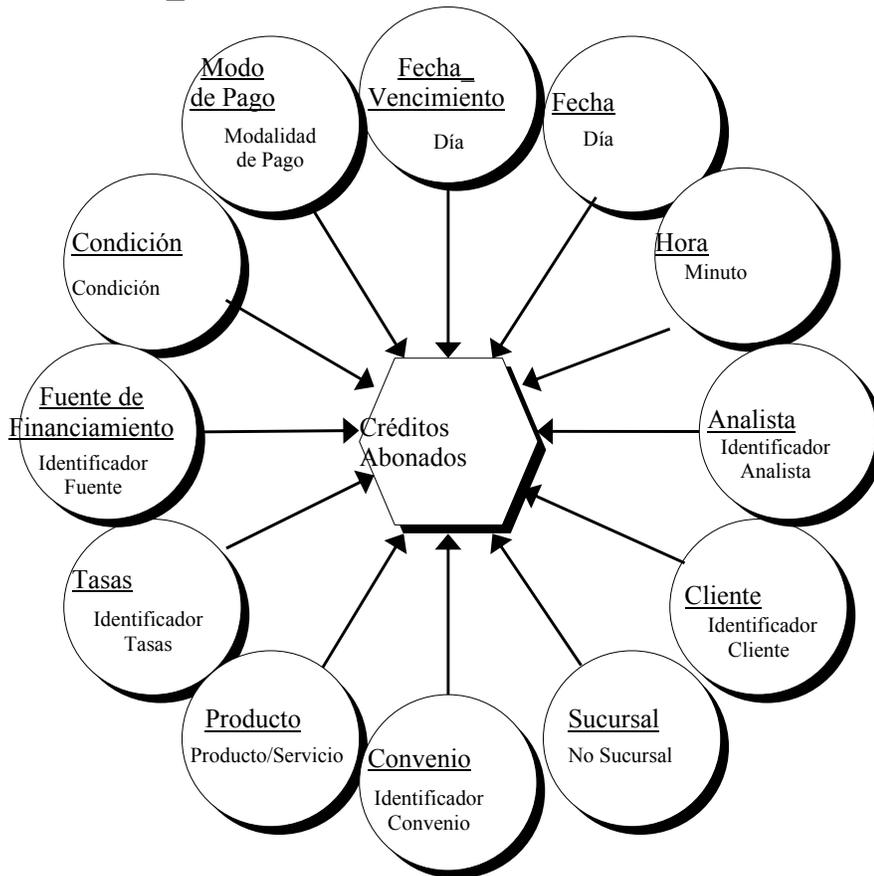
Detalle de los Hechos

Créditos _ gestionados



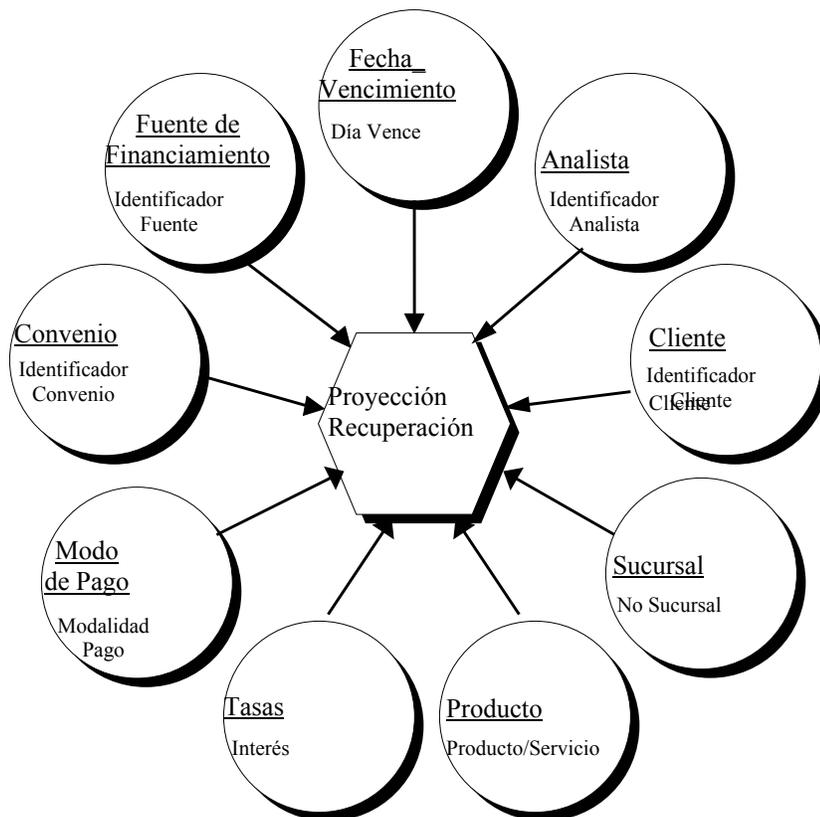
Hechos Descripción de los hechos Créditos Gestionados

Nombre Hecho	Descripción Hecho	RAD
No de crédito	Representa el número actual del crédito	N/A
Total Crédito	Monto total que pagará el cliente producto del financiamiento otorgado	Sum
Principal	Monto total aprobado para ser financiado en un crédito específico	Sum
Comisión	Cantidad total que se deduce en concepto de comisión	Sum
Servicios	Cantidad total que se cobra en concepto de servicios.	Sum
Mtto Valor	Cantidad total que se cobra en concepto de mantenimiento del Valor	Sum
Cargos Legales	Cantidad total que se paga en concepto de cargos legales: Pagarés, Contratos, Inscripciones	Sum
Intereses	Monto que se devenga en concepto de intereses por el crédito otorgado	Sum
Intereses moratorios	Monto que el cliente acumuló en concepto de intereses moratorios durante la vida del crédito	Sum
Valor Recargo	Monto total que el cliente ha pagado como multa por pagos tardíos	Sum
Días Plazo	El total de días de plazo para la cancelación del crédito.	N/A
Valor cuota	Indica el monto que abonara a su deuda el cliente cada vez	N/A
Monto Garantía	Indica el monto en que se valora la garantía presentada por el cliente	Sum

Créditos_Abonados Hechos**Créditos_Abonados Descripción de Hechos**

Nombre Hecho	Descripción Hecho	RAD
No Recibo de caja	Representa el número actual del recibo de caja que respalda el pago hecho por el cliente	N/A
No de crédito	Representa el número actual del crédito	N/A
No Desembolso	Representa el número de desembolso que se paga acualmente	
No de cuota	Indica el numero actual de la cuota que se paga	N/A
Abono final	Indica si el abono actual cancela el desembolso hecho	N/A
Monto total abono	Indica el total que el cliente abono a esta cuota en el día actual	Sum
Monto total cuota	Indica el monto total que el cliente debe pagar en esta cuota.	Sum
Abono a Principal	Monto que se abona al principal del crédito	Sum
Abono a Intereses	Monto que se abona a los intereses del crédito	Sum
Abono a Servicios	Monto que se abona a los servicios del crédito	Sum
Abono a Comisión	Monto que se abona a la Comisión del crédito	Sum
Mtto Valor	Monto que se paga en concepto de mantenimiento del valor	Sum
Dias_mora	Cantidad de días en que ha incurrido en mora esta cuota, contados después de la fecha de vencimiento de la cuota que paga	N/A
Monto Interes Mora	Cantidad pagada en concepto de interes moratorio	Sum
Recargo	Cantidad pagada en concepto de multa por pago tardío	Sum
Instrumento de pago	Indica el instrumento monetario con el que se efectúan pagos (Efectivo, Cheque personal, Cheque gerencia)	

Proyección_Recuperación hechos

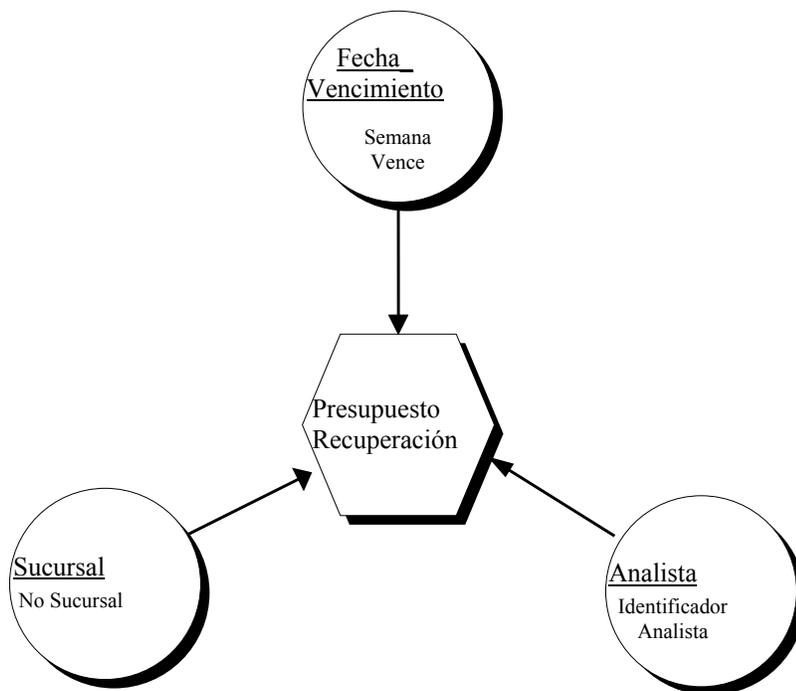


Proyección_Recuperación Descripción de Hechos

Nombre Hecho	Descripción Hecho	RAD
No de crédito	Representa el número del crédito que debe ser abonado	N/A
No Desembolso	Indica el numero desembolso de la cuota que se paga	N/A
No de cuota	Indica el numero actual de la cuota que se paga	N/A
Cuota final	Indica si la cuota actual cancela el desembolso hecho	N/A
Monto cuota	Indica el monto total que el cliente debe pagar de cuota.	Sum
Monto en Mora	Indica el monto total que el cliente debe pagar en esta cuota.	Sum
Principal	Monto que debe ser recuperado en concepto de principal del crédito	Sum
Intereses	Monto que debe ser recuperado en concepto de intereses del crédito	Sum
Servicios	Monto que debe ser recuperado en concepto de servicios del crédito	Sum
Comisión	Monto que debe ser recuperado en concepto de comisión del crédito	Sum
Mtto Valor	Monto que se paga en concepto de mantenimiento del valor	Sum

Nivel 2: grano Semanal

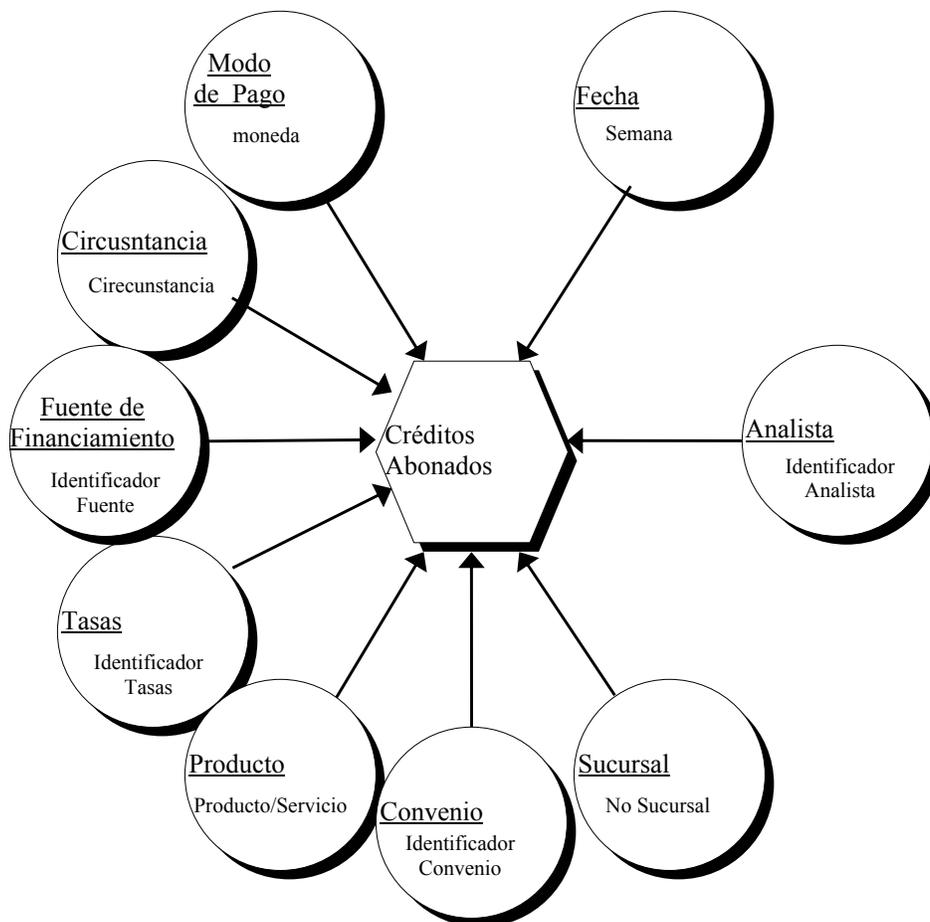
Presupuesto_Recuperación hechos



Presupuesto_Recuperación Descripción de Hechos

Nombre Hecho	Descripción Hecho	RAD
Presupuesto_Recup	Representa el valor que se presupuestó a ser recuperado para determinada semana del año	sum
Proyección_Recu	Indica el valor que se proyecta a recuperar para dicha semana del año (agregado semanal de la estrella proyección)	sum
Ejecutado_Recu	Indica el monto real que se recuperó en dicha semana	sum

Créditos_Abonados_Semanales Hechos



Créditos_Abonados Descripción de Hechos

Nombre Hecho	Descripción Hecho	RAD
Creditos cancelados	Cantidad de creditos que se cancelaron en la semana	sum
Monto total abono	Indica el total que la empresa recibio de abono a esta semana	Sum
Abono a Principal	Monto que se abona al principal	Sum
Abono a Intereses	Monto que se abona a los intereses	Sum
Abono a Servicios	Monto que se abona a los servicios	Sum
Abono a Comisión	Monto que se abona a la comisión	Sum
Mtto Valor	Monto que se paga en concepto de mantenimiento del valor	Sum
Monto Interes Moratorios	Cantidad pagada en concepto de interes moratorio	Sum
Recargo	Cantidad pagada en concepto de multa por pago tardío	Sum

ANEXO G: CORRESPONDENCIA FUENTE DESTINO Y RUTINAS ETL

Tabla de Correspondencia Fuente Destino

Tabla Destino	Columna Destino	Tipo de Datos	Long	Tabla / Archivo Fuente	Campo Fuente	Notas transformación de Datos
Cliente Dims	Cliente Clave	Integer				Campo identidad de la tabla
Cliente Dims	Cod Cliente	Varchar	9	climide	ccodcli	
Cliente Dims	Direccion Dom Cliente	Varchar	60	climide	cdirdom	separa direccion de barrio
Cliente Dims	Id Domicilio Cliente	Integer		climide	ccoddom	
Cliente Dims	Barrio Dom Cliente Descr	Varchar	40	climide	cdirdom	separar barrio de direccion
Cliente Dims	Distrito Dom Cliente	Varchar	5	climide		lista de barrio contra distrito
Cliente Dims	Sucursal Asignada	Varchar	20	climide	ccodofi	
Cliente Dims	Ciudad Dom Cliente	Varchar	40	climide	cdirdom	extraer ciudad de dirección domicilio
Cliente Dims	Depto Dom Cliente	Varchar	40	climide	ccodofi	asociar sucursal con depto
Cliente Dims	Pais Dom Cliente	Varchar	25	climide		prestamos solo a nacionales
Cliente Dims	Id Cliente	Varchar	17	climide	cnudoci	numero de cedula
Cliente Dims	Nombre Cliente	Varchar	100	climide	enomcli	
Cliente Dims	Tipo Cliente	Varchar	3	climide	cclaper	pesona juridica, natural con negocio, natural sin negocio
Cliente Dims	Fecha Primer Servicio	Datetime		climide	dregist	
Cliente Dims	Direccion Cobro Cliente	Varchar	60	climide		no disponible digitalmente por el momento
Cliente Dims	Barrio Cobro Cliente	Varchar	40	climide		no disponible digitalmente por el momento
Cliente Dims	Distrito Cobro Cliente	Varchar	5	climide		no disponible digitalmente por el momento
Cliente Dims	Ciudad Cobro Cliente	Varchar	40	climide		no disponible digitalmente por el momento
Cliente Dims	Depto Cobro Cliente	Varchar	20	climide		no disponible digitalmente por el momento
Cliente Dims	Calificacion Cliente	Varchar	3	climide	ccalcli	otras calificaciones son externas
Cliente Dims	Clasificacion Cliente	Varchar	10	climide	cclaci	
Cliente Dims	Sexo Cliente	Varchar	9	climide	csexo	cambiar codigo a etiqueta de sexo
Cliente Dims	Fecha Nacimiento Cliente	Datetime		climide	dnacimi	
Cliente Dims	Demog Cliente	Varchar	20	climide	cesteiv	cambiar soltero, casado a perfil demog
Cliente Dims	Activ Econo Cliente	Varchar	30	climide		no disponible digitalmente por el momento
Cliente Dims	Educacion Cliente	Varchar	12	climide		no disponible digitalmente por el momento
Cliente Dims	Casa Cliente	Varchar	5	climide		no disponible digitalmente por el momento
Cliente Dims	Industria	Varchar	20	climide		no disponible digitalmente por el momento
Cliente Dims	Grupo Industria	Varchar	20	climide		no disponible digitalmente por el momento
Sucursal Dims	Sucursal Clave	Integer				Campo identidad de la tabla
Sucursal Dims	Id Sucursal	Varchar	2	tabtofi	ccodofi	
Sucursal Dims	Nombre Sucursal	Varchar	20	tabtofi	ccodofi	establecer etiqueta de acuerdo al codigo
Sucursal Dims	Direccion Sucursal	Varchar	60			incluir direccion en rutina programada
Sucursal Dims	Barrio Sucursal	Varchar	35			incluir barrio en rutina programada
Sucursal Dims	Distrito Sucursal	Varchar	4			incluir distrito en rutina programada
Sucursal Dims	Ciudad Sucursal	Varchar	20	tabtofi	ccodofi	establecer etiqueta de acuerdo al codigo
Sucursal Dims	Depto Sucursal	Varchar	35	tabtofi	ccodofi	establecer etiqueta de acuerdo al codigo
Sucursal Dims	Gerente Sucursal	Varchar	60			incluir distrito en rutina programada
Sucursal Dims	Cant Empl Sucursal	Integer				incluir distrito en rutina programada
Hora Dims	Hora Clave	Integer				Campo identidad de la tabla
Hora Dims	Minuto	Integer		cjgdlog	chortrx	separar minuto
Hora Dims	Hora	Integer		cjgdlog	chortrx	separar hora y pasar a formato am/pm
Hora Dims	Bandera Am Pm	Varchar	2	cjgdlog	chortrx	determinar bandera
Analista Dims	Analista Clave	Integer				Campo identidad de la tabla
Analista Dims	Id Analista	Varchar	4	Tabtusu	ccodana	
Analista Dims	Nombre Analista	Varchar	60	Tabtusu	enomana	
Analista Dims	Estado Analista	Varchar	20	Tabtusu		
Analista Dims	Sexo Analista	Varchar	9			no disponible digitalmente por el momento
Analista Dims	Fecha Nac Analista	Datetime				no disponible digitalmente por el momento

Tabla Destino	Columna Destino	Tipo de Datos	Long	Tabla / Archivo Fuente	Campo Fuente	Notas transformación de Datos
Analista Dims	Educacion_Analista	Varchar	12			no disponible digitalmente por el momento
Analista Dims	Fecha_Orig_Contr_Analista	Datetime				no disponible digitalmente por el momento
Analista_Dims	Fecha_Actual_Contr_Analista	Datetime				no disponible digitalmente por el momento
Analista Dims	Direccion_Dom_Analista	Varchar	60			no disponible digitalmente por el momento
Analista Dims	Barrio_Dom_Analista	Varchar	40			no disponible digitalmente por el momento
Analista Dims	Distrito_Dom_Analista	Varchar	5			no disponible digitalmente por el momento
Analista_Dims	Ciudad_Dom_Analista	Varchar	40			no disponible digitalmente por el momento
Analista_Dims	Departamento_Analista	Varchar	40			no disponible digitalmente por el momento
Fecha Dims	Fecha Clave	Integer				Campo identidad de la tabla
Fecha Dims	Dia	Datetime				generado por procedimiento programado
Fecha Dims	No Dia En Mes Calendar	Integer				generado por procedimiento programado
Fecha Dims	No Dia En Anio Calendar	Integer				generado por procedimiento programado
Fecha Dims	No Dia En Mes Fiscal	Integer				generado por procedimiento programado
Fecha Dims	No Dia En Anio Fiscal	Integer				generado por procedimiento programado
Fecha Dims	Dia De Semana	Varchar	10			generado por procedimiento programado
Fecha Dims	Festivo Bandera	Varchar	2			generado por procedimiento programado
Fecha Dims	Tipo De Dia	Varchar	2			generado por procedimiento programado
Fecha Dims	Semana Calendar	Datetime				generado por procedimiento programado
Fecha Dims	No Semana Calendar	Varchar	4			generado por procedimiento programado
Fecha Dims	Mes Calendar	Varchar	16			generado por procedimiento programado
Fecha Dims	No Mes Calendar	Integer				generado por procedimiento programado
Fecha Dims	Trimestre Calendar	Varchar	7			generado por procedimiento programado
Fecha Dims	No Trimestre Calendar	Integer				generado por procedimiento programado
Fecha Dims	Anio Calendar	Integer				generado por procedimiento programado
Fecha Dims	Semana Fiscal	Varchar	13			generado por procedimiento programado
Fecha Dims	No Semana Fiscal	Integer				generado por procedimiento programado
Fecha Dims	Mes Fiscal	Varchar	18			generado por procedimiento programado
Fecha Dims	No Mes Fiscal	Integer				generado por procedimiento programado
Fecha Dims	Trimestre Fiscal	Varchar	9			generado por procedimiento programado
Fecha Dims	No Trimestre Fiscal	Integer				generado por procedimiento programado
Fecha Dims	Anio Fiscal	Integer				generado por procedimiento programado
Modo Pago Dims	Modo Pago Clave	Integer				Campo identidad de la tabla
Modo Pago Dims	Modo Pago	Varchar	7			establecer etiqueta de acuerdo al id
Modo Pago Dims	Id Periodo Pago	Varchar	2	credscs	ctipper	
Modo Pago Dims	Periodo Pago	Varchar	12	credscs	ctipper	establecer etiqueta de acuerdo al id
Modo Pago Dims	Id Frecuencia Pago	Varchar	2	credscs	cdiasug	
Modo Pago Dims	Descr Frecuencia Pago	Varchar	15	credscs	cdiasug	establecer etiqueta de acuerdo al id
Modo Pago Dims	Id Moneda Pago	Varchar	1	credscs	cmoneda	
Modo Pago Dims	Moneda Pago	Varchar	9	credscs	cmoneda	establecer etiqueta de acuerdo al id
Producto Dims	Producto Clave	Integer				Campo identidad de la tabla
Producto Dims	Id Producto	Varchar	2	cremcre	ccodprd	
Producto Dims	Descr Producto	Varchar	12	cremcre	ccodprd	establecer etiqueta de acuerdo al id
Producto Dims	Descr Producto Larga	Varchar	60	cremcre	ccodprd	establecer etiqueta de acuerdo al id
Producto Dims	Id Sector Econo	Varchar	2	cremcre	csececo	
Producto Dims	Descr Sector Econo	Varchar	30	cremcre	csececo	establecer etiqueta de acuerdo al id
Producto Dims	Id Destino	Varchar	2	cremcre	cdescre	
Producto Dims	Descr Destino	Varchar	30	cremcre	cdescre	establecer etiqueta de acuerdo al id
Producto Dims	Id Clase Producto	Varchar	2	cremcre		no establecidas todavia
Producto Dims	Descr Clase Producto	Varchar	60	cremcre		no establecidas todavia
Producto Dims	Id Categoria Producto	Varchar	2	cremcre		por el momento solo existe 01
Producto Dims	Descr Categoria Producto	Varchar	60	cremcre		categoria 01 corresponde a crédito

Tabla Destino	Columna Destino	Tipo de Datos	Long	Tabla / Archivo Fuente	Campo Fuente	Notas transformación de Datos
Abono_Credito_Facts	Fecha_Clave	Integer		credkar	dfecsis	asociar fecha con la clave en fecha_dim
Abono_Credito_Facts	Fecha_Vence_Clave	Integer		credkar	dfecpro	asociar fecha con la clave en fecha_vence_dim
Abono_Credito_Facts	Hora_Clave	Integer		cjgdlog	dfectrx	obtener la hora con respecto a la fecha
Abono_Credito_Facts	Analista_Clave	Integer		credkar	ccodcta	asociar la cta con el analista en cremcre
Abono_Credito_Facts	Cliente_Clave	Integer		credkar	ccodcta	asociar la cta con el cliente en cremcre
Abono_Credito_Facts	Sucursal_Clave	Integer		credkar	ccodcta	left(ccodcta,2) para el codigo oficina
Abono_Credito_Facts	Convenio_Clave	Integer		credkar	ccodcta	asociar la cta con el convenio en cremcre
Abono_Credito_Facts	Producto_Clave	Integer		credkar	ccodcta	asociar la cta con ccodpro, cdescre y csececo en cremcre
Abono_Credito_Facts	Tasas_Clave	Integer		credkar	ccodcta	
Abono_Credito_Facts	Modo_Pago_Clave	Integer		credkar	ccodcta	asociar la cta con datos en credscs
Abono_Credito_Facts	Fuente_Finan_Clave	Integer		credkar	ccodcta	asociar la cta con la cfuefin en cremcre
Abono_Credito_Facts	Circunstancia_Clave	Integer		credkar	ccodcta	asociar la cta con datos en cremcre
Abono_Credito_Facts	Monto_Total_Abono	Numeric	82	credkar	nmonto	revisar el cconcept el asociado con cj(caja)
Abono_Credito_Facts	Principal_Abono	Numeric	82	credkar	nmonto	revisar el cconcept el asociado con kp capital
Abono_Credito_Facts	Interes_Abono	Numeric	72	credkar	nmonto	in
Abono_Credito_Facts	Servicios_Abono	Numeric	72	credkar	nmonto	cconcept='02'
Abono_Credito_Facts	Interes_Mora	Numeric	72	credkar	nmonto	mo
Abono_Credito_Facts	Dias_Mora	Integer		credkar	dfecpro - dfecsis	efectuar calculo de dfecpro_dfecsis
Abono_Credito_Facts	No_Recibo_Caja	Varchar	10	credkar	cnrodoc	
Abono_Credito_Facts	No_Credito	Varchar	12	credkar	ccodcta	
Abono_Credito_Facts	No_Cuota	Integer		credkar	cnrocuo	
Abono_Credito_Facts	Monto_Total_Cuota	Numeric	72	credppg,credgas	ncapita-ncappag +intere - nintpag	
Abono_Credito_Facts	Abono_Final	Varchar	2	credppg		Numero max del abono para la cta en credppg
Gestion_Credito_Facts	Fecha_Clave	Integer		credkar	dfecsis	asociar fecha con la clave en fecha_dim
Gestion_Credito_Facts	Fecha_Vence_Clave	Integer		credkar	dfecpro	asociar fecha con la clave en fecha_vence_dim
Gestion_Credito_Facts	Fecha_Final_Clave	Integer		credppg	dfecven	asociar fecha con la clave en fecha_final_dim
Gestion_Credito_Facts	Analista_Clave	Integer		cremcre	ccodana	asociar analista con la clave en analista_dim
Gestion_Credito_Facts	Cliente_Clave	Integer		cremcre	ccodcli	asociar el cliente con la clave en cliente_dim
Gestion_Credito_Facts	Sucursal_Clave	Integer		cremcre	ccodcta	left(ccodcta,2) para el codigo oficina
Gestion_Credito_Facts	Convenio_Clave	Integer		cremcre	ccodgru	asociar el convenio con la clave en convenio_dim
Gestion_Credito_Facts	Producto_Clave	Integer		cremcre	ccodpro, cdescre, csececo	asociar ccodpro, cdescre y csececo con la clave en producto_dims
Gestion_Credito_Facts	Tasas_Clave	Integer		cremcre		
Gestion_Credito_Facts	Modo_Pago_Clave	Integer		credscs	ccodcta	asociar la cta con datos en credscs
Gestion_Credito_Facts	Garantia_Clave	Integer		cremcre	ccodcta	asociar la cta con la cfuefin en cremcre
Gestion_Credito_Facts	Estado_Clave	Integer		credkar	ccodcta	asociar la cta con datos en cremcre
Gestion_Credito_Facts	Fuente_Finan_Clave	Integer		cremcre	cfuefin	obtenidas de las claves de las dimensiones
Gestion_Credito_Facts	Monto_Total_Credito	Numeric	17.2			
Gestion_Credito_Facts	Principal	Numeric	14.2	cremcre	nmonapr	
Gestion_Credito_Facts	Comision	Numeric	8.2	cremcre		
Gestion_Credito_Facts	Servicios	Numeric	8.2	cremcre		
Gestion_Credito_Facts	Mtto Valor	Numeric	6.2			
Gestion_Credito_Facts	Cargos Legales	Numeric	6.2			
Gestion_Credito_Facts	Intereses	Numeric	8.2	cremcre		
Gestion_Credito_Facts	Intereses Mora	Numeric	7.2	cremcre		
Gestion_Credito_Facts	Recargos	Numeric	8.2	cremcre		
Gestion_Credito_Facts	Plazo Dias	Integer		cremcre		

Tabla Destino	Columna Destino	Tipo de Datos	Long	Tabla / Archivo Fuente	Campo Fuente	Notas transformación de Datos
Gestion_Credito_Facts	Cuota	Numeric	10.2	cremcre		
Gestion_Credito_Facts	Monto_Garantia	Numeric	19.2	cremcre		
Gestion_Credito_Facts	No_Credito	Varchar	12	cremcre	ccodcta	
Proyeccion_Recu_Facts	Fecha_Vence_Clave	Integer			dfecven	asociar fecha con la clave en fecha_final_dim
Proyeccion_Recu_Facts	Analista_Clave	Integer			ccodana	asociar analista con la clave en analista_dim
Proyeccion_Recu_Facts	Cliente_Clave	Integer			ccodcli	asociar el cliente con la clave en cliente_dim
Proyeccion_Recu_Facts	Sucursal_Clave	Integer			ccodcta	left(ccodcta,2) para el codigo oficina
Proyeccion_Recu_Facts	Convenio_Clave	Integer			ccodgru	asociar el convenio con la clave en convenio_dim
Proyeccion_Recu_Facts	Producto_Clave	Integer			ccodpro, cdescre, csececo	asociar ccodpro, cdescre y csececo con la clave en producto_dims
Proyeccion_Recu_Facts	Tasas_Clave	Integer				obtenidas de las claves de las dimensiones
Proyeccion_Recu_Facts	Modo_Pago_Clave	Integer			ccodcta	asociar la cta con datos en credscs
Proyeccion_Recu_Facts	Fuente_Finan_Clave	Integer			ccodcta	asociar la cta con la cfuefin en cremcre
Proyeccion_Recu_Facts	Monto_Cuota	Numeric	10.2	credppg		igual que en abono
Proyeccion_Recu_Facts	Principal	Numeric	10.2	credppg	ncapita- ncappag	
Proyeccion_Recu_Facts	Servicios	Numeric	8.2	credgas	nmongas- nmopag	para ctipgas = 02
Proyeccion_Recu_Facts	Intereses	Numeric	8.2	credppg	intere - nintpag	
Proyeccion_Recu_Facts	Mtto Valor	Numeric	6.2	credppg		
Proyeccion_Recu_Facts	No_Credito	Varchar	12	credppg	ccodcta	
Proyeccion_Recu_Facts	No_Cuota	Integer		credppg	cnrocuo	
Proyeccion_Recu_Facts	Cuota_Final	Integer		credppg		igual que en abono

Rutinas de Extracción, Transformación y Carga

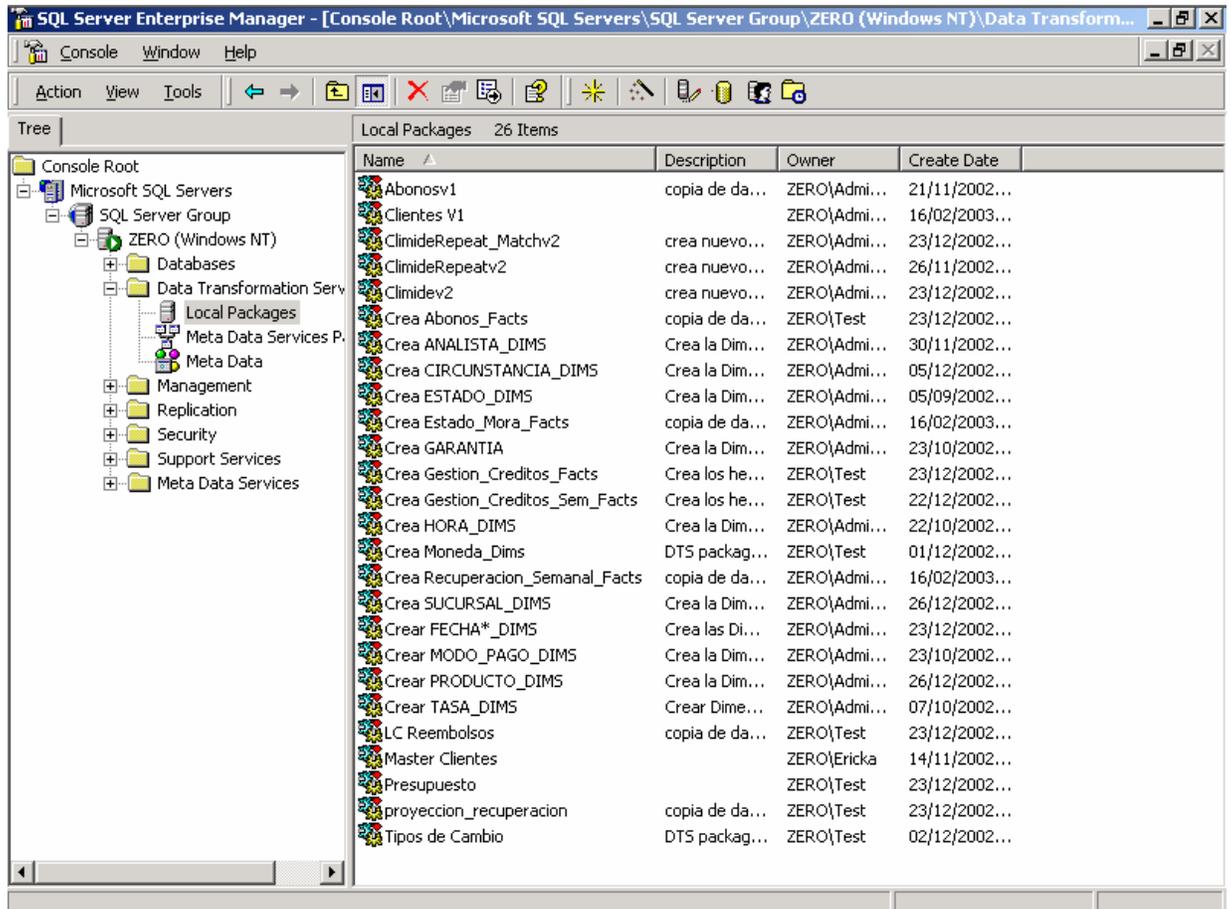


Figura 1. Conjuntos de Paquetes de DTS para la carga inicial de Datos.

En la figura anterior se puede observar dentro de la carpeta de Servicios de Transformación de Datos de Microsoft SQL Server 2000, los paquetes que se encargan de realizar el proceso de Extracción, Transformación y Datos (ETL). Los DTS se encargan de poblar el Data Mart con los datos que sean necesarios, de aplicar transformaciones que se requieran auxiliándose del área de preparación de datos, donde los datos intermedios son depositados temporalmente, mientras se encuentra en proceso, a la espera de ser transformados para llegar luego a su destino final.

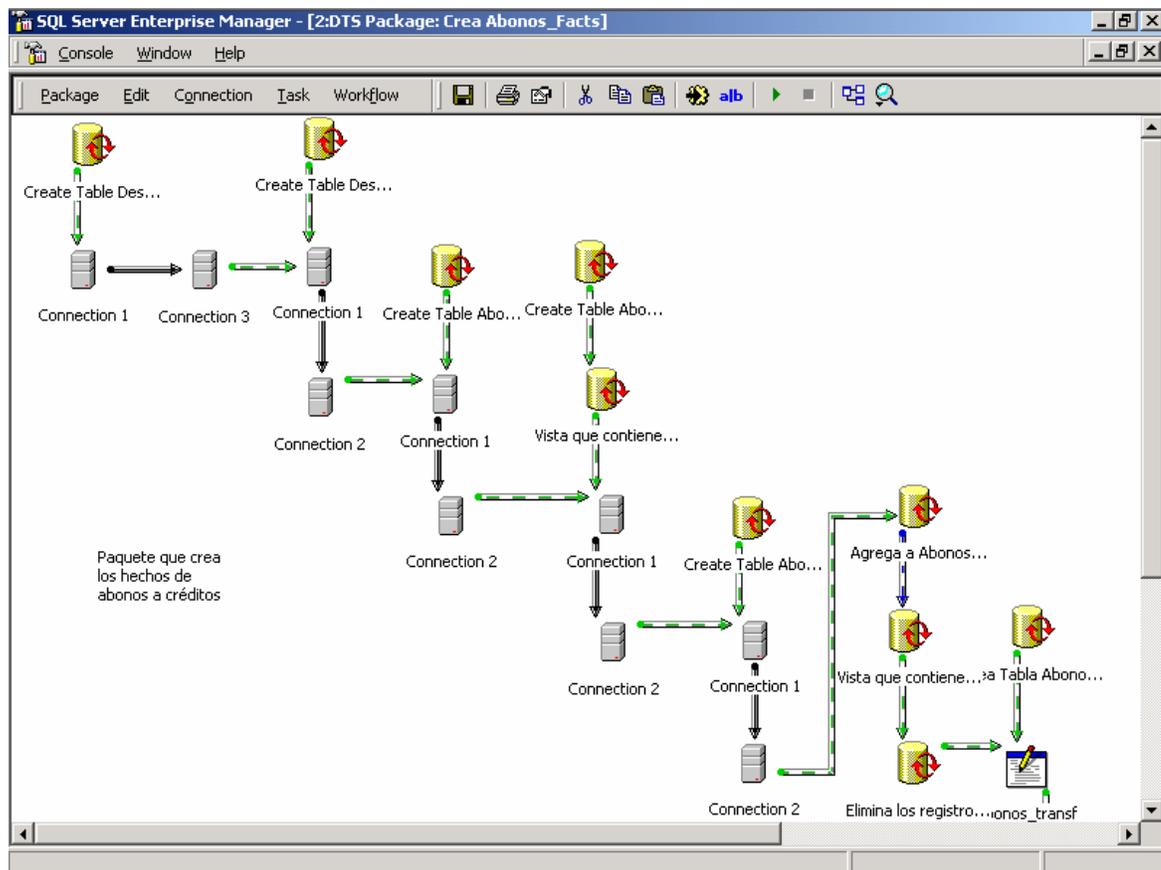


Figura 2- Paquete DTS del proceso de transformación de datos para la tabla Abono_Credito_Facts

En la figura se puede apreciar que un paquete DTS está conformado de un conjunto de tareas de diversa índole y conexiones a fuentes de datos. Para ejecutar las transformaciones requeridas se utiliza la ejecución de un proceso de tareas secuenciales, y en algunos casos en paralelo, la definición de la secuencia se puede observar claramente con los símbolos de flechas que muestran el orden que sigue el flujo de datos, éstas flechas indican cuáles son las siguientes tareas en ejecutarse en dependencia del éxito, finalización o fracaso de la tarea.

Los iconos de conexiones hacen referencia a orígenes de datos especificando una fuente y un destino para el proceso de transformación, también hay tareas que ejecutan sentencias SQL. Otro elemento que se observa en el extremo inferior derecho de la figura es el icono que corresponde a una llamada a programa y que permite el paso de los parámetros respectivos; esta herramienta es muy útil en los casos que necesiten la aplicación de programas que puedan realizar las transformaciones necesarias.

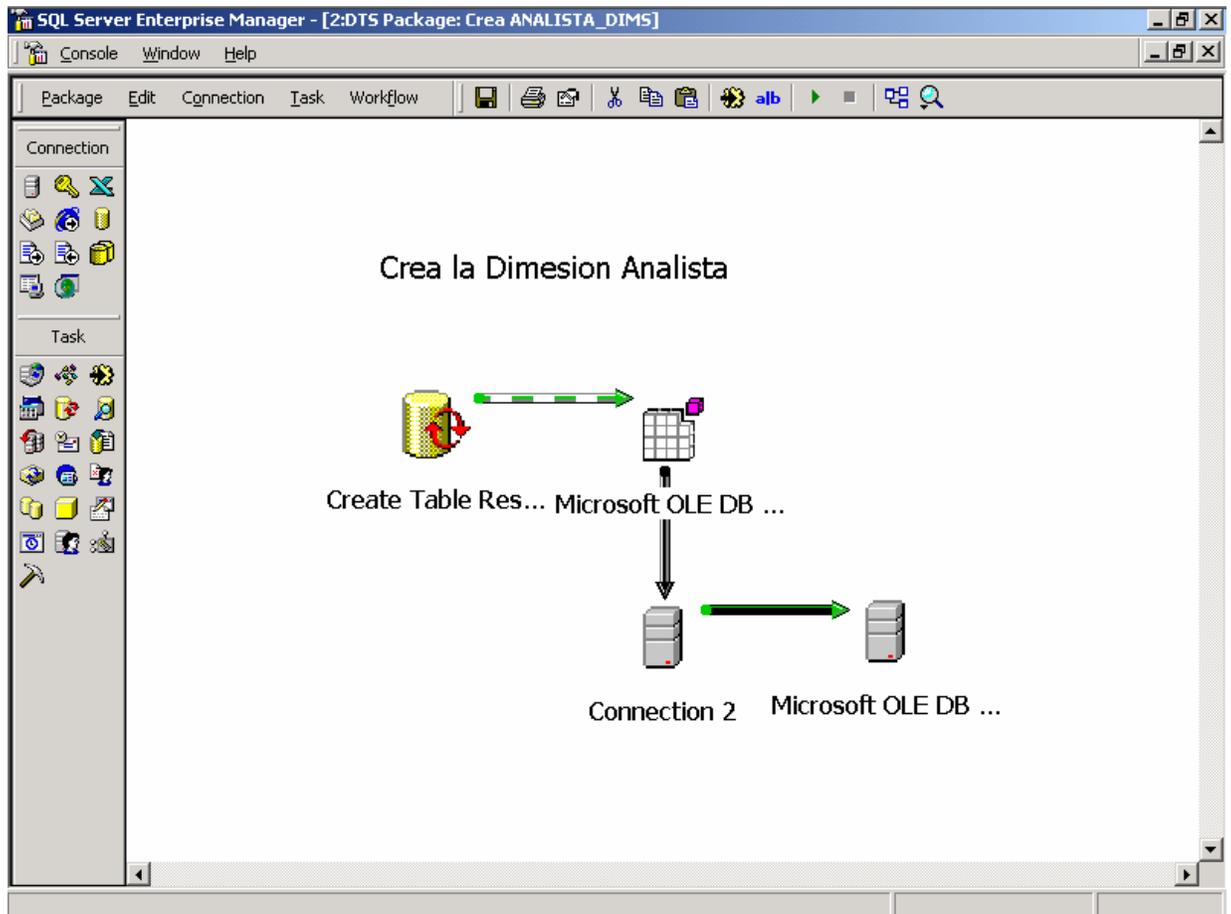


Figura 3- Paquete que crea la Dimension Analista

En la figura 3 se aprecia que cierta información del Data Mart debe ser poblada a partir de datos almacenados en tablas auxiliares que pertenecen al sistema transaccional, pero que no eran administradas por un Sistema Gestor de Bases de Datos, por lo que se necesitaba que la herramienta ETL contará con conectores que pudieran acceder a dicha fuente. Este es un elemento importante a la hora de elegir una herramienta de automatización de la ETL, la herramienta seleccionada debe ser capaz de obtener los datos necesarios a partir de todas las fuentes de Datos que se requieran, de lo contrario, se perdería mucha flexibilidad en el tratamiento que se le diera a los datos.

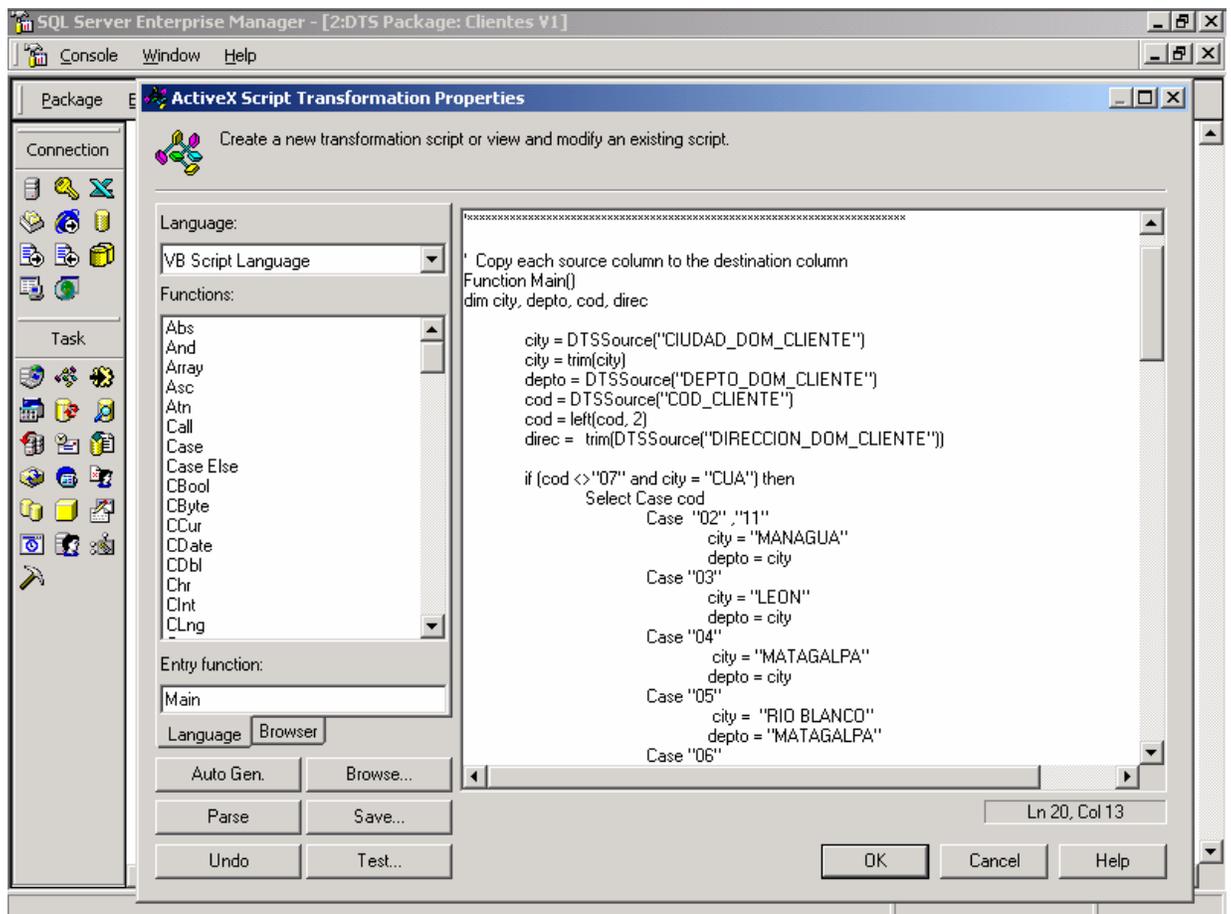


Figura 4- Código de tranformacion de una tarea

En la figura 4 se aprecia una muestra del código que define las transformaciones que sobre los registros de una tabla realiza una tarea que pertenece a un determinado paquete. En el paquete de la figura, en concordancia con lo previamente definido en la tabla de correspondencia se obtiene el valor que guardaran los campos ciudad y departamento de la dimensión Clientes, sobre la base de la dirección de domicilio del Cliente.

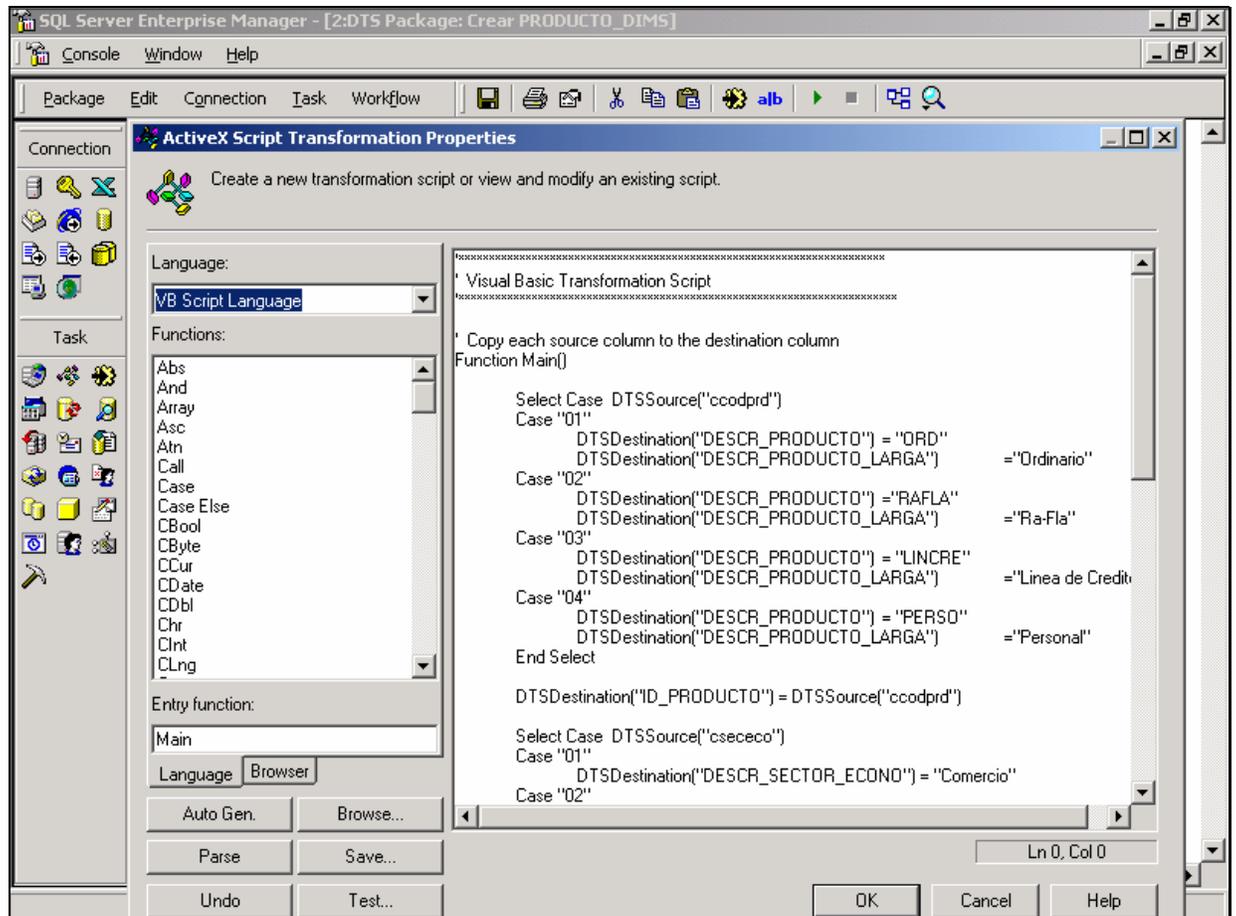


Figura 5 - Código en lenguaje Visual Basic Script

ANEXO H: REPORTES OBTENIBLES

INFORME DE RECUPERACIONES DE CREDITOS

Datos Demográficos de Clientes - Abonos Realizados

Edad	Toda Edad	Datos		Moneda Credito	
		Principal Abono		Principal Cor	
Genero	Demog Cliente	CORDOBAS	DOLARES	CORDOBAS	DOLARES
FEMENINO		8351305.81	390228.39	8351305.81	5652027.79
	CASADO	4809182.55	243297.36	4809182.55	3524759.83
	DIVORCIADO	79893.52	5055.33	79893.52	73262.67
	SOLTERO	2891173.05	113303.4	2891173.05	1640226.92
	UNION LIBRE	391393.74	9247.41	391393.74	133902.77
	VIUDO	178048.32	19324.89	178048.32	279875.6
	NO ESPECIFICA	1614.63	--	1614.63	--
MASCULINO		8816323.18	457045.11	8816323.18	6621769.47
	CASADO	6332986.36	355905.58	6332986.36	5157248.67
	DIVORCIADO	35595.58	3817.09	35595.58	55076.2
	SOLTERO	1960477.17	83449.66	1960477.17	1208361.27
	UNION LIBRE	460562.52	10111.85	460562.52	146624.09
	VIUDO	22436.62	366.18	22436.62	5313.37
	NO ESPECIFICA	4264.93	3394.75	4264.93	49145.87
Total general		17167628.99	847273.5	17167628.99	12273797.26

INFORME DE RECUPERACIONES DE CREDITOS

Producto, Sector Economico y Destino del Credito por Periodo y Sucursal - Abonos Realizados

Fecha Dia	Toda Fecha Dia			
Sucursal	Toda Sucursal			
Producto Clav	Sector Econo	Destino	Principal Abono	Principal Cor
Linea de Credito			11183519.38	11183519.38
	Agricola		391383.49	391383.49
		Activo Fijo	383787.27	383787.27
		Capital de Trabajo	7596.22	7596.22
	Comercio		8079886	8079886
		Activo Fijo	7892524.98	7892524.98
		Capital de Trabajo	178204.73	178204.73
		Mixto (CT+AF)	9156.29	9156.29
	Ganadero		485274.03	485274.03
		Activo Fijo	473804.6	473804.6
		Capital de Trabajo	11469.43	11469.43
	Hipotecario		2308.72	2308.72
		Activo Fijo	2308.72	2308.72
	Industria		974110.09	974110.09
		Activo Fijo	939771.83	939771.83
		Capital de Trabajo	26090.4	26090.4
		Mixto (CT+AF)	8247.86	8247.86
	Personal		210115.98	210115.98
		Activo Fijo	210115.98	210115.98
	Servicios		1040441.07	1040441.07
		Activo Fijo	1001812.33	1001812.33
		Capital de Trabajo	8509.59	8509.59
		Mixto (CT+AF)	30119.15	30119.15
Ordinario			2654235.64	2654235.64

INFORME DE GESTION DE CREDITOS

Producto, Sector Economico y Destino del Credito por Periodo y Sucursal

Fecha_Semana	2002
Sucursal	Toda Sucursal

Producto	Sector Econo	Destino	Creditos	Principal_cor
Linea de Credito			3157	C\$ 93,174,477.68
	Agricola		55	C\$ 4,301,304.48
		Activo Fijo	54	C\$ 4,290,208.40
		Capital de Trabajo	1	C\$ 11,096.08
	Comercio		2274	C\$ 66,864,562.28
		Activo Fijo	2246	C\$ 65,670,022.63
		Capital de Trabajo	27	C\$ 1,173,704.92
		Mixto (CT+AF)	1	C\$ 20,834.73
	Ganadero		82	C\$ 4,469,975.01
		Activo Fijo	82	C\$ 4,469,975.01
	Hipotecario		1	C\$ 30,000.00
		Activo Fijo	1	C\$ 30,000.00
	Industria		254	C\$ 6,820,669.20
		Activo Fijo	252	C\$ 6,685,810.74
		Capital de Trabajo	2	C\$ 134,858.46
	Personal		167	C\$ 2,778,328.90
		Activo Fijo	167	C\$ 2,778,328.90
	Servicios		324	C\$ 7,909,637.81
		Activo Fijo	323	C\$ 7,893,543.04
		Capital de Trabajo	1	C\$ 16,094.77
Ordinario			849	C\$ 43,095,446.28
	Agricola		31	C\$ 1,949,984.93
		Activo Fijo	31	C\$ 1,949,984.93
	Comercio		252	C\$ 13,519,478.34
		Activo Fijo	248	C\$ 13,458,923.77
		Capital de Trabajo	1	C\$ 19,996.99
		Mixto (CT+AF)	3	C\$ 40,557.58

INFORME DE GESTION DE CREDITOS

Sucursal y Estado por Periodo y Moneda del Credito

Moneda	Toda Moneda
Fecha_Semana	NOVIEMBRE

Sucursal	Id Estado		Datos	
	Vigente		Cancelado	
	Creditos	Principal_cor	Creditos	Principal_cor
Managua	242	C\$ 3,928,566.20	54	C\$ 1,162,109.79
Leon	174	C\$ 3,237,359.26	36	C\$ 661,592.13
Matagalpa	82	C\$ 2,833,788.55	52	C\$ 546,753.97
Rio blanco	179	C\$ 5,099,335.45	50	C\$ 822,634.16
Masaya	56	C\$ 1,415,076.96	42	C\$ 389,976.23
Esteli	76	C\$ 1,926,296.23	29	C\$ 504,324.85
Rivas	101	C\$ 1,896,087.74	49	C\$ 530,600.58
Sebaco	45	C\$ 1,950,734.72	22	C\$ 706,829.25
Chinandega	108	C\$ 2,485,051.60	17	C\$ 306,834.62
Ivan Montenegro	115	C\$ 2,945,076.26	24	C\$ 251,490.12
Total general	1178	C\$ 27,717,372.97	375	C\$ 5,883,145.70

ANEXO I: GLOSARIO

GLOSARIO

Agregación: Actividad de combinar datos desde múltiples tablas para formar una unidad de información más compleja, necesitada frecuentemente para responder consultas del Data Warehouse en forma más rápida y fácil.

Atributos: Conjunto de campos de las tablas de una base de datos transaccional o analítica que definen la información que se almacena en una tabla en particular

Base de datos multidimensional: Base de datos diseñada para procesamiento analítico on-line (*OLAP*). Estructurada como un hipercubo con un eje por dimensión.

Data Mart: Conjunto de hechos y datos organizados para soporte a la toma de decisiones basados en la necesidad de un área o departamento específico. y sus datos no tienen porque tener las mismas fuentes que los de otro DataMart

Data Mining: La extracción de información predecible escondida en grandes bases de datos para el análisis de los datos con el objetivo de descubrir relaciones, patrones, o asociaciones desconocidas.

Data Warehouse: Sistema de Base de Datos que almacena una gran cantidad de datos transaccionales integrados y enriquecidos para ser usados para análisis por usuarios especializados (encargados de tomar decisiones en la empresa).

Dato: Es la representación computarizada de la información en todo negocio. Existen muchas variedades de datos que se pueden almacenar en las computadoras. Los datos que deben ser incluidos en el Data Warehouse son aquellos datos comprensibles y consistentes en si mismos que son necesarios para manejar el negocio como un todo y en sus partes individuales (indicadores).

Datos anormales: Datos que resultan de errores (por Ej. Errores en el ingreso de datos durante la carga) o que representan eventos inusuales.

Datos Derivados: Datos que resultan luego de efectuar operaciones de transformación sobre un conjunto dado de datos origen, con el objetivo de ser de mayor significado para el proceso de toma de decisiones.

Datos Periódicos: Conjunto de datos generados o extraídos con cierta periodicidad a partir de los sistemas transaccionales para alimentar el entorno analítico de la organización

DBMS (Data Base Management System): Sistema Relacional que se encarga del control y seguridad de los datos organizados en tablas proveyéndoles de un entorno en que las aplicaciones clientes pueden realizar sus operaciones sin poner en riesgo la estructura de datos del modelo. Además realiza control de integridad y presenta la capacidad de realizar consultas a los datos almacenados.

Diccionario de Datos: Un compendio de definiciones y especificaciones para las categorías de datos y sus relaciones.

Dimensión: Conjunto de entidades similares agrupadas en una tabla; que sirve como llave de búsqueda (actuando como índice), o como mecanismo de selección de datos. Dentro del esquema estrella, corresponde a las tablas que están unidas a la tabla central a través de sus respectivas llaves. La cantidad de estas tablas le otorgan la característica de multidimensionalidad a esta estrategia.

Drill Down: Exponer progresivamente un nivel de mayor detalle del elemento que resulte seleccionado dentro de un reporte o consulta

Drill-Up: Es el efecto contrario a drill-down. Significa ver menos nivel de detalle, sobre la jerarquía significa generalizar o sumarizar, es decir, subir en el árbol jerárquico.

Entregables: Resultado que se espera obtener como producto de la aplicación de un procedimiento enunciado previamente.

ETL (Extraction, Transformation and Load): Pasos por los que atraviesan los datos para ir desde el sistema OLTP (o la fuente de datos utilizada) al entorno analítico. Extracción, se refiere al mecanismo por medio del cual los datos son leídos desde su fuente original. Transformación (también conocida como limpieza) es la etapa por la que puede atravesar una base de datos para estandarizar los datos de las distintas fuentes, normalizando y fijando una estructura para los datos. Finalmente está la Carga, que consiste básicamente en llevar los datos leídos y estandarizados al Data Warehouse (puede ser remota o localmente).

Hechos: Conjunto de medidas de negocio interrelacionadas que se agrupan en una tabla; con el objetivo de representar una evaluación cuantitativa del desempeño de la organización. Dentro del esquema estrella, corresponde a la tabla central a la que se encuentran unidas las dimensiones por medio de llaves externas y cuyo conjunto conforma la llave primaria de la tabla de hechos, lo cual permite que las dimensiones actúen como índice de búsqueda.

Jerarquía: Es un conjunto de atributos descriptivos que permite que a medida que se tenga una relación de muchos a uno se ascienda en la jerarquía. Por ejemplo: los Centros de Responsabilidad están asociados a un Tipo de Unidad, el cual pueden corresponder a una gerencia, subgerencia, superintendencia, etc.; por otro parte, cada CR está asociado a otro CR a nivel administrativo y, también existe una clasificación a nivel funcional.

Metadatos: Son datos que describen el significado de la estructura de los datos del negocio, así como también la manera en que éstos son creados, accedidos y usados, lo que asegurará que los datos del negocio sean usados de manera completa y consistente.

Modelo de Datos Corporativo: Representa la definición de datos que corresponde a la descripción de las actividades en las que se desempeña la empresa y de los objetos del mundo real (clientes, localidades y productos) con los que se interrelaciona la organización. Este modelo de datos sirve de fundamento para sistemas de procesamiento de transacciones y sistemas de soporte a la toma de decisiones.

ODS (Operational Data Store): es la construcción arquitectónica donde los datos operacionales que han sido integrados colectivamente son almacenados. Es una colección de datos integrados, orientados a temas (comúnmente basada en el cliente), tomada directamente de los sistemas operacionales para soportar monitoreo operacional.

OLAP - Procesamiento analítico on-line (On Line Analytical Processing): Conjunto de principios que proveen un ambiente de trabajo dimensional para soporte decisional. se refiere a aplicaciones de bases de datos orientadas a arreglos que permite a los usuarios ver, navegar, manipular y analizar *bases de datos multidimensionales*.

OLTP (On-line Transaction Processing): Sistema transaccional diario (o en detalle) que mantiene los datos operacionales del negocio.

Script: Programa o secuencia de instrucciones que es interpretado o ejecutado por otro programa en lugar del procesador de la computadora.

Sistema de Soporte de Decisiones: Sistema de aplicaciones automatizadas que asiste a la organización en la toma de decisiones mediante un análisis estratégico de la información histórica.

SGBD (Sistema Gestor de Bases de Datos): ver DBMS

Snapshot: Imagen instantánea de los datos contenidos en un sistema en un tiempo dado.

Sumarización: Actividad de incremento de la granularidad de la información en una base de datos. La sumarización reduce el nivel de detalle, y es muy útil para presentar los datos para apoyar al proceso de Toma de Decisiones.