



# UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE ESTUDIOS SUPERIORES  
ARAGÓN

ANÁLISIS, DISEÑO E IMPLEMENTACIÓN DE UN  
DATA MART DE DATOS OPERATIVOS DEL FONAES  
PARA MEDIR AVANCES FÍSICOS Y FINANCIEROS

**T E S I S**

QUE PARA OBTENER EL TÍTULO DE :  
INGENIERO EN COMPUTACIÓN  
P R E S E N T A :  
DIANA HERNÁNDEZ VARGAS



ASESOR:

M. EN C. JESÚS HERNÁNDEZ CABRERA

MÉXICO

2007



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **Agradecimientos**

A Dios, por darme la oportunidad de ser, de equivocarme y aprender.

A mi Padre, por ser siempre mi modelo a seguir, por enseñarme a creer en mí y lo más importante: su infinito amor.

A mi hija, por ser el principal impulso para ser una mejor persona, mujer y madre.

A Iván, por estar a mi lado.

A mis hermanos, por contar siempre con su apoyo y cariño.

A mi asesor, por su infinita paciencia e invaluable guía para concluir este trabajo.

A mis entrañables amigos: Rebeca, Walter y Alejandro, por tener la fortuna de conocerlos y contar con su amistad.

# Análisis, diseño e implementación de un Data mart de datos operativos del FONAES para medir avances físicos y financieros

<b>Introducción</b> .....	1
<b>1. Conceptos de Data Warehouse</b> .....	4
1.1. Definición de un DW.....	5
1.2. Propósito y naturaleza de un DW.....	7
1.3. Comparación entre un DW y una BD operacional.....	8
1.4. Características de un DW.....	9
1.5. DW y Datamart.....	12
1.6. Herramientas OLAP.....	19
<b>2. Metodología para la construcción de un DW</b> .....	21
2.1. Definición de Objetivos.....	22
2.2. Definición de Requerimientos.....	23
2.3. Diseño y modelado.....	26
2.4. Implementación.....	36
2.4.1. Proceso ETL.....	37
2.4.2. Explotación de la Información.....	42
2.5. Revisión.....	43
2.6. Administración y mantenimiento.....	44
<b>3. Caso práctico: Análisis, diseño de un Data mart de datos operativos del FONAES para medir avances físicos y financieros</b> .....	45
3.1. Planteamiento detallado del problema.....	45
3.2. Definición de Objetivos.....	46
3.3. Definición de Requerimientos.....	47
3.3.1. Requerimientos de Negocio.....	47
3.3.2. Requerimientos arquitectónicos.....	52
3.4. Diseño y Modelado.....	54
3.4.1. Diseño arquitectónico.....	54
3.4.2. Diseño y modelado de la base de datos.....	56
<b>4. Caso práctico: Implementación de un Data mart de datos operativos del FONAES para medir avances físicos y financieros</b> .....	66
4.1. Proceso ETL.....	67
4.2. Explotación de la información.....	72
4.2.1. Reporte de resultados por entidad Federativa e Instrumento.....	72
4.2.2. Reporte de metas y avance financiero.....	81
<b>Conclusiones</b> .....	85
<b>Referencias bibliográficas</b> .....	87

## Introducción

A través de los años en que he trabajado con sistemas informáticos, me he dado cuenta de toda la información que es desaprovechada en la mayoría de ellos; los datos que son generados por los sistemas operacionales<sup>1</sup> pocas veces son usados para más que simples reportes informativos. Esto, significa perder oportunidades de mejora y de buenos negocios que podrían hacerse con un manejo inteligente de esta información en las empresas mexicanas. El darle forma a esta información requiere de ciertas habilidades de diseño y abstracción, además de metodologías diferentes a las usadas en un sistema informático operacional. Business Intelligence (BI) o Inteligencia de Negocios por su traducción al español, es el área de la informática que se encarga del manejo inteligente de la información que es recolectada y procesada por los sistemas operacionales en una empresa. Los almacenes de información llamados en inglés Data Warehouse o Data mart, son un concepto fundamental de esta rama de la informática. Como su nombre lo indica, un almacén de datos es un repositorio que sirve como fuente principal para el análisis de la información. Su objetivo es unificar la información de diferentes fuentes de la misma empresa o externas, creando datos homogéneos que sirvan como materia prima de análisis.

El proyecto que se expone en el siguiente trabajo es un almacén de datos de la parte operativa de la empresa, que surgió como una solución práctica a problemas de extracción de la información que se presentaban de manera recurrente. Antes de describir el proyecto daré una breve descripción de las actividades que realiza la Institución gubernamental para la cual fue desarrollado, y la problemática que se presentó y que dio origen al proyecto. El FONAES (Fondo Nacional de Apoyo a Empresas Sociales) es un órgano desconcentrado de la Secretaría de Economía que se dedica a dar créditos sociales a sectores de la población mexicana que tienen un alto nivel de marginación como son campesinos, artesanos, personas con alguna discapacidad, etc. y que no son

---

<sup>1</sup> **Sistemas operacionales.**- Sistemas que sirven específicamente para el negocio operacional de la empresa.

sujetos de crédito para la banca privada o los diferentes créditos ofrecidos por otros órganos gubernamentales. Debido al carácter estratégico de la Institución, es necesario generar constantemente informes que reflejen los avances de la institución en materia de inversión y de acciones generadas mensualmente. Estos reportes gerenciales, le son solicitados actualmente al FONAES por diferentes áreas internas y principalmente por otras instituciones gubernamentales, además de la Presidencia de la República. El FONAES cuenta con un sistema operacional en línea, en el cual se puede acceder a través de Internet o de la Intranet Institucional, pero por ahora no cuenta con toda la información necesaria para generar estos reportes y por lo mismo las áreas involucradas obtienen los reportes en forma manual, sacando la información parcialmente del sistema y completándola con los datos proporcionados por las diferentes Representaciones Estatales en el país.

La solución propuesta fue la creación de un Datamart <sup>2</sup> con información operacional y la creación de un módulo de captura que alimentará a la base de datos del sistema operacional de la Institución con las metas físicas y financieras que le son establecidas a la Institución.

El presente trabajo se concentra básicamente en el Data mart desarrollado, sus diferentes fases y la explotación del mismo. En el primer capítulo se da un marco teórico de las necesidades de carácter informático que dieron origen al concepto de Data warehouse<sup>3</sup>, su definición, características fundamentales, comparación con una base de datos operacional, concepto de un Datamart, procesos de la implementación de un ambiente Data warehouse, una descripción de las herramientas OLAP que nos servirán como apoyo para la explotación de la información y en este caso de Analysis Services que será la herramienta OLAP utilizada para este proyecto.

---

<sup>2</sup> **Datamart.-** Almacén de datos empresariales específicos de un área o departamento de la Organización, en el que el análisis se limita al área para la que es construido.

<sup>3</sup> **Data warehouse.-** Almacén de datos empresariales de toda una Organización cuyo objetivo es el análisis de esta información con fines estratégicos.

En el segundo capítulo se describirá brevemente la metodología de Análisis y Diseño utilizada para la implementación de un Data Warehouse con ejemplos claros y sencillos, que ayuden a entender poco a poco los pasos necesarios para la implementación que se seguirá más adelante en el caso práctico del Datamart.

En el tercer capítulo se hará el planteamiento detallado del problema en el FONAES, donde se analizarán los requerimientos específicos, se diseñará el DataMart y se definirá el proceso de carga de la información (ETL)<sup>4</sup>.

Continuando con la implementación piloto y pruebas del diseño de la Base de datos, en el cuarto capítulo en un ambiente de desarrollo, que servirá como ejemplo práctico de explotación de los datos cargados en el Datamart.

En el último capítulo se describirán las conclusiones y resultados del caso práctico implementado, y la parte de trabajo a futuro para terminar con la implementación en ambiente de producción.

---

<sup>4</sup> **ETL.-** Proceso de extracción, transformación y carga de información. (Extract/Transform/Load por sus siglas en inglés)

## 1. Data warehouse (Almacén de datos)

En el momento en que las grandes empresas necesitaron usar la información que generaban en sus bases de datos operacionales para *planear, evaluar o tomar decisiones* importantes, es que aparece el concepto de Data warehouse (Almacén de datos). Para realizar estas tareas era necesario contar con hechos y cifras que sirvieran como base confiable que dieran un panorama real del estado y funcionamiento de la empresa. Algunos de los problemas a los que se enfrentaron las organizaciones y que dieron origen a la creación de este concepto se describen a continuación:



### Información dispersa



En la mayoría de las empresas la información se encuentra dispersa en varios sistemas operacionales, departamentos, bases de datos o archivos (por mencionar sólo algunas fuentes). Esto sin tomar en cuenta a las grandes empresas que cuentan con varias sucursales en diferentes países o localidades.

### Pérdida de tiempo en la recopilación de los datos

Esto genera graves problemas como la inversión de tiempo para recopilar los datos necesarios de los analistas de sistemas y todos los departamentos involucrados, tiempo que en la mayoría de los casos se tiene que volver a emplear las veces que sea requerida la información.



### Entorpecimiento de los sistemas operacionales

También se tiene la dificultad de tener que usar los datos que se encuentran en las bases de datos operativas y debido a la gran cantidad de registros que se tienen que procesar para obtener las sumas, acumulados y totales requeridos, el desempeño

de estos sistemas es entorpecido por los procesos de los informes ejecutivos, generando así, retardos en las respuestas y conflictos en las actualizaciones de las bases de datos. Esto, debido a que las bases de datos operativas no están diseñadas para el análisis de datos.

### **Inconsistencia de la información**

Debido a que cada fuente de datos maneja diferentes conceptos, tipos de datos y bases de datos, la estandarización de la información no siempre es exitosa, y los reportes pueden contener datos no muy confiables.

Debido a todos estos problemas se llegó a la conclusión de que es mejor *separar la base de datos operacional de la que es utilizada para la generación exclusiva de reportes y toma de decisiones*, ya que de esta manera no se afecta el desempeño de los sistemas operacionales y se puede tener información histórica según se requiera, además de tener la información importante centralizada.

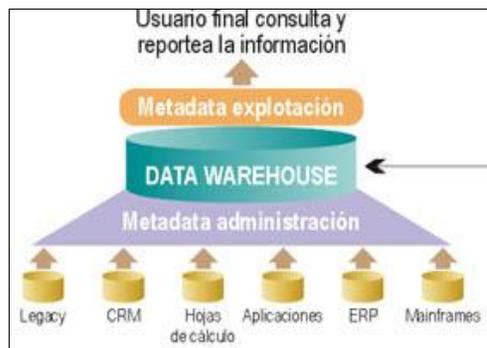
#### **1.1 Definición de Data warehouse**

Un Data Warehouse podría definirse como un conjunto de procesos y acciones, es una colección de datos orientados a un tema, integrados y no volátiles en el soporte al proceso de toma de decisiones. Un Data Warehouse es construido y diseñado expresamente para almacenar información de tipo ejecutivo.

El tener un almacén de datos centralizado que contenga la información necesaria de diversas fuentes para los reportes gerenciales, acelera el proceso de acceso y su análisis. Hace que sea más práctica y sencilla la explotación de los datos para tomar - de la mejor manera- las decisiones estratégicas de la empresa.

La construcción de un Data Warehouse se basa en la integración de datos desde una o varias fuentes de datos, la mayoría de las veces, bases de datos operacionales heterogéneas. El objetivo principal es satisfacer los requerimientos de información de la empresa para una mejor gestión, con eficiencia y facilidad de acceso.

En la figura 1.1 se muestra un ejemplo de un ambiente de Data warehouse en donde varios sistemas y fuentes de información alimentan al Data warehouse empresarial.



**Figura 1.1 Ejemplo de un ambiente Data warehouse**

Gracias a su arquitectura, un Data warehouse permite independencia de la fuente de datos. No importa, entonces, que la fuente de datos sea modificada, ya que los datos históricos, las aplicaciones de análisis y el modelo de datos no cambian.

La información pasa por un proceso de extracción, de eliminación de inconsistencias y de resumen, antes de ser cargados al almacén de datos. De esta manera elimina la información no necesaria conteniendo sólo la de carácter estratégico para la empresa u organización.

Existen muchas definiciones para el Data warehouse, la más conocida fue propuesta por Inmon <sup>1</sup> en 1992: *"Un Data warehouse es una colección de datos orientados a temas, integrados, no-volátiles y variante en el tiempo, organizados para soportar necesidades empresariales"*.

---

<sup>1</sup> Inmon, William H. (1996). - **Building the Data Warehouse**, Wiley Computer, New York.

## 1.2 Propósito y naturaleza de un Data Warehouse

El propósito de un Data warehouse, como se dijo anteriormente, es servir para la toma de decisiones de los ejecutivos principales de una empresa. A continuación se detallan las ventajas más importantes:

- Los procesos de toma de decisiones pueden ser mejorados, ya que se tiene una mayor disponibilidad de la información. Las decisiones se toman más rápido al contar con la información adecuada.
- Los procesos empresariales pueden ser optimizados, ya que un Data Warehouse puede servir para medir los niveles de desempeño de la empresa.
- Las conexiones y dependencias entre procesos empresariales se vuelven más claras y entendibles. Secuencias de procesos empresariales pueden ser optimizadas para ganar eficiencia y reducir costos.
- Los datos son organizados y estructurados para tener significado empresarial. De esta manera, se pueden identificar errores o defectos de las aplicaciones actuales y corregirlas o mejorarlas.
- La información compartida conduce a un lenguaje común, conocimiento común y mejoramiento de la comunicación de la empresa. Es decir, conduce a una mejor integración de la organización en sus diferentes áreas y departamentos.

Un Data warehouse está orientado al sujeto, lo que significa que es diseñado alrededor de los *sujetos o personas más importantes* del negocio de la empresa, como puede ser el cliente o el vendedor. Las bases de datos operacionales, por otra parte, están orientadas a los procesos basados en la funcionalidad de la aplicación para la cual fueron diseñadas. Por ejemplo, un sistema operacional podría estar diseñado basado en los procesos como ahorro, créditos y otras transacciones. Mientras que un Data warehouse podría estar organizado alrededor de sujetos como los datos del cliente.

### 1.3 Comparación entre un Data warehouse y una Base de Datos operacional

Un Data warehouse difiere de una base de datos operacional de muchas maneras. Una de las diferencias principales entre estos dos tipos de sistemas son los datos almacenados en cada uno de ellos. En los sistemas operacionales (sistemas OLTP), la información es manejada y transformada constantemente, mientras la información almacenada en el Data warehouse es para la toma de decisiones y es relativamente estática. En la tabla 1.1 se muestran las principales diferencias entre estos dos tipos de sistemas:

Sistemas operacionales	Sistemas Datawarehouse
Está orientado esencialmente a las transacciones operativas.	Está orientado al análisis de la información.
Están orientados a la aplicación, es decir, que los datos sirven para un proceso en particular o alguna funcionalidad del negocio	Están orientados al sujeto, los datos sirven para determinado sujeto del negocio como la información del cliente
Los datos se encuentran detallados	Datos resumidos o refinados con cálculos complejos en el tiempo. Por ejemplo, un sumario de las transacciones bancarias por día, semana o mes.
La estructura de la base de datos es usualmente estática	La estructura es dinámica, ya que se pueden crear nuevas estructuras según sea necesario.
La entrada de datos apunta a las personas que usarán el sistema y las ayudará sólo en sus labores operacionales.	La información está estructurada a manera de que se pueda analizar en diferentes niveles y en diferentes áreas de negocio de la empresa.
La información es volátil, cambiante.	La información es no volátil, no cambia una vez que ha sido cargada
Los requerimientos son conocidos antes de diseñar el sistema	Los requerimientos no son totalmente conocidos antes de diseñar el sistema
Sigue el clásico desarrollo de ciclo de vida, dónde una iteración del diseño es completada a través de normalización de datos y revisión de los requerimientos del usuario.	Tiene un ciclo de desarrollo totalmente distinto, donde se toma la estructura de datos de una aplicación existente y se hace el diseño para el análisis de los datos. Menos ciclos se involucran a partir de que el usuario define el sistema.
El desempeño es importante debido a la gran cantidad de usuarios concurrentes que accedan a los datos.	El desempeño no es tan importante, ya que no se esperan muchos usuarios que accederán a la información al mismo tiempo.

Tabla 1.1

## 1.4 Características de un DATA WAREHOUSE

### 1.4.1 Orientado al tema

Una de las principales características de un Data warehouse es que se orienta a determinado tema del negocio del cual se espera conocer más, en contraste con un sistema operacional que se basa fundamentalmente en aplicaciones.

Un ambiente operacional se diseña alrededor de las aplicaciones y funciones para las que fue creado.

Un Data warehouse se diseña en base a sujetos o actores principales del negocio de la empresa, como puede ser: vendedor, cliente, acreedor o proveedor para una industria de comercialización; doctor, enfermera o paciente para un hospital; alumno, maestro o clase, para una escuela.

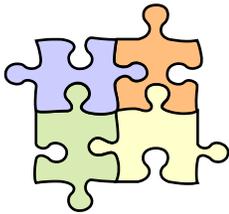


Los datos de un sistema operacional al ser orientados a la aplicación, contiene todo el detalle necesario para las funciones y procesos para los cuales fue definido. Un Data warehouse excluye toda la información que no es relevante para la toma de decisiones.

### 1.4.2 Integrado

El aspecto más importante de un Data warehouse es que su información siempre esté integrada. Las convenciones de los nombres son consistentes, medidas de variables uniforme y atributos de los datos de los campos consistentes a pesar de que existan fuentes múltiples de donde se obtengan los datos.

La mayoría de las veces, las aplicaciones operacionales muestran cómo cada diseñador ha estructurado cada una de las bases de datos, mostrando su estilo personal. La mayor parte de las empresas, en lugar de desarrollar su propio software lo compran para reducir costos y tiempo, esto dificulta la integración de sus bases de datos operacionales, teniendo entonces un ambiente heterogéneo de información.



Diferentes bases de datos tienen a veces el mismo concepto de dato diseñado de manera muy distinta, ya sea en nombre o en tipo de campo. Es aquí donde se define de qué manera se irán integrando las diferentes fuentes de datos.

### **1.4.3 De tiempo variante**

Toda la información almacenada en un Data warehouse es solicitada alguna vez. Esta es una de las características básicas de un almacén de datos. En un sistema operacional, por el contrario, la información debe estar actualizada y disponible en el instante en que se acceda.

Debido a que la información de un almacén de datos es solicitada en cualquier momento y no es necesario que contenga los datos de ese “momento” en particular, los datos son entonces llamados de “tiempo variante”.

En la mayoría de los casos los datos históricos no son muy usados en un ambiente operacional. Por el contrario, en el caso de un Data Warehouse es fundamental este tipo de información para identificar patrones o tendencias de la organización, que sirvan a los ejecutivos para evaluar o tomar decisiones en la empresa.

Datos de tiempo variante	
Operacional	Data Warehouse
El horizonte de tiempo de este tipo de sistemas tiene un promedio 60-90 días	El horizonte de tiempo es de 5 a 10 años
La clave de las tablas puede o no tener un elemento de tiempo	Cada estructura clave debe de tener un elemento de tiempo
Los datos, pueden ser eliminados, agregados, consultados o actualizados	Una vez que los datos son correctamente cargados no pueden ser alterados, sólo consultados

Tabla 1.2 Diferencias en datos de tiempo variante entre un Data warehouse y un sistema operacional

- Los sistemas operacionales tienen un horizonte mucho más corto (de sesenta a noventa días), comparado con el horizonte de tiempo de un Data Warehouse que va de cinco a diez años en promedio.
- La segunda manera en la que se muestra el *tiempo variante* en un ambiente Data Warehouse, es que en cada estructura clave debe de tener explícita o implícitamente un elemento de tiempo como: día, semana, mes, etc.
- La información en un ambiente Data Warehouse, una vez que es cargada correctamente no puede ser alterada. De manera opuesta, en un ambiente operacional la información puede ser agregada, eliminada, actualizada según sea necesario.

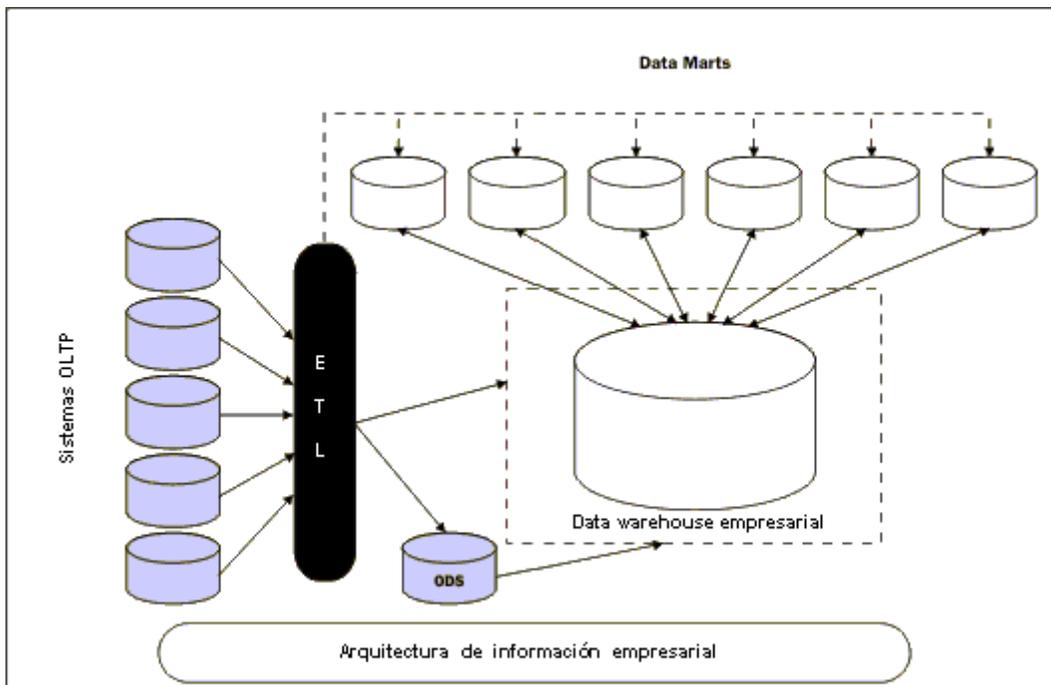
Se puede definir entonces la información contenida en un Data Warehouse como una larga serie de fotografías de los datos de la empresa.

#### 1.4.4 No volátil

La no volatilidad en un ambiente de Data Warehouse se refiere a la estabilidad de la información que contiene. En un sistema operacional los datos son constantemente alterados por los usuarios y sistemas que interactúan con él.

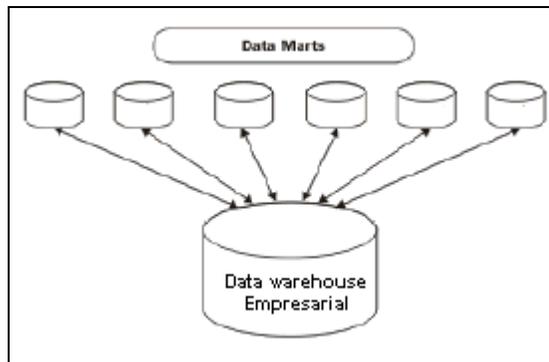
## 1.5 Data warehouse y Datamart

El ambiente de Data warehouse está creciendo en una arquitectura separada de los sistemas operacionales o transaccionales. Este tipo de ambiente cuenta con diferentes componentes que sirven para diferentes propósitos y tipos de negocio. La arquitectura que ha crecido desde los inicios del Data warehouse es ahora algo que podría llamarse *Arquitectura de Información Empresarial*, como se muestra en la figura 1.2. En esta arquitectura, los sistemas operacionales son usualmente sistemas de procesos transaccionales en línea (OLTP) que sirven a aplicaciones de usuarios finales.



**Figura 1.2 Arquitectura de Información Empresarial**

Los Data marts están orientados a las necesidades y requerimientos específicos, como servir a las necesidades de un departamento en particular, o quizás como una parte de un Data warehouse por un período de tiempo específico.



**Figura 1.3 Data marts y Data warehouse**

Se podría decir entonces que un Data mart es una estructura que reensambla un Data warehouse y está diseñado para ayudar con el soporte a la toma de decisiones, pero contiene un subconjunto de los datos del Data warehouse, ya que en éste se encuentran datos de requerimientos específicos.

Una de las ventajas de un Data mart es su movilidad, por ejemplo un vendedor podría cargar en su Laptop sólo la información necesaria para sus viajes de ventas. El Data mart puede contener sólo la información que le sea útil para su estrategia de ventas, como un viaje a determinado país, se podría cargar la información de los clientes más importantes de ese país y las ventas más importantes a éstos en los últimos cinco años.

Existen dos diferencias fundamentales entre un Data mart y un Data warehouse. La primera es que un Data mart es un subconjunto de un Data warehouse. La información en un Data mart está dirigida a requerimientos específicos y éste, está diseñado para satisfacer esos requerimientos, como datos departamentales, datos de grupos de clientes, o de un determinado período de tiempo, etc. Un Data warehouse por el otro lado es un repositorio empresarial sin ningún requerimiento específico.

El Data warehouse está diseñado como parte de una solución estratégica de planeación empresarial con todas las áreas, un data mart está diseñado como parte de una planeación táctica para satisfacer un requerimiento en particular de una o algunas áreas. Esto no significa, que una empresa pueda crear data marts aleatoriamente sin una estrategia empresarial. Si esto llegara a pasar, el resultado serían data mart

desconectados y sería muy difícil su integración dentro del Data warehouse de la empresa. Sin embargo, el diseño de una estrategia de Data warehouse puede incluir criterios y estándares que permitan una integración posterior de los Data mart creados.

La otra diferencia entre un Data mart y un Data warehouse es que la información en un Data mart se encuentra más granulada que en un Data warehouse. A partir de que los requerimientos del Data mart están mejor definidos, se pueden definir mejor las consultas que serán requeridas y guardarlas para cuando sean requeridas, de esta manera la extracción de la información será mas rápida y eficiente.

### **Arquitecturas de solución o implementación**

Los expertos en Data warehouse y Data marts desarrollaron tres principales soluciones para la construcción de Data marts. Cada una de estas soluciones tiene sus ventajas y desventajas. Algunos, sugieren tener un data warehouse empresarial que alimente a un conjunto de data marts dependientes acorde a sus requerimientos específicos. Otros, sugieren primero construir data marts y después integrarlos en el data warehouse empresarial, inclusive en un ambiente data warehouse "distribuido". Finalmente, otros expertos toman una solución híbrida, donde data marts independientes puedan ser construidos a la par de data marts dependientes que se alimenten de la información del data warehouse empresarial.

### **Arquitectura de "arriba a abajo" o "central"**

Esta solución implica construir un data warehouse y tenerlo para alimentar a un subconjunto de Data mart dependientes. Por ejemplo, datos relacionados con recursos humanos pueden ser extraídos del data warehouse a uno de los data marts, y datos pertenecientes a los clientes de Sudamérica en otro. Así que, acorde a esta solución, el flujo de datos va de los sistemas operacionales al área donde se transforma la información. Esta información pasa luego al data warehouse, de donde se alimentan los data marts dependientes. Algunos de sus beneficios son:

- Emplea una rigurosa metodología para modelar, obtener e implementar los requerimientos de procesos de decisión de los usuarios finales.

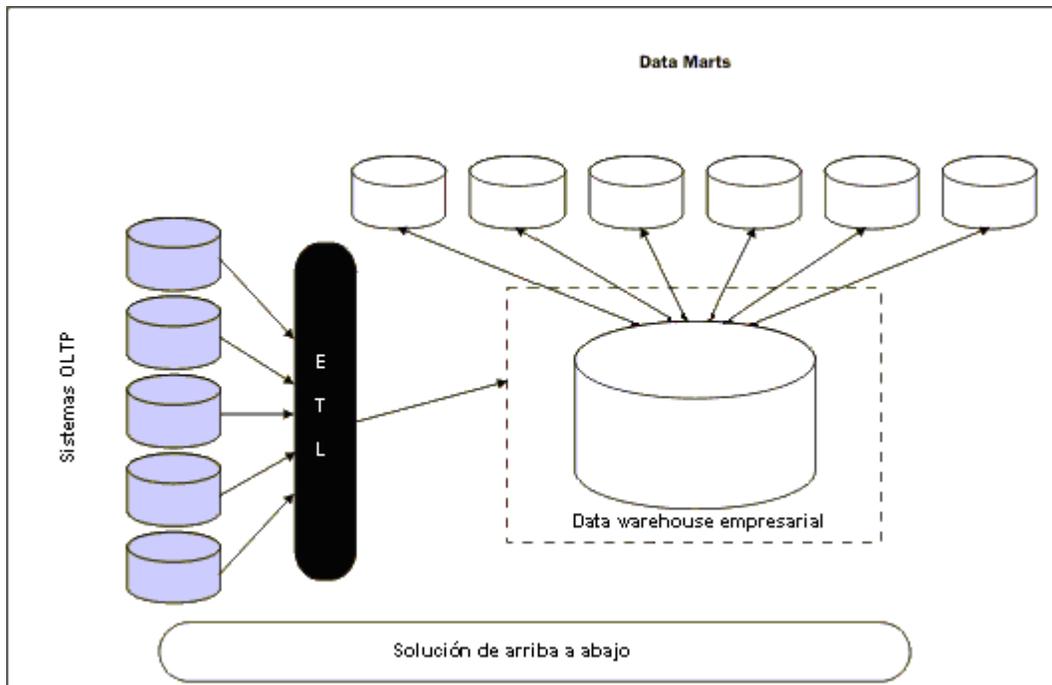
- Crea un sistema de Data warehouse que da a los usuarios finales la capacidad de tener una perspectiva de las operaciones de la empresa en su totalidad, temas, y oportunidades potenciales para el desarrollo del negocio.

- Está basado en el modelado orientado al sujeto, que minimiza los problemas de integración entre los proyectos de data warehouse. Principalmente porque todos los data marts seguirán los mismos procesos de diseño e implementación.

Pero además de tener estas ventajas, tiene también varias desventajas importantes que podrían influir en decidir o no construir bajo esta arquitectura.

- Este tipo de solución regularmente requiere de tiempos muy largos de desarrollo, costos también muy altos, además de una pobre funcionalidad final, ya que depende del Data warehouse empresarial.

- En un Data warehouse empresarial con sus largos ciclos de liberación, no es posible liberar soluciones rápidas y que satisfagan completamente las necesidades tan cambiantes en las organizaciones actuales.

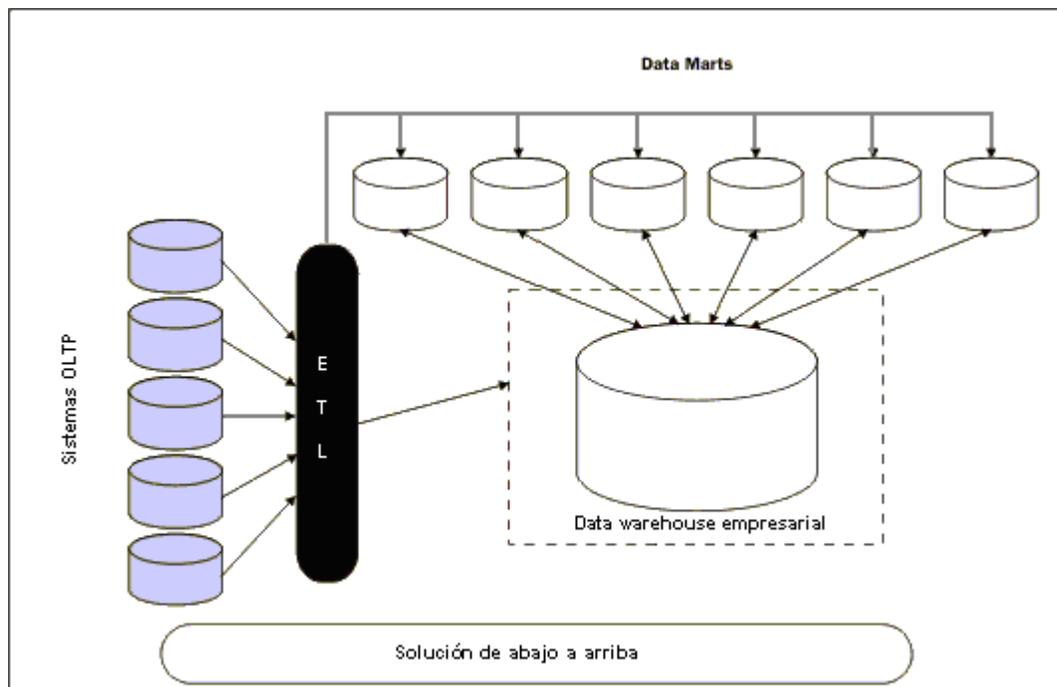


**Figura 1.4 Solución de arriba a abajo**

### **Solución de “abajo a arriba” o “distribuida”**

Esta solución consiste en construir los data marts primero y luego agregar la información de éstos en el Data warehouse empresarial. Se sugiere que no se tenga un Data warehouse centralizado, sino que sea un Data warehouse distribuido que se componga de data marts independientes que se conecten por medio de algún software, middleware o herramientas para soporte de decisiones que hagan una vista consistente de todos estos data marts.

Según esta solución, los datos fluyen directamente de los sistemas operacionales al área de extracción, transformación y carga, y de ahí directamente a los data marts independientes. Estos Data marts alimentan el Data warehouse empresarial, como se muestra en la figura 1.4.



**Figura 1.5 Solución de abajo a arriba o distribuida**

Las desventajas de este tipo de solución son:

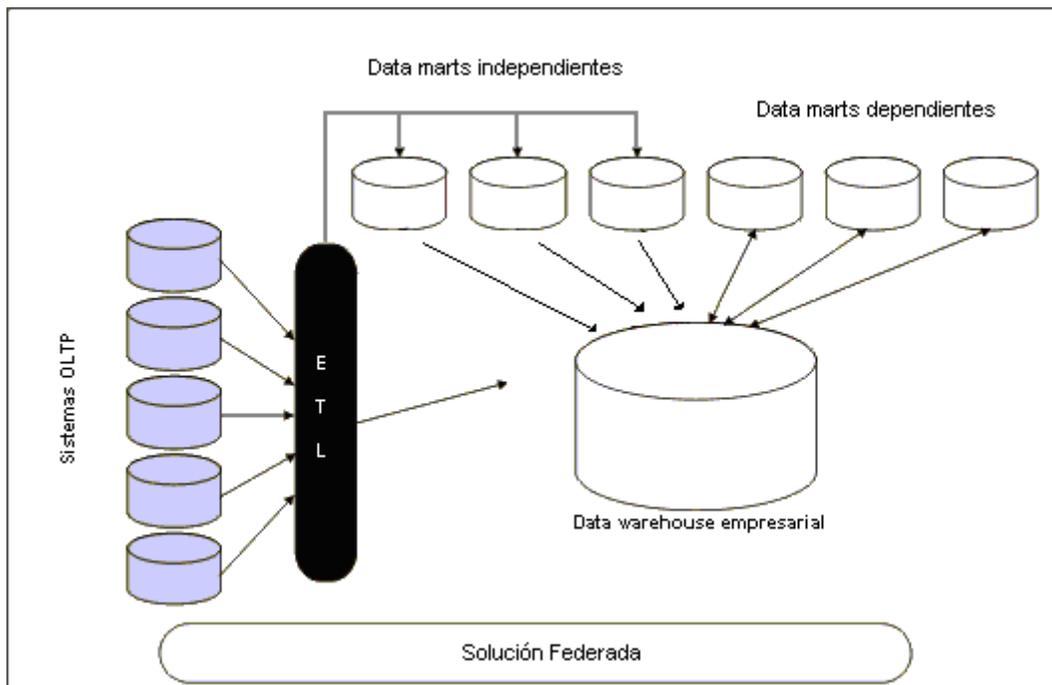
- Una incontrolable construcción y proliferación de data marts independientes podría resultar en problemas de integración entre estos data marts y el data warehouse empresarial. De hecho, el desarrollo del data warehouse empresarial no es posible sin cambios mayores del diseño de los data marts. Estos problemas pueden resolverse usando un modelo de información de Data warehouse común en el desarrollo de los data marts independientes.
- Debido a la proliferación de Data marts, los usuarios de negocio querrán acceder a los data marts pertenecientes a otros departamentos. Esto traería como consecuencia un bajo rendimiento al querer conectar varios data marts en una misma consulta.
- La necesidad de desarrollar los data marts tan pronto como sea posible, frecuentemente lleva a basar los data marts en los datos de los sistemas operacionales y no en las necesidades de negocio del usuario.

### **Solución “federada” o de “dos niveles”**

Las soluciones anteriores tienen sus fortalezas y debilidades. Una buena solución sería entonces una que tomara ambas arquitecturas, tomando lo mejor de ellas. A este tipo de solución se le llama federada.

Con esta solución, el desarrollo del sistema Data warehouse se convierte en un proceso iterativo. El sistema de Data warehouse en este caso contiene Data marts independientes que obtienen su información de los sistemas operacionales. Este sistema debe de incluir también un Data warehouse empresarial que alimenta a Data marts dependientes, como se puede ver en la figura 1.6.

La clave de una buena integración en un Data warehouse de este tipo es un modelo de información común. Este modelo de información se utiliza para construir los nuevos Data marts independientes o cuando nuevos Data marts dependientes son agregados al Data warehouse.



**Figura 1.6 Solución Federada**

Los Data warehouses o Data marts pueden estar dirigidos por dos factores básicos: necesidades de datos operacionales y requerimientos de negocio para soporte de decisiones. De esta manera, es que se puede tomar la decisión de la mejor opción de arquitectura que cumpla no sólo con las necesidades actuales de información de la empresa, sino también para un crecimiento futuro.

## 1.6 Herramientas OLAP

OLAP significa procesamiento analítico en línea (On-Line Analytical Processing), una herramienta OLAP sirve para explotar la información de las estructuras definidas en una base de datos, ya sea transaccional o Data warehouse.

Debido a su gran poder de análisis de información y flexibilidad, las herramientas OLPA son usadas comúnmente para explotar los datos en almacenes de datos. Éstas, permiten un mejor manejo de la información debido a que pueden manejar grandes cantidades de datos y tienen respuestas rápidas en procesamientos complejos.

Estas herramientas tienen la capacidad de realizar operaciones como vistas de resumen a detalle (drill down), de detalle a resumen (drill up), filtrar y rotar (Slice and Dice), etc. Generan también distintas vistas con agregaciones que permiten tener un mejor análisis de los datos. Algunos ejemplos de herramientas OLAP podrían ser Microsoft Analysis Services ® , Bussines Object ®, Essbase ®.

Las soluciones OLAP pueden mejorar la productividad de la organización entera enfocándose en lo esencial para su crecimiento y por transferir la responsabilidad para el análisis a las partes operacionales de la organización.

Este tipo de herramientas se podrían dividir en cuatro tipos esenciales:

- Herramientas ROLAP (Relational On-line Analytical Process). Son herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos SQL ordinarias, es decir, relacionales. Estas herramientas simulan los datos multidimensionales.
- Herramientas MOLAP (Multidimensional On-line Analytical Process). Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arrays de varias dimensiones, llamados cubos. La

respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales.

- Herramientas HOLAP (Hybrid On-line Analytical Process). Nos permiten un análisis híbrido de la información, es decir, que pueden obtener la información de bases de datos relacionales o multidimensionales.

- Herramientas DOLAP ( Desktop OLAP). Estas herramientas nos permiten visualizar la información de los cubos sin necesidad de tener un servidor central por el cual acceder a la base de datos, por ejemplo, la computadora portátil del director de ventas, que requiere solamente la información de alguna región en particular.

## 2. Metodología para el desarrollo de un Data warehouse

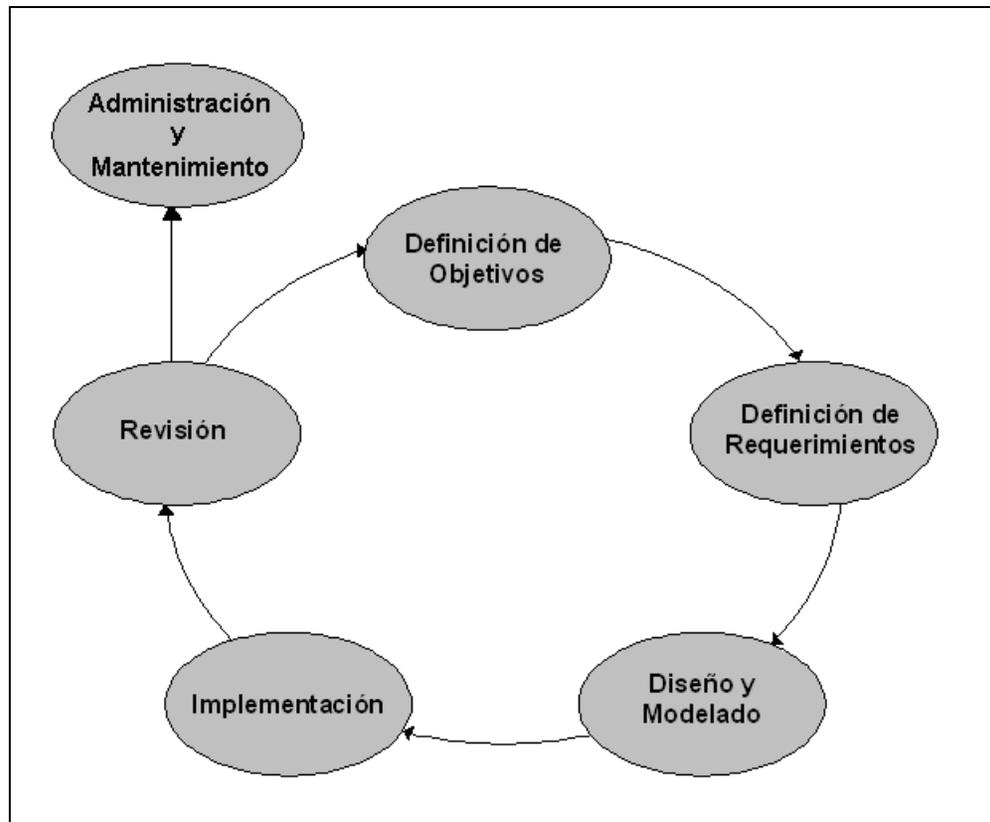
Un método de implementación iterativa para el desarrollo de aplicaciones se basa en una implicación constante de los usuarios durante todo el ciclo de desarrollo del proyecto. Una de estas metodologías, enfocada al desarrollo de Data warehouse y de las más usadas en la actualidad, es la creada por el Instituto SAS ©<sup>2</sup> llamada “Rapid Warehousing Methodology” RWM ®<sup>3</sup>. Aunque esta metodología no será descrita ni implementada totalmente en el presente proyecto, se describirán brevemente las fases que lo componen y en la que se basan la mayoría de las metodologías de desarrollo de Data warehouse.

La metodología RWM © sigue el modelo iterativo de retroalimentación de los usuarios finales que explotarán la información contenida en el Data warehouse a los desarrolladores. Esta metodología abarca las fases de Definición de objetivos (Assessment), Definición de Requerimientos (Requeriments), Diseño y modelado (Design), Implementación (Construction) , Revisión (Review) y ; Admnistración y mantenimiento.

---

<sup>2</sup> **Instituto SAS.**- Compañía inglesa dedicada al desarrollo de software, principalmente en Gestión Corporativa Inteligente (Business Intelligence).

<sup>3</sup> **Rapid Warehousing Methodology.**- Metodología rápida para la creación de Datawarehouses creada por el Instituto SAS.



**Figura 2.1 Metodología Rápida de Warehousing ®**

El ciclo iterativo entre estas fases se puede ver claramente en la figura 2.1., después de que concluye la fase de implementación sigue la fase de revisión en la que se evalúan los resultados de todas las fases anteriores y de esta manera permite que el sistema se pueda desarrollar en pequeñas fases iterativas hasta que se cumplan los objetivos del mismo. Una vez que estos objetivos o requerimientos son alcanzados, el ciclo iterativo termina, y se sigue entonces con la fase de administración y mantenimiento del Data warehouse.

### **Definición de objetivos**

La fase de definición de objetivos o evaluación preliminar es una fase crucial para el acertado desarrollo del proyecto de Data warehouse, ya que es de aquí de donde parten todas las demás fases del proyecto. En ella, se descubren e identifican los objetivos. Se define el *por qué* del Data warehouse, los beneficios esperados, la infraestructura técnica y organizativa. Se parte entonces de las

necesidades generales de la organización y la situación actual, que permitirán identificar la mejor solución para la empresa.

### **Definición de requerimientos**

Un paso fundamental para llevar a cabo cualquier proyecto, es el análisis del problema a resolver. Para adecuar la solución a las necesidades que dan origen al problema, es necesario recolectar y analizar los requerimientos.

Para el análisis de un Data warehouse o Data mart los requerimientos se pueden dividir en dos tipos fundamentales:

- De negocio
- Arquitectónicos

## Requerimientos de Negocio

Los requerimientos de negocio son los que se definen en relación a las necesidades y procesos de negocio de la empresa. Esta tarea incluye identificar los objetivos de información del Data warehouse o data mart, determinando la mejor solución y cualquier requerimiento de información especial del usuario.

Después de haber identificado las reglas de negocio más importantes de la organización, se definen las métricas que permitirán conocer o medir el estado de algún proceso del negocio. Como por ejemplo, la cantidad de facturas generadas o rechazadas, tiempo de facturación para el área de facturación. El volumen de ventas para un área de comercialización en cierto período de tiempo, o ventas en determinado lugar, etc.

A partir de la identificación de las reglas de negocio se determinan las métricas a utilizar en el Data warehouse. Una métrica o identificador es información que permite conocer el estado de algún proceso, evento o situación en particular de la empresa. Estas métricas son muy útiles para evaluar o definir el desempeño del negocio para mejorar procesos o identificar buenas prácticas y promoverlas.

Existen varios tipos de métricas que pueden ser identificadas en un proceso de negocio:

Calidad de la información: Promedios, máximos, mínimos

Métricas de desempeño: Tendencias, gráficos

Análisis de impacto: Montos, promedios, tendencias

Cada una de estas métricas evalúa distintos aspectos del negocio y la combinación de ellas puede generar información adicional, las cuales pueden dirigir a acciones de capacitación, control de calidad, mejora de procesos dependiendo de los resultados que sean obtenidos.

## Requerimientos Arquitectónicos

Los objetivos de tipo arquitectónico se refieren a las características de construcción del sistema. Un mecanismo arquitectónico es una solución a un requerimiento o un problema frecuente. Este mecanismo arquitectónico se define en distintos niveles de abstracción. De esta manera se llega a especificar de manera más concreta los productos que satisfagan esa necesidad. Por ejemplo, para el requerimiento de *persistencia* de la información de algún sistema, en este caso, el nivel de diseño sería un manejador de base de datos (relacional u orientado a objetos), ya en el siguiente nivel se define específicamente el producto que satisfaga esa necesidad, como Oracle®, Sybase® u ObjectStore®.

Nivel Análisis	Nivel Diseño	Nivel Implantación
Persistencia	RDBMS	Oracle®
		SQL Server®
	OODBMS	ObjectStore®

**Tabla 2.1 Niveles de abstracción arquitectónica**

Regularmente son suficientes tres niveles para identificar los productos que se incorporarán a nuestro mecanismo arquitectónico. En el nivel de análisis se definen las características genéricas que debe de cumplir el producto. En el nivel de diseño se identifican las herramientas que cumplen con el requerimiento de análisis. En el nivel de implantación o instrumentación se definen las marcas de productos que cumplen con las necesidades especificadas en los niveles anteriores.

Los requerimientos de análisis arquitectónicos más frecuentes se dividen en requerimientos de funcionalidad, utilización, confiabilidad, rendimiento, soporte, diseño, instrumentación, interfaz y físicos. A continuación se muestra una tabla en la que se describen algunos ejemplos de cada uno de los tipos de requerimientos antes mencionados.

Tipo	Mecanismo	Descripción
Funcionalidad	Impresión	Forma en que se le permitirá al usuario imprimir los informes del sistema.
	Presentación	Manera en la que se le permitirá al usuario acceder a la información.
Utilización	Accesibilidad	Facilidad con la que se puede acceder a las funciones del sistema
	Estética	Estilo de diseño de la interfaz de usuario
Confiabilidad	Exactitud	Exactitud y precisión con que se deben de hacer los cálculos del sistema.
	Recuperación	Capacidad de recuperación del sistema frente a fallos
Rendimiento	Tiempo de respuesta	Tiempo que tarda el sistema en responder las solicitudes.
	Tiempo de carga	Tiempo en que tarda el sistema en el proceso ETL, para actualizar los datos del data warehouse
Soporte	Sincronización	Capacidad de mantener los datos sincronizados con los sistemas operacionales
	Escalable	Facilidad en la que el sistema puede crecer.
Diseño	Impresión	Manera en la que se le permitirá al usuario emitir reportes impresos del sistema.
	Seguridad	Manera en que se permitirá el acceso discrecional a la información.

**Tabla 2.2 Tipos de mecanismos arquitectónicos**

## Diseño y modelado

Los requerimientos de negocio identificados durante la anterior fase proporcionarán las bases para realizar el diseño y el modelado del Data Warehouse o Data mart. En esta fase se identificarán las fuentes de los datos como sistemas operacionales, fuentes externas, etc. y las transformaciones que serán necesarias para su integración al Data warehouse. A partir de dichas fuentes, obtener el modelo lógico de datos del Data Warehouse.

El modelo lógico se traducirá posteriormente, en el modelo físico de datos que se almacenará en el Data Warehouse y que definirá la arquitectura de almacenamiento adaptándose al tipo de explotación que se realice del mismo.

El modelado de datos es un paso delicado en la construcción de cualquier aplicación de base de datos. Éste, usualmente describe el proceso de negocio y las

relaciones entre las diferentes entidades que interactúan en el ambiente de negocio. Para tener una idea más clara de lo que es el modelado de datos, se describirá brevemente lo que es el modelado tradicional llamado también Entidad-Relación.

## Modelado Entidad-Relación

El modelo Entidad Relación (ER) es el tipo de modelado que usualmente se implementa en las bases de datos operacionales. Es de este tipo de modelos, de donde parte el modelado dimensional.

El modelo ER es una herramienta de abstracción que puede ser usada para entender y simplificar las relaciones de datos del proceso de negocio. El modelado ER está basado en dos conceptos básicos: *entidades* y las *relaciones* entre estas entidades. Los modelos ER son usados para producir un modelo de datos para el área específica de Interés. En la figura 2.1 se muestra un ejemplo de un diagrama típico de ER.

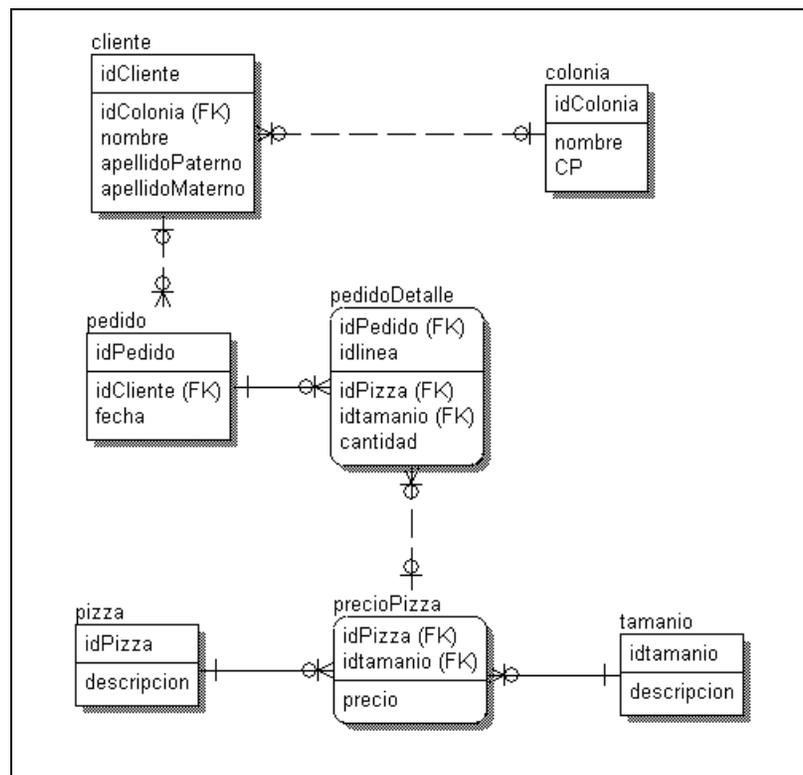


Figura 2.1 Ejemplo de diagrama Entidad-Relación

Las *entidades* representan objetos o conceptos del mundo real que pueden ser observados y clasificados por sus propiedades o características. En el ejemplo de la figura 2.1 se pueden observar las entidades principales de negocio de una pizzería: el cliente, la pizza y el pedido. Las propiedades o características de estas *entidades* son llamadas *atributos*. Estos *atributos* deben de tener un nombre único que se describa a sí mismo, para el ejemplo anterior, los *atributos* para la entidad pizza son el tamaño y el precio, en el caso de la entidad cliente sus atributos pueden ser nombre y dirección.

### **Modelado dimensional**

El modelado dimensional es una técnica para conceptualizar y visualizar modelos de datos como un conjunto de métricas que describen los aspectos más comunes del negocio. Es especialmente útil para resumir y reagrupar los datos y presentar vistas que ayude al análisis de la información. El modelado dimensional se enfoca en datos numéricos como precios, cuentas, balances, montos, ventas o número de eventos. En Data warehouse, el modelado dimensional es más simple, más expresivo y más fácil de entender que el modelado entidad-relación (ER).

Los conceptos fundamentales del modelado dimensional son los *hechos* y las *dimensiones*. Estos conceptos sirven como componentes básicos de los *cubeos*<sup>4</sup> de datos. Los cubos de información y el modelado dimensional son herramientas muy poderosas para describir los procesos de negocio del usuario desde la perspectiva de base de datos.

### **Hechos**

Una tabla de hechos es una colección de datos relacionados, con métricas e información conectada. Cada tabla de hechos representa un objeto de negocio,

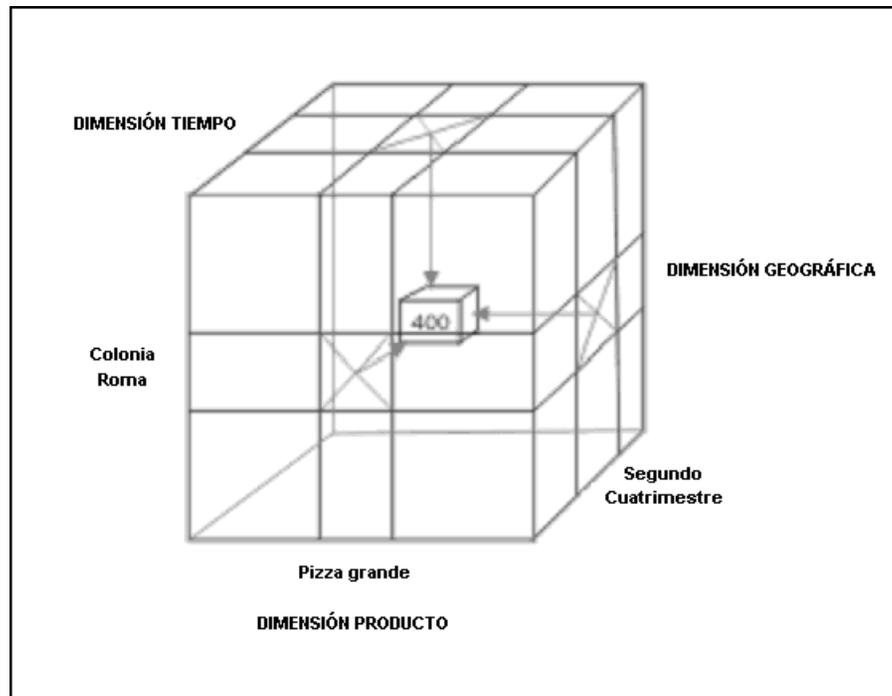
---

<sup>4</sup> **Cubo.**- Es la unidad de almacenamiento de un Data warehouse, es el equivalente a las tablas en las bases de datos relacionales. Los cubos son subconjuntos de datos de un almacén de datos, organizado y sumariado dentro de una estructura multidimensional.

como una transacción, un evento, o algún proceso que pueda ser usado para analizar el negocio. Como por ejemplo: las ventas de alguna compañía por tienda o algún período de tiempo. Este es el tipo de datos que interesan. Las tablas de hechos son las que contienen los datos numéricos a almacenar y se derivan usualmente de los sistemas operativos de la empresa. La tabla de *hechos* es la abstracción de las métricas o requerimientos de negocio. En nuestro ejemplo podrían ser las ventas por zona o colonia.

### ***Dimensiones***

Una dimensión es una colección de propiedades a través de la cual se conduce el análisis de *hechos*. Las dimensiones permiten ver los *hechos* en diferentes contextos. En este caso las ventas se podrían ver por tienda, ciudad, región o tipo de producto. Estos tres niveles son miembros de una dimensión geográfica del proceso de ventas. En un modelo dimensional cada referencia en la tabla de hechos está asociada con uno y sólo un registro de las múltiples dimensiones. Por ejemplo, en la gráfica 2.2, en la celda que tiene el valor de 400 corresponde a un conjunto único de dimensiones: segundo cuatrimestre, Pizza grande hawaiana y la Colonia Roma. Muchos procesos analíticos son usados para cuantificar el impacto de las *dimensiones* en los *hechos*, para hacer más clara la relación entre los hechos y las dimensiones. Por ejemplo, un registro en la tabla de *hechos*, éste tiene muchas ligas en las tablas de dimensiones. Estas ligas identifican al registro como único.



**Figura 2.2 Ejemplo de cubo**

Algunas de las dimensiones más comunes en ventas podrían ser: tiempo, localización geográfica, cliente y vendedor.

### **Jerarquías dimensionales**

Las jerarquías dimensionales es el ordenamiento lógico de los atributos dentro de cada una de las dimensiones definidas. Es posible que existan varias jerarquías dentro de una dimensión, pero siempre es posible identificar una jerarquía principal y jerarquías secundarias, como se puede observar en la figura 2.3, en la cual la jerarquía principal o de primer nivel es el año.

De una manera más clara, se puede definir el ordenamiento de los atributos (jerarquías) como el camino de la navegación.

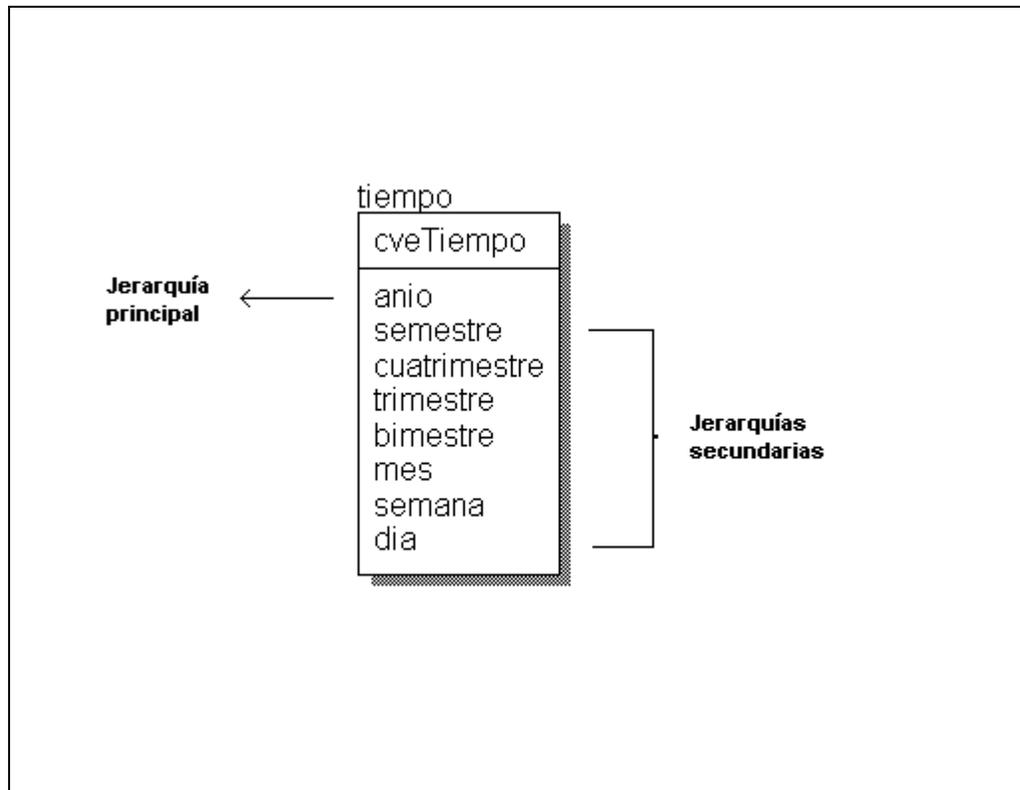
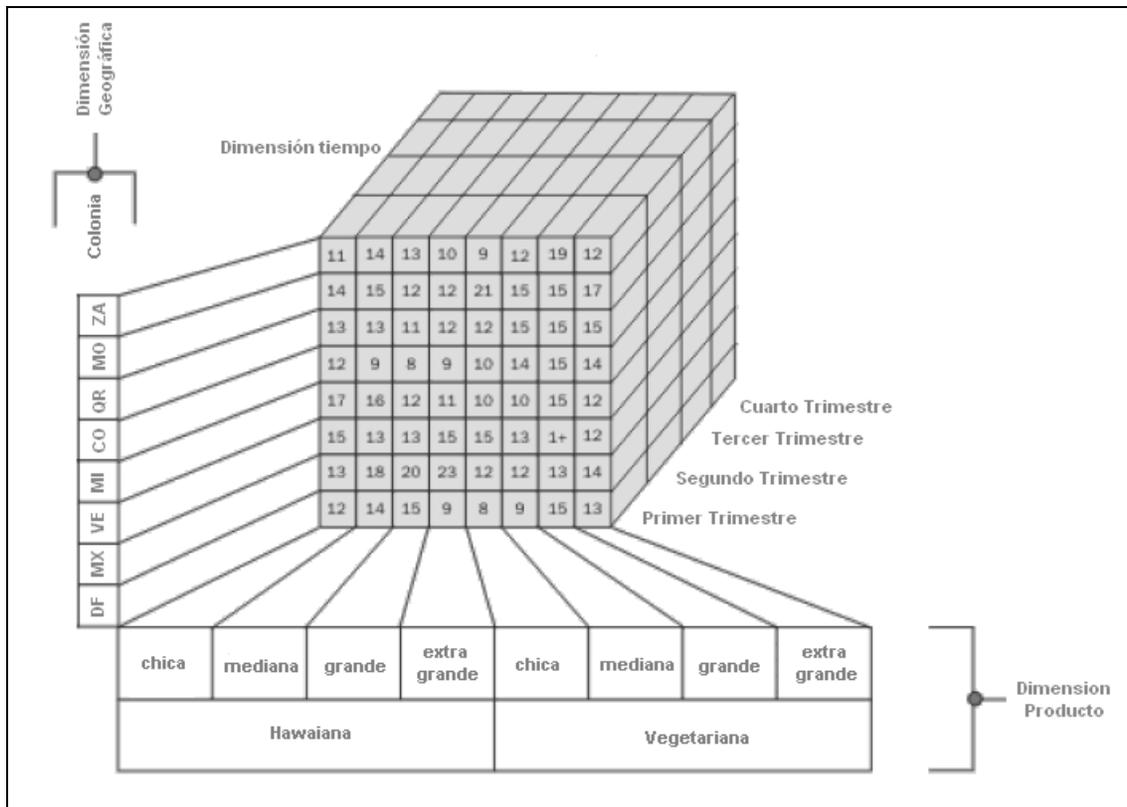


Figura 2.3 Ejemplo de Jerarquías en la dimensión tiempo

### Diseño de los cubos

Una buena manera de visualizar el modelo de datos, las jerarquías dimensionales y los hechos es construyendo un *cubo*. Las dimensiones del cubo serán las dimensiones de la tabla de hechos. Podemos representar un modelo tridimensional usando un *cubo*. Usualmente un modelo dimensional consiste en más que tres dimensiones y se refiere a éste como un *hipercubo*. Sin embargo, un hiper-cubo es imposible de ver gráficamente, es así como *cubo* es el término más usado. En la figura 2.4, la métrica usada es el volumen de ventas de una pizzería que atiende a varias colonias. Esta métrica está determinada por la combinación de tres dimensiones: geográfica, producto y tiempo. La dimensión geográfica y de producto tiene sus propios niveles jerárquicos. Por ejemplo, la dimensión geográfica tiene el nivel de colonia, mientras que la de producto los de tamaño y especialidad.



**Figura 2.4 Ejemplo de cubo en tres dimensiones**

La dimensión de tiempo muestra los niveles de trimestre. Se puede visualizar fácilmente que la especialidad que se vende más es la hawaiana y dónde más se vende es en la colonia México. De esta manera es más sencillo visualizar y analizar la información, es una excelente herramienta para observar el resumen y los totales de la información de manera más clara.

### **Diseño de base de datos**

Los Data marts son muy similares en estructura y naturaleza a un Data warehouse. La diferencia esencial entre data mart y data warehouse es que los data marts están más enfocados en un aspecto específico del negocio o alguna área de la empresa. Por ejemplo, las ventas de una tienda departamental, por diferentes períodos de tiempo o productos.

Los Data marts son usualmente mucho menos complejos y más fáciles de manejar que los Data warehouses. La cantidad de datos es menor y el modelo de datos es también más pequeño, ya que contiene menos tablas.

Modelar un Data mart debe de estar más enfocado al usuario final. Los usuarios finales deben de estar involucrados en el proceso de modelado del Data mart.

El modelado dimensional, es conveniente para el desarrollo de Data marts que soporten análisis de datos dimensional para un proceso de negocio. Dicho análisis puede incluir usar una herramienta OLAP para analizar los datos y organizarla en cubos.

Generalmente, los Data marts así como los Data warehouses, siguen esquemas específicos en su estructura, los dos tipos más importantes de esquemas son el de *estrella* y el de *copo de nieve*.

### Esquema de estrella (star schema)

Enfocándonos en el ejemplo que vimos anteriormente, podemos ver que el modelo se basa en tres dimensiones. Estas dimensiones son: localización geográfica, tiempo y producto. La tabla de hechos, que incluye las métricas, como ventas, inventario, número de clientes, etc., también incluye llaves foráneas que se relacionan con las tablas de dimensiones.

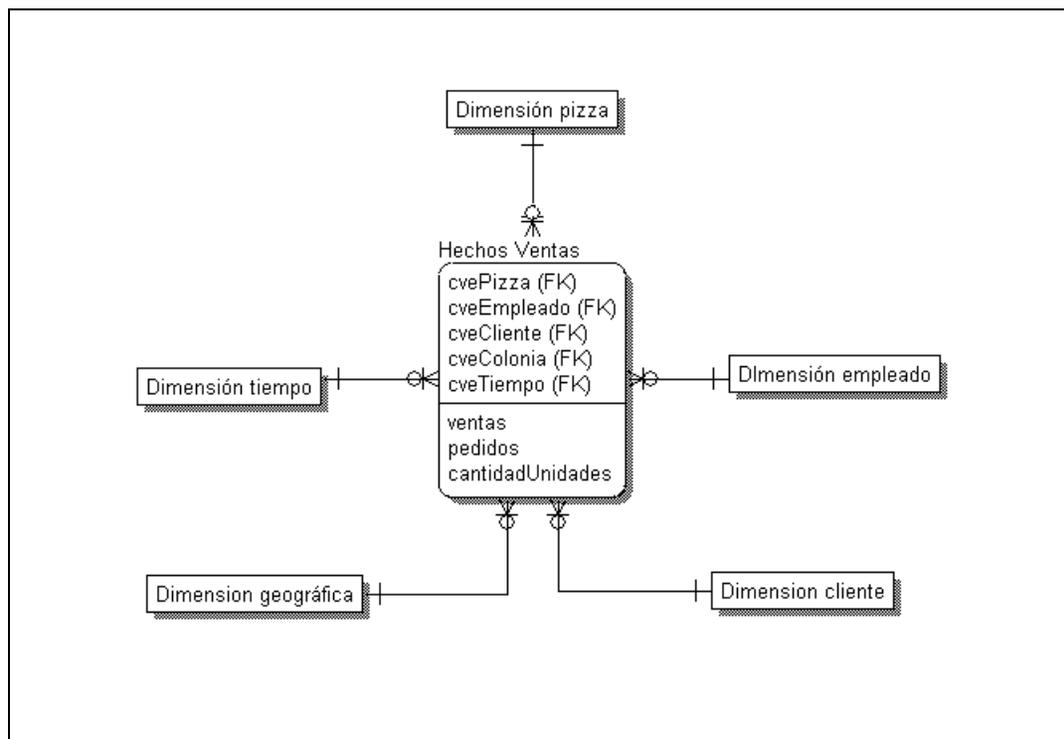


Figura 2.5 Esquema de estrella

El esquema de estrella se caracteriza por lo siguiente:

- Una sola tabla de hechos que contiene una llave primaria compuesta, con un campo para cada dimensión y columnas adicionales para los hechos numéricos.

- Una tabla de dimensión sencilla por cada dimensión con una llave generada o también llamada sintética, y un indicador de nivel que describa el nivel del atributo por cada registro. Por ejemplo, si la dimensión es pizza, los registros en la tabla deben de referirse a la ESPECIALIDAD (Vegetariana, peperoni, hawaiana), TAMAÑO (chica, mediana, grande) , etc. Debemos entonces asignar un nivel a cada atributo: especialidad =1, tamaño=2, etc. y colocarlos en la tabla de hechos en ese mismo orden.

- La tabla de hechos debe de contener información detallada, como ventas para una determinada colonia, para determinada pizza, o determinado período de tiempo.

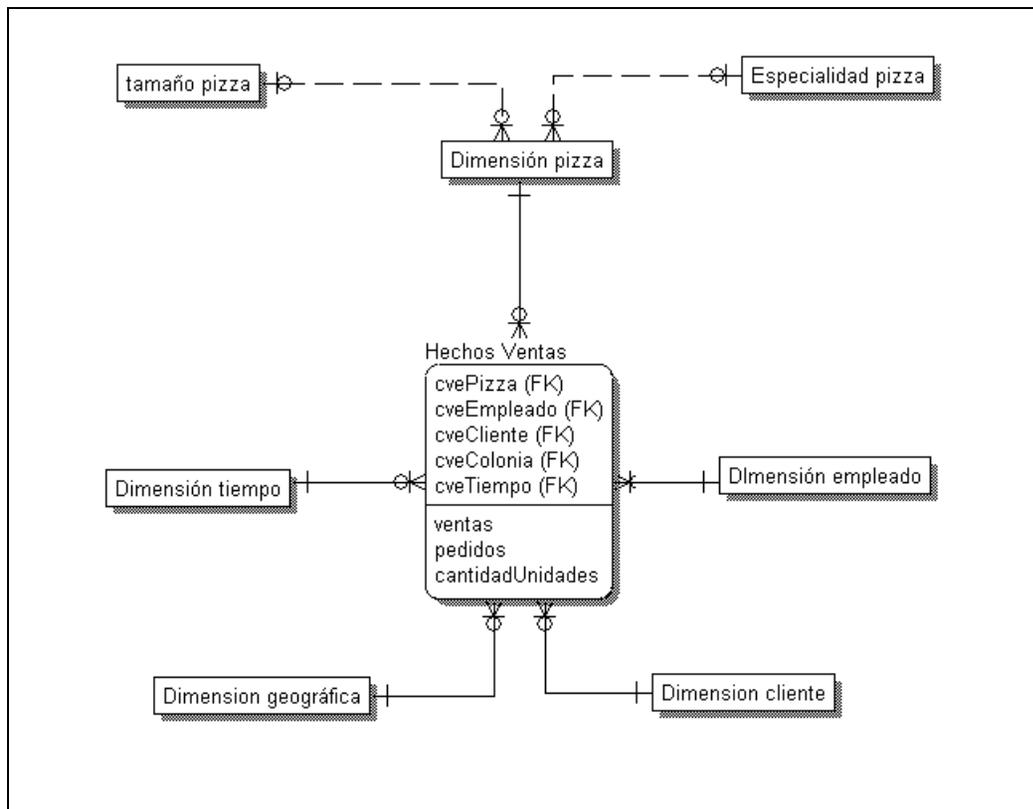
- La tabla de hechos debe de contener datos parcialmente consolidados, como ventas para una zona, para una especialidad-tamaño para determinado período de tiempo.

### **Esquema de copo de nieve (Snowflake schema)**

El esquema de estrella es la más simple representación de las entidades de negocio y su relación en el modelo dimensional. Sin embargo, algunas veces, una o más de las tablas de dimensiones podrían crecer y convertirse en una gran tabla, lo que impulsa al diseñador a utilizar el modelo de copo de nieve.

El esquema de copo de nieve es el resultado de descomponer una o más de las dimensiones, las cuales a veces contiene sus propias jerarquías. Podemos definir las relaciones de muchos-a-uno entre miembros de una tabla de dimensión como una tabla de dimensión separada con una jerarquía. Tomemos nuevamente el ejemplo de la pizzería, la dimensión de pizza está compuesta de dos sub-dimensiones: especialidad

y tamaño. En la gráfica 2.6 se puede observar la descomposición de la dimensión pizza para tener dos dimensiones más: tamaño y especialidad.

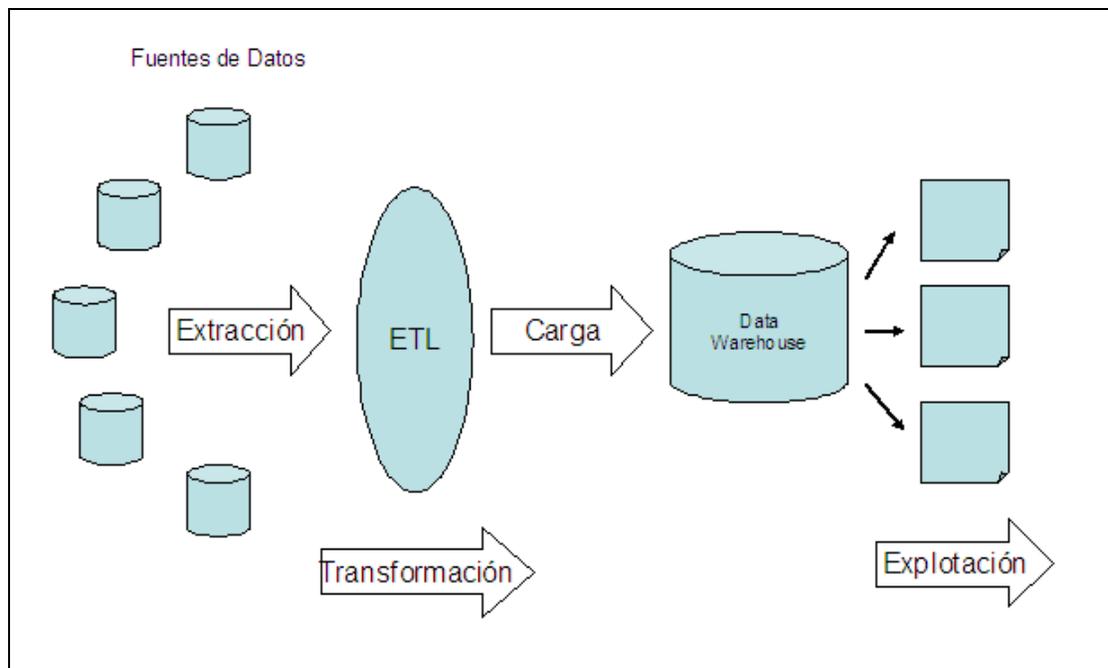


**Figura 2.6 Ejemplo de esquema de estrella**

La descomposición de la dimensión del cliente hace que el modelo parezca un copo de nieve, de ahí el nombre de este tipo de esquema.

## Implementación

Durante esta etapa del desarrollo de una Data warehouse se codifica y se carga la información en la base de datos que ha sido diseñada en la fase de modelado. Se desarrollan, también, las aplicaciones que explotarán la información para los usuarios. En la figura 2.7 se muestran los procesos principales que conforman esta fase: extracción, transformación, carga y explotación de la información. Cada uno de estos procesos está compuesto por tareas específicas, las cuales serán descritas a detalle más adelante.



**Figura 2.7 Fases de la Implementación de un Data Warehouse**

La implementación de un Data warehouse o Data mart consiste en:

- 1) Extraer la información de los sistemas operacionales y de las diversas fuentes definidas en la fase de diseño, los datos importantes que formarán parte del almacén de datos.
- 2) Transformar los datos que necesiten ser modificados.
- 3) Carga de los datos validados, que deberá planificarse según las necesidades de actualización de los usuarios.

A las tres primeras fases de la etapa de implementación se les llama proceso ETL por sus siglas en inglés: “*Extraction, Transformation and Loading*”.

4) Explotación de la información que consiste en el desarrollo o implementación de alguna aplicación que sirva como canal para la utilización práctica de los datos que contiene el Data Warehouse y que cumpla con los objetivos para la cual fue construido.

### **Proceso ETL**

Como se mencionó en el primer capítulo, la integración es la característica más importante de un Data warehouse, es por esto que la etapa de la implementación es de crucial importancia. El proceso ETL consiste en las tareas que extraerán la información de las diferentes fuentes de datos, la ajustarán a las necesidades del Data Warehouse y la cargarán al mismo. A continuación se describe cada una de estas importantes tareas de la implementación.

### **Extracción de datos**

Son los procesos encargados de obtener la información necesaria para el Data Warehouse, estos procesos pueden extraer los datos de múltiples orígenes o fuentes, por ejemplo:

- Bases de datos de sistemas operacionales
- Bases de datos externas (públicas y privadas)
- Archivos planos
- Datos en formatos tradicionales: documentos, facturas, recibos, etc.
- Internet

En este proceso se deben de extraer los datos de las diferentes fuentes de información que se hayan definido en la fase de diseño y modelado. Algunas veces, es desde aquí que se obtienen resúmenes de los datos dependiendo del diseño del Data Warehouse, y de esta manera, evitar tener información a nivel de detalle que no es relevante para la empresa u organización.

## Transformación de datos

El proceso de transformación de datos es crucial en la implementación de un Data warehouse, ya que es aquí en donde se asegura la integridad de los datos a cargar en la base de datos, que es la característica fundamental en un Data warehouse, como se mencionó en el primer capítulo.

La transformación se encarga de eliminar inconsistencias en los *formatos*, *medidas de atributos*, *convenciones de nombramiento* y *codificación* de los datos que se obtienen de las múltiples fuentes de datos.

## Codificación

Un ejemplo claro de este tipo de inconsistencia es el campo de género, en donde cada diseñador lo codifica de manera distinta: desde una “M” y “F” hasta un campo de sólo un bit donde 1 es hombre y 0 mujer. No importa entonces cómo esté definido el dato en las aplicaciones operacionales, lo importante es que al ser cargado en el Data warehouse el dato llegue en un estado uniforme. Esto se puede ver claramente en la figura 1.2.

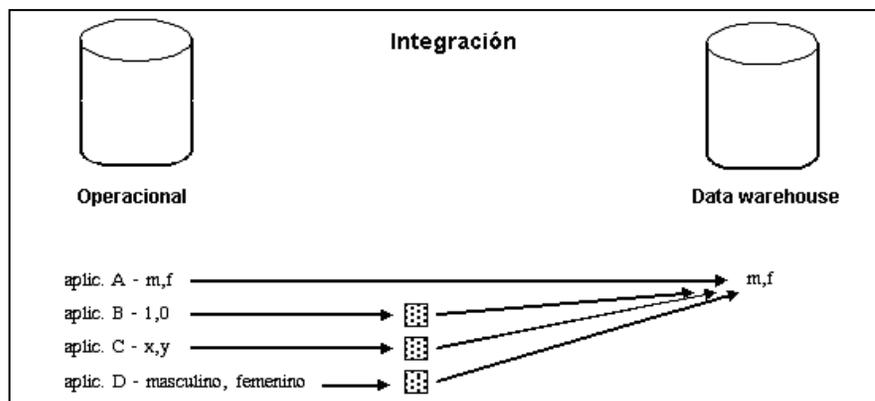


Figura 2.8 Codificación

## Medida de atributos

Otro problema importante de integración es el tipo de medida que se usa en el negocio de la empresa. En el caso, por ejemplo, de una empresa que se dedique a la renta de autobuses en distintos países, las medidas de distancia podrían ser millas o kilómetros, de esta manera se tendría entonces que establecer un estándar, que haga homologar la información en el Data warehouse.

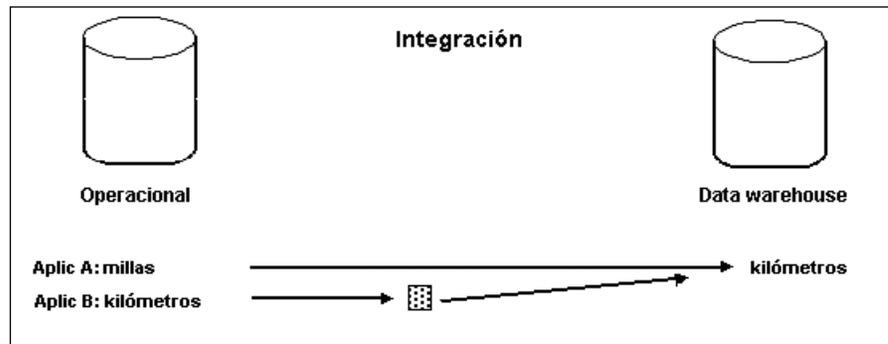


Figura 2.9 Medida de atributos

## Convenciones de nombramiento

Algunas veces existe el mismo campo en varias bases de datos operacionales, que, aunque contengan la misma información, tienen un nombre distinto. Para una adecuada integración, es necesario unificar los nombres de los campos, como se muestra en la figura 1.4.

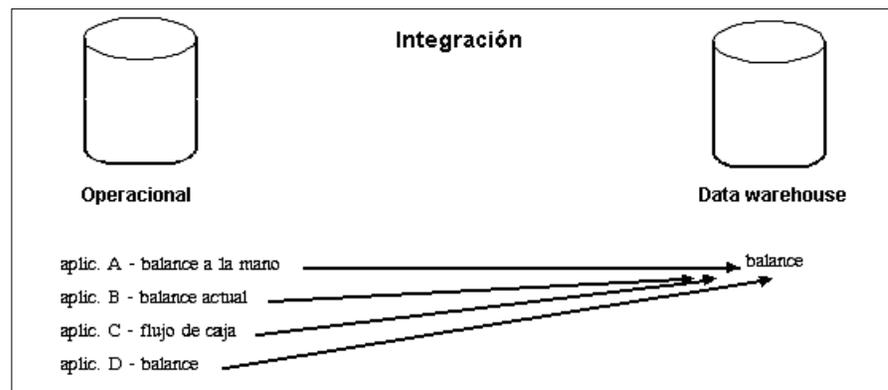


Figura 2.10 Convenciones de Nombramiento

## Fuentes múltiples

El mismo dato se encuentra a veces en distintas bases de datos, ya que no existe una relación o integración entre las mismas aplicaciones operacionales. Es, entonces, de gran importancia que el proceso de transformación y carga de datos obtenga la información de la fuente de datos más apropiada para pasarla al Data warehouse.

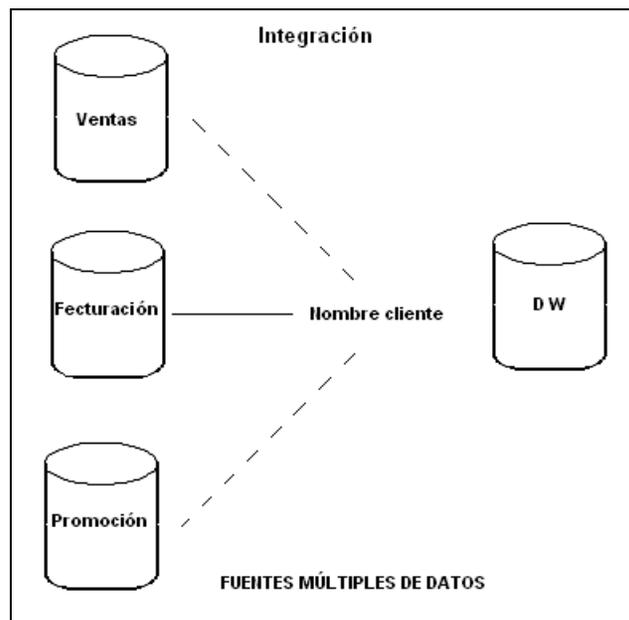


Figura 2.11 Fuentes múltiples de datos

Además de la integración de datos, estos procesos se encargan también de limpiar los datos de información no relevante, creación de llaves sintéticas<sup>5</sup>, obtener agregados y resúmenes de datos.

<sup>5</sup> Llave sintética.- Número consecutivo que elimina las llaves compuestas para ganar velocidad en las consultas, llamada también llave artificial.

## **Carga de datos**

Una vez que los procesos de transformación crearon datos consistentes, se procede a cargar la información a la base de datos del Data warehouse, y se hace la indexación de los registros. En este proceso de carga de información es en donde se realizan las pruebas de calidad que sean necesarias, así como también el manejo de errores.

Carga de los datos validados en el Data Warehouse. Esta carga deberá ser planificada con una periodicidad que se adaptará a las necesidades de actualización detectadas durante las fases de diseño.

## **Meta datos**

Una parte también muy importante de estos procesos de transformación es la creación de los meta datos de carga, que son el soporte de los procesos de almacenamiento y negocio de la empresa.

Los meta datos son la información que define los datos que alimentarán, transformarán y actualizarán el Data warehouse. Se usan para documentar el contenido del almacén de datos, manejar y cambiar procesos, mantener la seguridad y controlar estándares. Aunque no existe un modelo específico de definición de meta datos, en la mayoría de ellos, se debe de incluir lo siguiente:

- Estructura de los datos origen.
- Flujo de fuente-destino de la información.
- Especificaciones de las transformaciones de datos.
- El modelo dimensional del destino (dimensiones, relaciones y jerarquías del Data warehouse).
- Registro de actualización de nuevos datos al Data warehouse.
- Los niveles y el método de sumarización.
- Periodicidad de carga de información.

## 2.4.2 Explotación

El objetivo principal de cualquier Data warehouse es la explotación de la información que éste contiene. Para ello, se utilizan diversas técnicas dependiendo de los requerimientos del usuario y las necesidades de la empresa. A continuación se enuncian algunas de estas técnicas:

- Reportadores.- Sistemas que extraen información de bases de datos relacionales o data warehouse, la muestran generalmente en forma tabular, con funciones de concatenación, agrupamiento, sumarización, gráficas, etc.
- Herramientas OLAP (On-line analytical processing)
- Sistemas de Información ejecutiva EIS (Executive Information System).- Son aquellos programas de información que tienen herramientas asociadas que da a los directivos información estratégica y de administración.
- Sistemas de soporte de decisión DSS (Decision Support Systems).- Son sistemas automatizados que ayuda a la organización mediante un análisis estratégico de la información histórica.
- Visualización de la información en algún sistema desarrollado especialmente para ello.
- Minería de Datos (Data Mining).- Proceso de extracción de información de la base de datos, que permite identificar patrones, modelos, tendencias, relaciones, etc.



## Revisión

La construcción del Data Warehouse no finaliza con la implantación del mismo, sino que es una tarea iterativa en la que se trata de incrementar su alcance aprendiendo de las experiencias anteriores.

Identificar los aciertos y errores que pudieron existir en las fases anteriores de la metodología, serán muy útiles para continuar y mejorar el siguiente ciclo iterativo de desarrollo. La calidad de datos es un factor determinante en el éxito de un proyecto de Data warehousing. Es en esta etapa en donde deben sanearse los inconvenientes relacionados con la calidad de los datos fuente.

Después de implantarse, la metodología SAS ® sugiere que se haga una revisión del Data Warehouse planteando preguntas que permitan, después de los seis o nueve meses posteriores a su puesta en marcha, definir cuáles serían los aspectos a mejorar o potenciar en función de la utilización que se haga del nuevo sistema.

Nuevamente, después de tres o seis semanas, se realiza una revisión de la implementación para asegurarse de que la transición se ha hecho correctamente y de que los usuarios tienen acceso al Data warehouse, la cual deberá contener información consistente. No es hasta después de 18 o 24 meses de que haya concluido la construcción, que se percibirán los beneficios y se evaluará si el Datawarehouse sigue siendo vigente para los requerimientos de negocio de la empresa.



## **Administración y mantenimiento**

Después de terminar con la fase de implementación y revisión, se continúa con la fase que se encarga de mantener la carga constante de información, según la definición de los metadatos de carga.

Una vez implementado el Data warehouse, éste requiere mantenimiento constante. A continuación se enumeran las tareas más importantes de esta fase:

- \* Agregar nuevos usuarios
- \* Agregar consultas, tablas, vistas o información
- \* Administración de la calidad de la información
- \* Monitoreo del almacenamiento de los datos
- \* Disponibilidad de la aplicación
- \* Capacidad y desempeño
- \* Mejoras menores que no impliquen grandes cambios en los requerimientos

### **3. Caso práctico: Análisis, diseño de un Data mart de datos operativos del FONAES para medir avances físicos y financieros**

El FONAES (Fondo Nacional de Apoyo a Empresas Sociales) es un órgano desconcentrado de la Secretaría de Economía que inició actividades en el año de 1991. El objetivo principal de este órgano gubernamental es otorgar créditos y capacitación a empresas sociales de sectores de la población con altos grados de marginación, y que no tienen acceso a créditos bancarios o de otras entidades gubernamentales. La Secretaría de Hacienda y Crédito Público (SHCP) asigna recursos a la institución, el cual es administrado por el FONAES según sus propios criterios. Debido al carácter estratégico de la Institución y la ley de transparencia, es necesario generar constantemente informes que reflejen los avances de la misma en materia de Inversión y de acciones generadas. Estos reportes gerenciales le son solicitados actualmente al FONAES por diferentes áreas internas y principalmente por otras instancias gubernamentales, además de la Presidencia de la República.

#### **Planteamiento detallado del problema**

El FONAES cuenta con una representación en cada estado de la República mexicana, cada una de estas oficinas es la encargada de recibir las solicitudes de crédito y darles el seguimiento correspondiente según las reglas de operación vigentes. Actualmente se cuenta con un sistema operacional en línea llamado SIEL en la extranet<sup>6</sup> institucional. La mayoría de las representaciones no capturan todas las solicitudes recibidas en el SIEL, sino solamente las que cumplen con las reglas de operación y por lo cual serán aceptadas. Otro factor importante es que la asignación del presupuesto y algunos otros parámetros son llevados en un archivo por separado, ya que el sistema actualmente no tiene pantallas de captura para esta información.

El área usuaria y promotora del proyecto es la Dirección General de Planeación y Asistencia Técnica (DGPAT), esta dirección tiene, entre una de

---

<sup>6</sup> **Extranet.**- Es el término relacionado con la implementación de la tecnología de Internet para beneficio exclusivo de comunicación entre la organización y en donde el público en general no tiene acceso.

sus actividades funcionales, la elaboración de informes de metas e indicadores relativos a los tipos de apoyo con los que cuenta la institución.

Dicha información es solicitada a las Direcciones Generales Operativas<sup>7</sup>, a la Dirección General de Operación Regional y a la Dirección General de Administración y Finanzas, misma que es conciliada entre las áreas involucradas en el proceso, con el propósito de identificar posibles diferencias o errores en los datos. Una vez terminada la recolección de la información, se obtiene el cierre definitivo y se elaboran los informes correspondientes de indicadores y metas. Es importante mencionar que este proceso de conciliación implica utilizar una cantidad de tiempo importante, lo que puede provocar el retraso en la entrega de los informes en los tiempos establecidos.

La DGPAT ha solicitado entonces, a la Dirección de Informática y específicamente a la Subdirección de Desarrollo de Sistemas, una solución para obtener estos reportes institucionales de manera más rápida y eficiente, aprovechando al máximo la infraestructura y recursos con los que ya cuenta el FONAES.

### **Definición de objetivos**

Los cambios al sistema SIEL para ajustarlo a las necesidades actuales de la institución no serán tocados en el presente trabajo (éstos ya se encuentran implementados en una primera fase), se enfocará solamente en los requerimientos de negocio de los informes que le son solicitados, la creación del repositorio de datos y el desarrollo de la solución para tener acceso a la información. Este proyecto podría dividirse en dos objetivos específicos:

Crear un repositorio de información operativa desde el cual puedan obtenerse los reportes que le son solicitados a la Institución y que sirva además para medir su desempeño.

Crear una solución arquitectónica que permita explotar la información que será almacenada en el repositorio de datos.

---

<sup>7</sup> **Dirección General Operativa.**- Es cada una de las direcciones que se encarga de asignar recursos del FONAES a las diferentes organizaciones que se lo solicitan.

## **Definición de requerimientos**

Como se mencionó en el segundo capítulo, los requerimientos de un Data warehouse se dividen en los *requerimientos de negocio* o de *dominio específico* que se basan en las reglas de negocio de la empresa y los *requerimientos arquitectónicos* que se refieren a las características de la construcción del sistema. Se definirán entonces primero los requerimientos de negocio, que son los que dieron origen al desarrollo del proyecto.

### **Requerimientos de negocio**

Antes de entrar de lleno con el análisis de los requerimientos de negocio, se definirán los conceptos de la organización y que serán mencionados de manera recurrente a lo largo del proyecto.

**Apoyo.**- Es el crédito o aportación que se otorga a una empresa u organización.

**Instrumento.**- Son los tipos de créditos o aportación que se otorgan a las empresas. Existen varios tipos de instrumentos como: Programa de desarrollo productivo de la mujer, Apoyo financiero a Microempresas, Comercializadora social, etc

**Solicitud.**- Es la petición que hacen las organizaciones solicitando apoyos

**Presupuesto.** – Es el monto que le asigna el gobierno a FONAES para ser asignado a los diferentes apoyos que ofrece la Institución

**POA** .- Es el **P**rograma **O**perativo **A**nual, es decir, el presupuesto que le asigna el gobierno federal a la Institución con el objetivo de mejorar las condiciones de extrema pobreza en el país.

**Avance.**-Es la relación que existe entre las metas a las que tiene que llegar la institución y lo que se ha realizado en determinado momento expresado en porcentaje.

**Socio.**- Es cualquier persona que forma parte de alguna empresa social que aporte recursos financieros, habilidades y/o conocimientos, que sirvan como base para el establecimiento de la empresa social.

**Dirección General Operativa.-** El FONAES se encuentra dividido en diferentes direcciones y éstas a su vez son las encargadas de distribuir y darles seguimiento a los recursos que son asignados a las diferentes empresas sociales. Cada Dirección tiene que asignar determinados tipos de créditos, que son llamados Instrumentos.

**Empresa.-** Es un conjunto de personas o entidad que realiza alguna actividad empresarial, agrícola, artesanal, etc.

**Acciones.-** Son las actividades que realiza el FONAES con las diferentes empresas apoyadas, como son: cursos, auditorías, consultorías, etc.

**Proyecto.-** Son los proyectos que presentan las organizaciones con el objetivo de que el FONAES les asigne parte de los recursos necesarios para el desarrollo del mismo.

**Indicador.-** Es el tipo de unidad de medida que será utilizada con el objetivo de medir el desempeño del FONAES.

**Inversión.-** Es el monto asignado a los grupos u organizaciones que solicitan apoyo al FONAES.

Una herramienta muy útil para identificar los requerimientos de negocio de una empresa, es un mapa mental<sup>8</sup> que servirá como guía en la definición de las necesidades del área usuaria, en este caso la DGPAT formó parte del grupo de trabajo que definió los aspectos del negocio que serán medidos.

Para el análisis de los requerimientos de negocio, se creó un mapa mental<sup>3</sup> (figura 3.1), el cual nos ayudará a identificar de manera más precisa los requerimientos de negocio del área usuaria. Tomando como base los reportes que le son solicitados a la DGPAT, se define el mapa mental con los parámetros de medición que serán necesarios incluir en el repositorio de datos.

---

<sup>8</sup> **Mapa mental.-** Es una técnica que permite la organización y la manera de representar la información en forma fácil, espontánea, creativa, en el sentido que la misma sea asimilada y recordada por el cerebro. Creada por el Dr. Tony Buzman.

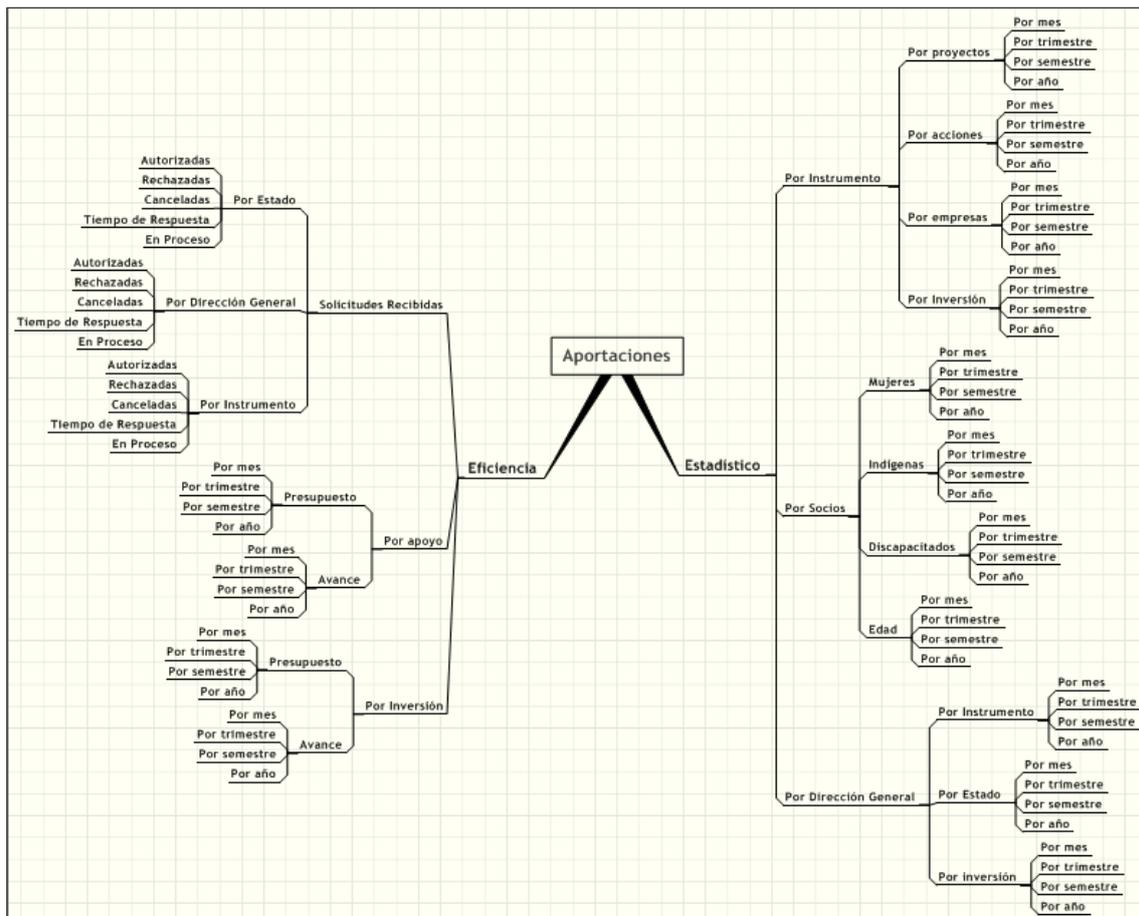
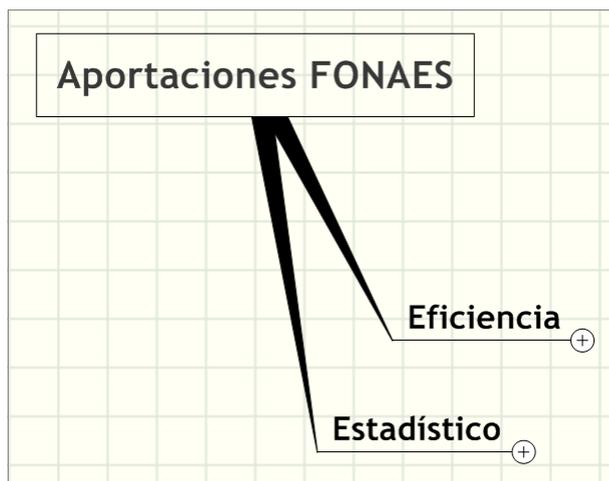


Figura 3.1 Mapa mental

Como se puede observar en el diagrama, se han separado dos aspectos a evaluar de la empresa: eficiencia y estadísticas (ver figura 3.2). En el primer grupo se medirán principalmente los tiempos de respuesta de la organización contra las solicitudes recibidas, el número de apoyos otorgados y la inversión. En el segundo, se medirá por tres rubros: Instrumento, tipos de socios y Dirección General. De esta manera se podrán identificar los instrumentos que son asignados, la eficiencia de cada Dirección General y el tipo de población que es objetivo de inversión.



**Figura 3.2**

Los reportes que se crearán para la DGPAT según los formatos entregados a la Dirección de Informática son:

1. -Avance del POA por Dirección General y Tipo de Apoyo.- Este reporte contendrá el resumen del avance de la Institución por presupuesto, Dirección General y el tipo de apoyo.

2.- Reporte de Resultados por Entidad Federativa e Instrumento.- Es el resumen anual, que se va evaluando constantemente por estado e Instrumento financiero. Este reporte sirve para tener un panorama del avance en apoyos e inversión de Institución por Entidad Federativa.

3.- Reporte de metas y avance financiero.- Es un resumen de las metas de inversión establecidas en el POA por estado e instrumento financiero.

4.- Programa Operativo Anual por Dirección General y Tipo de Apoyo.- Este reporte desplegará el resumen del presupuesto asignado a la Institución por Dirección General y tipo de apoyo.

5.- Apoyos en Micro regiones<sup>9</sup> por entidad federativa.- Este informe contiene la información resumida por estado de la República y Micro regiones de los apoyos asignados por la Institución.

6.- Apoyos en Micro regiones.- Cantidad de acciones e inversión por Micro región en el País.

7.- Apoyos a la población indígena. Inversión por tipo de apoyo a la población indígena del país.

<sup>9</sup> **Micro región.**- Subconjunto de una región identificada como de alta marginación social.

8.- Apoyos a la población indígena por entidad federativa. Apoyos e inversión por estado de la República a la población indígena.

9.- Empleos generados por Instrumento de apoyo.- Resumen de los apoyos generados por año por Instrumento.

10.- Distribución de socios por grupo de edad.- Resumen de los socios beneficiados por rango de edad.

11.- Distribución de socios por género.- Resumen también del género de los socios de las organizaciones apoyadas por el FONAES.

12.- Proyectos de impacto regional.- Resumen de la inversión que se les ha dado a los proyectos que se tienen identificados como de Impacto Regional.

13.- Proyectos de impacto regional por entidad federativa.- Resumen por estado de la República de los proyectos de impacto regional apoyados por el FONAES.

14.- Apoyos para el fomento y la difusión de la artesanía nacional por instrumento.- Existe un programa especial en el FONAES que se encarga de difundir y apoyar a la artesanía en el país, este reporte filtraría estos apoyos por Instrumento.

15.- Apoyos para el fomento y la difusión de la artesanía nacional por entidad federativa.-Desglose de este tipo de apoyos, pero por estado de la República.

16.- Apoyos a Fondos y Fideicomisos. – Son los apoyos que son identificados así.

## **Requerimientos arquitectónicos**

### **Requerimientos de funcionalidad**

**Impresión.-** Se requiere que se pueda imprimir desde cualquier máquina y que se tenga la posibilidad de exportar a Excel ®, para que la información pueda ser usada para otros reportes no prediseñados.

**Presentación.-** Es necesario que los reportes generados tengan el mismo formato que los reportes que actualmente son generados por la Institución, en los cuales se respete el logo actual, los espacios, el tipo de letra, los colores, etc.

### **Requerimientos de uso**

**Accesibilidad.-** Se requiere que se pueda acceder a los reportes desde cualquier centro regional de la república, al igual que el Sistema Integral en Línea (SIEL).

**Estética.-** La manera de presentar los reportes debe ser igual que los que actualmente está generando el SIEL.

### **Requerimientos de confiabilidad**

**Exactitud.-** Estos reportes deben de coincidir con la información que se esta capturando y generando constantemente en la Institución

**Recuperación.-** El nivel de recuperación de la información no es tan crítico como el de la información de la operación de la Institución, ya que la información para la generación de los reportes es sólo de carácter informativo.

### **Requerimientos de rendimiento**

**Tiempo de respuesta.-** El tiempo de respuesta debe de ser de menos de 10 segundos, ya que es un sistema web.

**Tiempo de carga.-** El tiempo de carga de la información del sistema operativo al Data mart no es crítico, ya que la información será procesada y cargada al Data mart en horarios de no operación de la Institución.

### **Requerimientos de soporte**

**Sincronización.-** Debido a que es necesario contar con información lo más actualizada posible, respecto a la operación del FONAES.

**Escalable.-** Los reportes que le son solicitados actualmente a la Institución no varían mucho año con año. Por ello, el almacén de información debe de contener la mayor cantidad de datos operativos que permitan obtener cualquier tipo de reporte que le sea solicitado posteriormente a la Institución.

### **Requerimientos de diseño**

**Impresión.-** Se le permitirá al usuario ver el reporte en pantalla desde dónde podrá ser enviado a impresión directamente o exportarlo a los diferentes.

**Seguridad.-** La seguridad será manejada a través de la seguridad implementada en el SIEL, en donde se tienen que dar permisos por usuario a cada reporte.

## Diseño y modelado

### Diseño arquitectónico

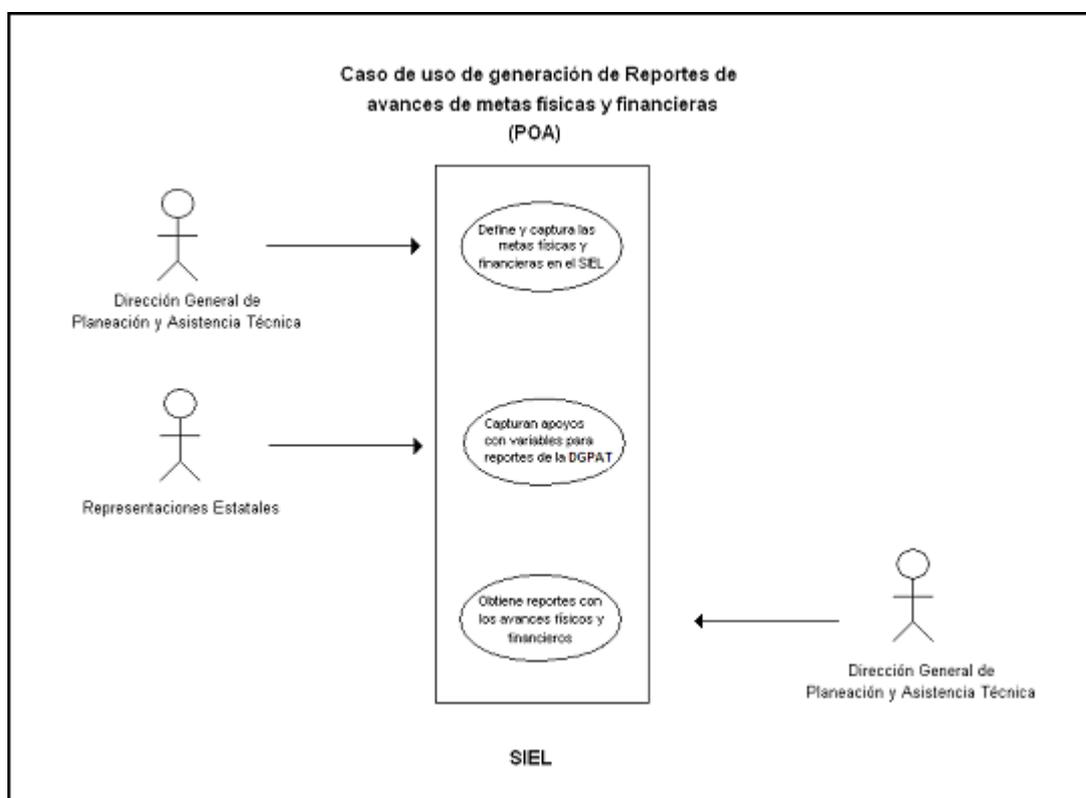
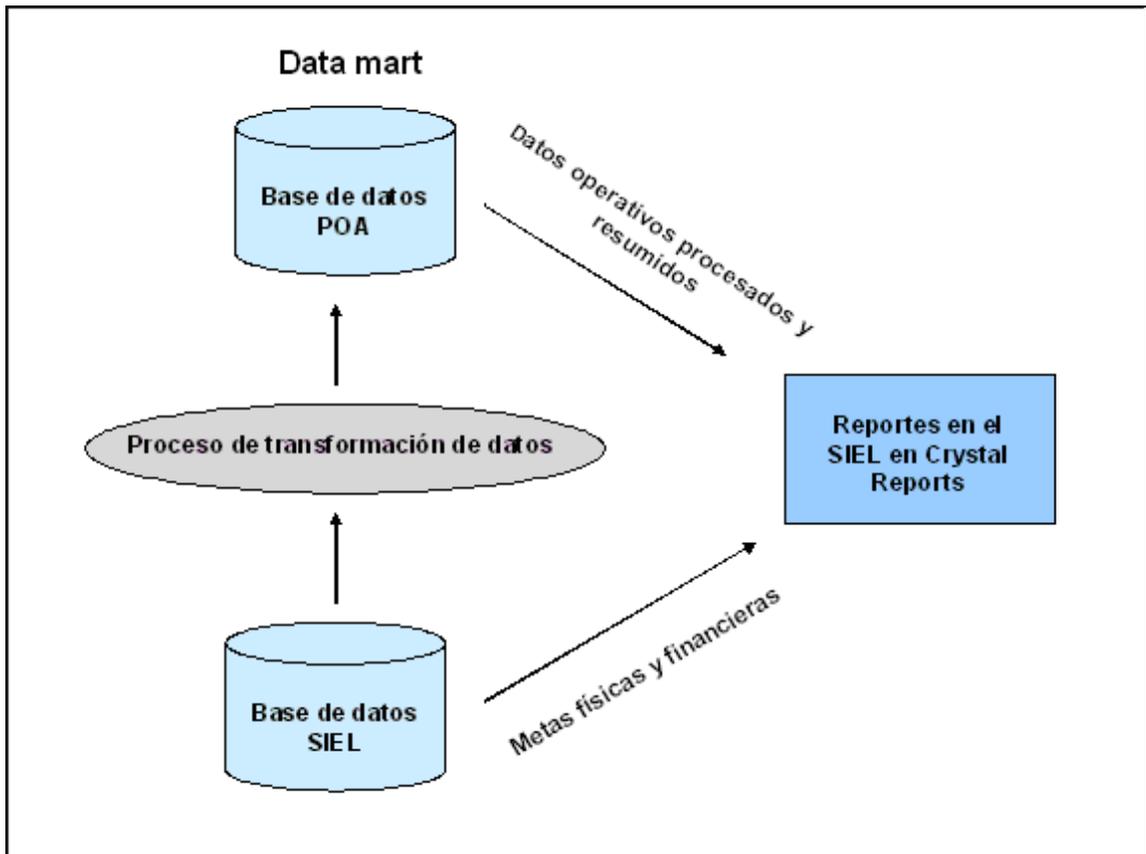


Figura 3.3 Casos de uso

La figura 3.3 muestra los casos de uso que muestra la interacción entre los actores que usarán la solución del datamart y el sistema SIEL. La solución será transparente para el usuario, ya que se usará el mismo módulo de reportes que se tiene implementado en el SIEL institucional. Este módulo se encuentra en ASP's y Crystal Reports como servidor de reportes. Para la publicación de reportes se usará la estructura ya implementada en donde sólo es necesario colocar el reporte en el servidor y configurar los parámetros de entrada en las tablas de configuración del sistema.



**Figura 3.4 Diseño arquitectónico**

Debido a que las partidas que son asignadas para cada tipo de apoyo son modificadas constantemente, se decidió tomar estas métricas desde el sistema operacional en línea SIEL, como se puede observar en la figura 3.4. Este sistema cuenta con un módulo llamado “Métricas e indicadores POA”, el cual es mantenido por la DGPAT según se vayan modificando los recursos al FONAES. También desde la base de datos SIEL se tomarán los datos que alimentarán al Data mart después del proceso de transformación y carga. Los reportes serán generados desde el sistema SIEL, tomando la información operativa almacenada en el Data mart y las métricas de referencia desde el SIEL en tiempo real.

## Diseño y modelado de base de datos

La fuente única de datos para el data mart será la base de datos del SIEL, que es en donde todos los usuarios operacionales deberán capturar la información. También de la base de datos SIEL, se obtendrán los datos de los presupuestos asignados por Dirección General Operativa, Instrumento y modalidad. Y de esta manera, permitir que los parámetros de comparación sean modificados según se modifique el presupuesto anual, y así, permitir que sean dinámicos y ajustables a las necesidades del FONAES.

### Definición de fuentes de datos

Fuente	Negocio	Dueño	Plataforma	Ubicación	Descripción de la fuente de datos
SIEL	Operaciones FONAES	Todas las Direcciones operativas	SQL Server	Servidor de producción	Datos operacionales del FONAES
SIEL	Definición de POA	DGPAT	SQL Server	Servidor de producción	Métricas e indicadores POA

Tabla 3.1 Definición de fuentes de datos del Data mart

### Definición de tablas de Hechos

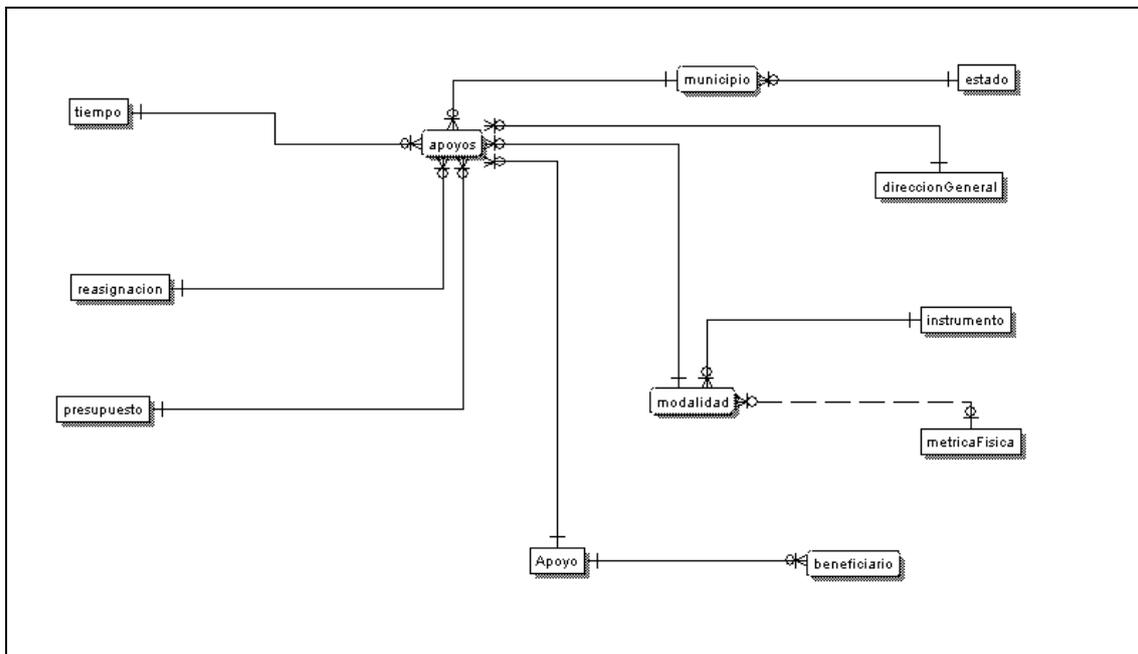


Figura 3.5. Hechos "Apoyos"

Debido a la cantidad de información que será manejada como carga inicial (alrededor de 60,000 apoyos) y los apoyos que serán asignados cada año (alrededor de 3000 registros), se decidió tener en el data mart en dos tablas de hechos. Una de detalle con el nivel más bajo de granularidad llamada “Apoyos” (Figura 3.5), para los reportes solicitados y la otra de resumen llamada “Avance” (Figura 3.6), para tener un panorama global de la información generada por la Institución.

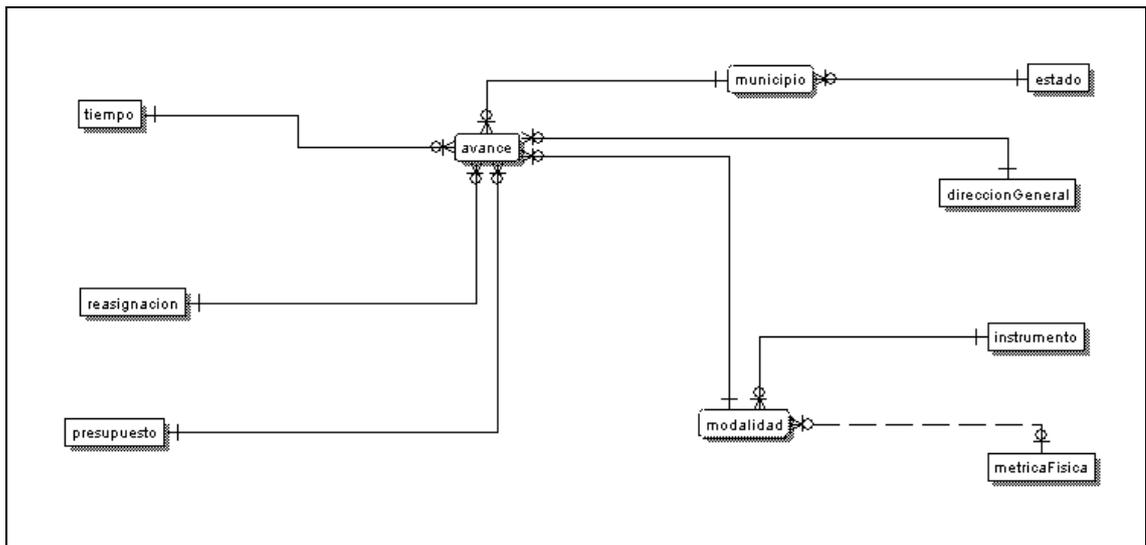


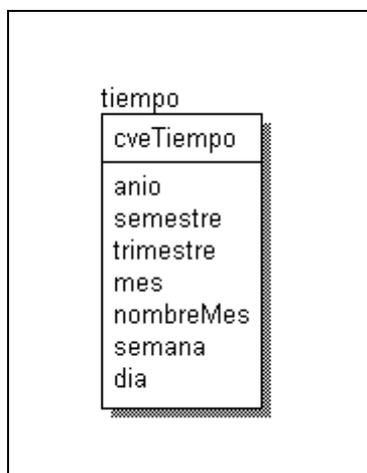
Figura 3.6 Hechos “Avance”

## Dimensiones del Data mart

Nombre	Descripción
tiempo	Dimensión base de cualquier data mart o datawarehouse. En esta dimensión se separan las unidades de tiempo: año, semestre, trimestre, mes, semana, día, etc
municipio	Municipio de la República Mexicana, en dónde se puede identificar también el estado al que pertenece con una llave foránea a la tabla estado.
Dirección general	Dirección del FONAES, la cual maneja determinados tipos de apoyos o instrumentos.
apoyo	Es el crédito o inversión que es asignado por el FONAES a las organizaciones.
beneficiario	Individuo que es beneficiado con algún apoyo, en dónde las jerarquías identificarán el sector de la población que esta siendo apoyado, como edad, sexo, etnia, etc
Instrumento	Es el tipo de crédito o inversión que se les asigna a las organizaciones por el FONAES, en donde una de las jerarquías es también la modalidad, o sea un subconjunto de cada apoyo.
presupuesto	Es la partida del presupuesto que se le asigna al FONAES. Una de las jerarquías de esta dimensión sería la reasignación que es la sub-partida del presupuesto que se le asigna al FONAES.

**Tabla 3.2 Dimensiones del Data mart**

### Dimensión tiempo



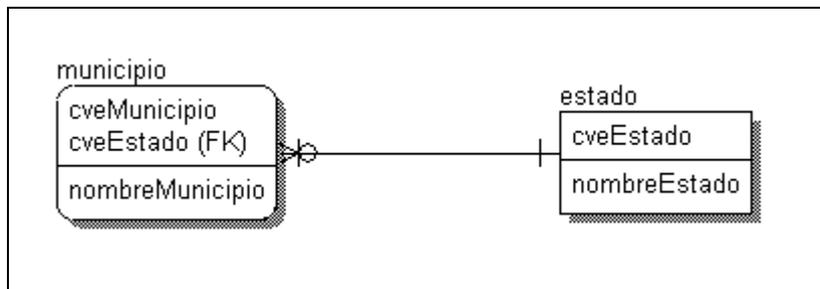
**Figura 3.7 Dimensión tiempo**

Nombre	Descripción	Valores ejemplo
cveTiempo	Llave sintética de la dimensión, consecutivo.	1,2,3,4,5...n
año	Año del registro de la fecha insertada	2000,2001, etc
semestre	Semestre del año	1,2. Primer o segundo semestre del año

Nombre	Descripción	Valores ejemplo
trimestre	Trimestre del año	1,2, 3. Primer, segundo o tercer trimestre
mes	Mes del año	1-12.
Nombre mes	Descripción del mes de la fecha insertada	Enero, febrero, marzo, abril....etc
semana	Número de la semana del año de la fecha insertada	1,2,3....52.
día	Día de la fecha insertada	1, 2, 3...31.

**Tabla 3.3 Detalle de la dimensión tiempo**

### Dimensión municipio

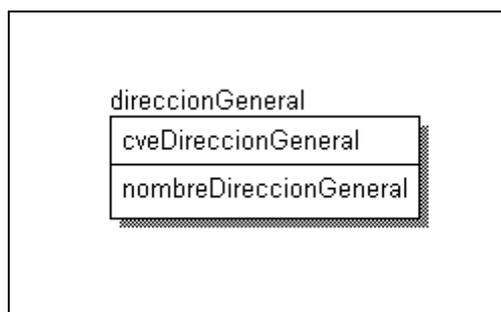


**Figura 3.8 Dimensión municipio**

Nombre	Descripción	Valores ejemplo
cveMunicipio	Llave sintética de la dimensión, consecutivo.	1,2,3,4,5...n
cveEstado	Llave foránea de la tabla estado	1.2. 3...32
nombreMunicipio	Nombre del Municipio	Ecatepec, Gustavo A Madero, Coyoacan, etc

**Tabla 3.4 Detalle de la dimensión municipio**

### Dimensión Dirección General

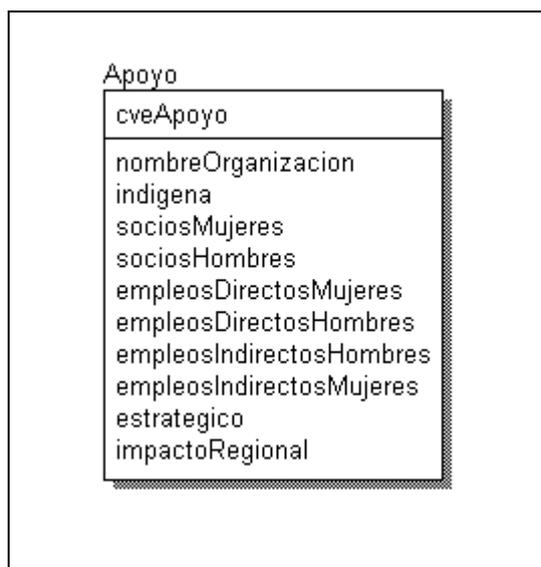


**Figura 3.9 Dimensión Dirección General**

Nombre	Descripción	Valores ejemplo
cveDireccionGeneral	Clave única de la Dirección General	1,2 ,3,4
nombreDireccionGeneral	Descripción de la Dirección general	DGPAT, Operación Regional, etc

**Tabla 3.5 Detalle de la dimensión Dirección General**

## Dimensión Apoyo



**Figura 3.10 Tabla de hechos Apoyo**

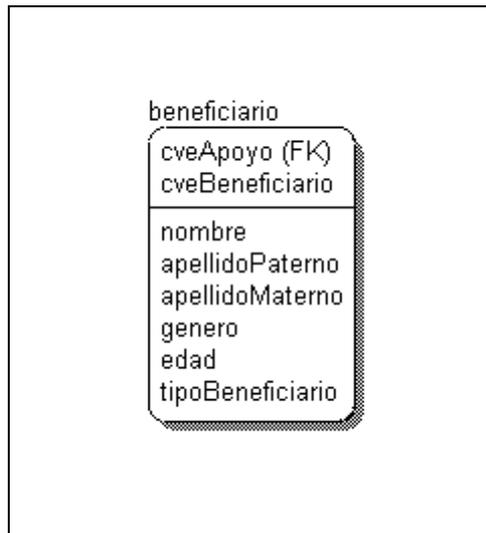
Nombre	Descripción	Valores ejemplo
cveApoyo	Identificador único del apoyo.	1,2,3,4
nombreOrganizacion	Nombre del grupo u organización social	Comuneros de Occidente, Artesanías Lydia
Indígena	Campo binario que identifica al apoyo como indígena	0,1
sociosMujeres	Número de socios mujeres	12,5,2,etc
sociosHombres	Número de socios hombres	1,10, 7 etc
empleosIndirectosMujeres	Empleos generados a mujeres indirectamente por el apoyo otorgado.	1,0,42,etc
empleosIndirectosHombres	Empleos generados a hombres indirectamente por el apoyo otorgado.	0,10, 20, 30, etc
estrategico	Sí se trata o no de un apoyo estratégico para el desarrollo del entorno.	0,1
impactoRegional	Este campo identifica si el apoyo	0,1

	tendrá un impacto a nivel regional	
--	------------------------------------	--

**Tabla 3.6 Detalle de la tabla de hechos Apoyo**

Esta es la dimensión más importante del Data mart, ya que contiene los apoyos que son otorgados por la institución a las diferentes organizaciones que los solicitan.

### Dimensión Beneficiario



**Figura 3.11 Dimensión beneficiario**

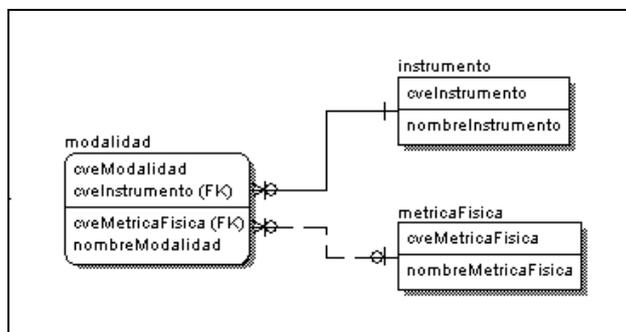
Nombre	Descripción	Valores ejemplo
cveApoyo	Llave foránea del apoyo al cual pertenece el beneficiario	1,2,3,4...n
cvebeneficiario	Llave única del beneficiario	1,2,3,4...n
nombre	Nombre del beneficiario	Juan, Fernanda, etc.
apellidoPaterno	Apellido paterno del beneficiario	García, Campos, etc.
apellidoMaterno	Apellido materno del beneficiario	López, García, etc.
genero	Género al cual pertenece el beneficiario	Femenino, masculino
Edad	Edad del beneficiario	20, 25, 18, etc.
tipoBeneficiario	El rol que juega el beneficiario en la Organización	Socio, jornalero, comunero, etc.

**Tabla 3.7 Detalle de la dimensión beneficiario**

Como se puede observar en la figura 3.9, esta dimensión está formada por los integrantes o beneficiarios que conforman cada una de las organizaciones que reciben apoyos del FONAES. Con esta dimensión, se

podrá identificar el tipo de población que esta siendo beneficiada por la Institución.

### Dimensión Modalidad



**Figura 3.12 Dimensión modalidad**

#### Tabla modalidad

Nombre	Descripción	Valores ejemplo
cveInstrumento	Clave del Instrumento	1,2, 3, 4, 5
cveModalidad	Clave de la modalidad	1, 2, 3,4, 5
cveMetricaFisica	Clave de la métrica física que corresponde medir para este instrumento.	1,2, 3, 4, 5
nombreModalidad	Nombre del instrumento financiero	Procomer, Cajas Solidarias, ADE, Comercializadora social

**Tabla 3.8 Detalle de la tabla modalidad**

#### Tabla Instrumento

Nombre	Descripción	Valores ejemplo
cveInstrumento	Clave de la modalidad	1, 2, 3, 4, 5
nombreInstrumento	Nombre de la modalidad	APORTACIONES PARA CAPACITACION , PDPM , ADE Directo, ADE Indirecto

**Tabla 3.9 Detalle de la tabla instrumento**

### Tabla metricaFisica

Nombre	Descripción	Valores ejemplo
cveMetricaFisica	Clave de la métrica física	1, 2, 3, 4, 5
nombreMetricaFisica	Nombre de la métrica física	Acción, apoyo, etc.

Tabla 3.10 Detalle de la tabla metricaFisica

### Dimensión presupuesto

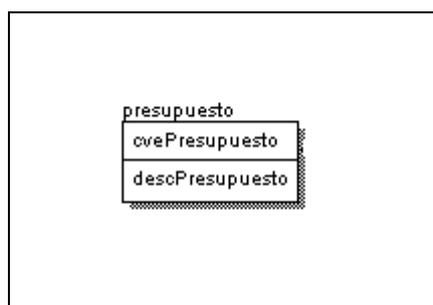


Figura 3.13 Dimensión presupuesto

### Dimensión presupuesto

Nombre	Descripción	Valores ejemplo
cvePresupuesto	Clave del presupuesto	1, 2, 3, 4, 5
descPresupuesto	Descripción del presupuesto	Base FONAES, Base PEC

Tabla 3.11 Detalle de la dimensión presupuesto

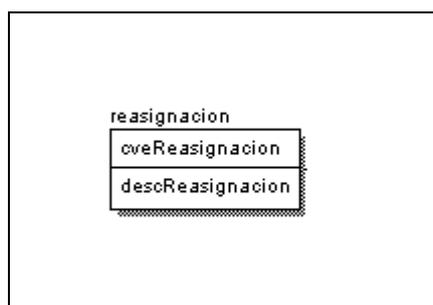


Figura 3.14 Dimensión reasignación

### Tabla reasignación

Nombre	Descripción	Valores ejemplo
cveReasignacion	Clave de la reasignación	1, 2, 3, 4, 5
descReasignacion	Descripción de la reasignación	FONAES, ADE, Microrregiones

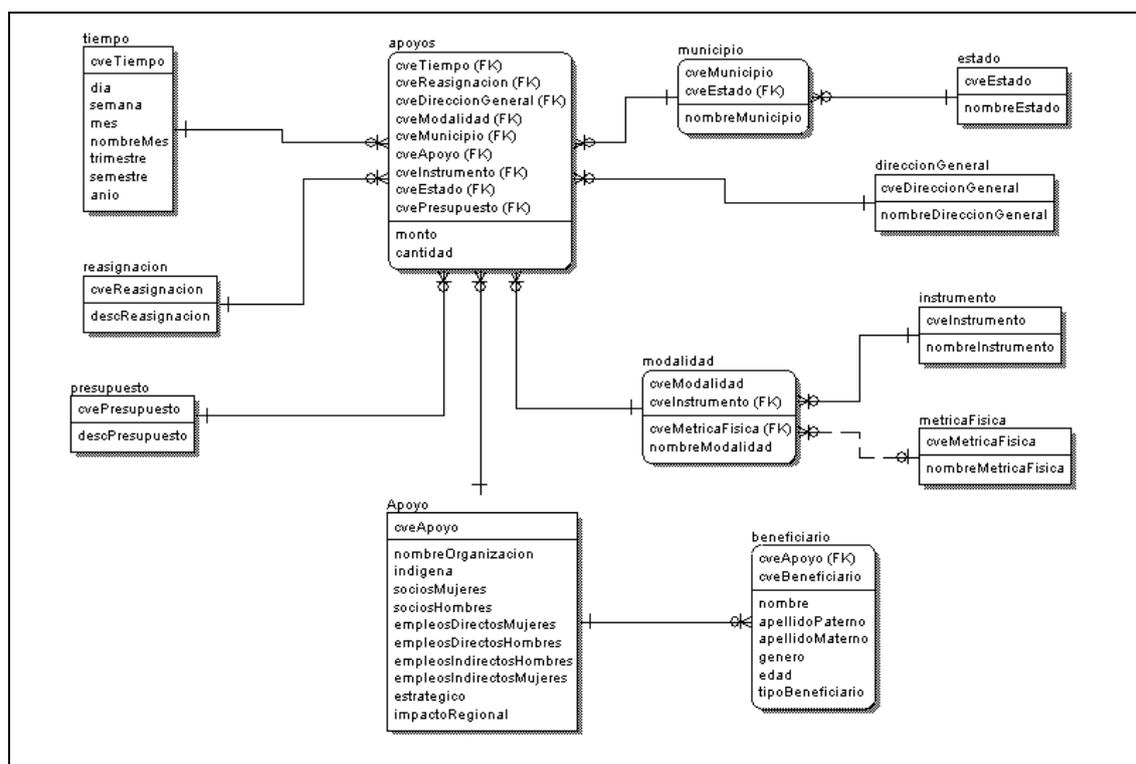
Tabla 3.12 Detalle de la dimensión reasignación

## Descripción de las tablas de hechos

Nombre	Descripción	Regla de Agrupación por default
Apoyos	Tabla detalle del data mart con el máximo nivel de granularidad <sup>10</sup> : por apoyo otorgado	Por presupuesto
Avance	Tabla de hechos resumen a nivel de instrumento.	Por instrumento

**Tabla 3.13 Detalle de las tablas de Hechos**

En la figura 3.15 se puede observar el esquema de la tabla de hechos Apoyos con las llaves foráneas de las diferentes dimensiones que la conforman. Las tablas de dimensiones con los campos definidos y sus llaves.



**Tabla 3.15 Esquema de la tabla de hechos Apoyos**

<sup>10</sup> **Granularidad.**-Nivel de detalle con el cual se manejará la información.

En la figura 3.16 se muestra la tabla de hechos Avance, como se puede observar, tiene las relaciones de las dimensiones de municipio, Dirección General, modalidad, presupuesto, reasignación y la básica del tiempo. En este caso, no se incluyó la dimensión apoyos que es el nivel de granularidad más bajo.

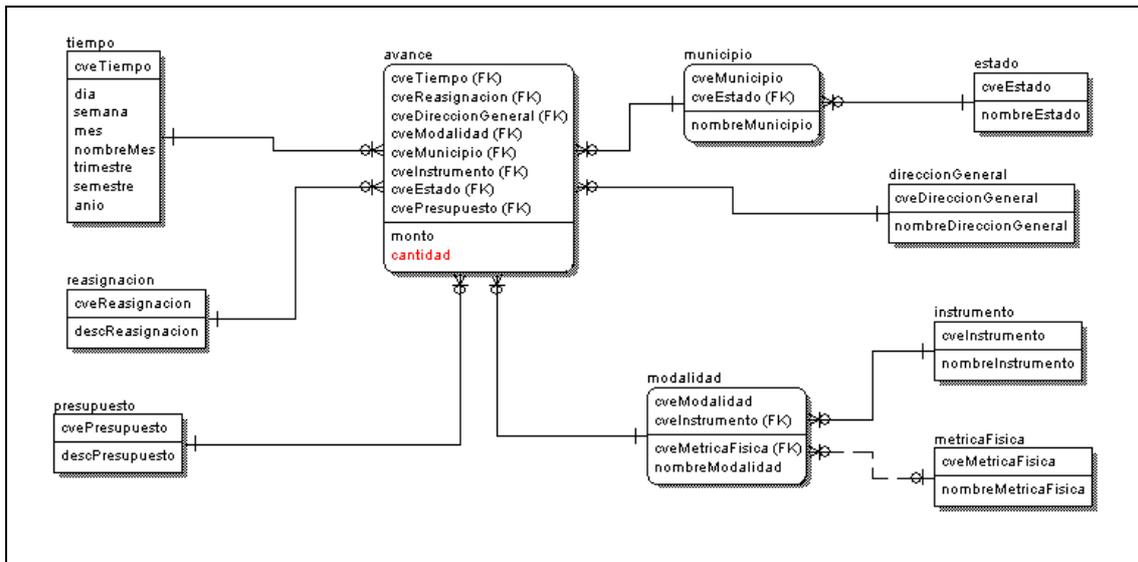


Tabla 3.16 Esquema de la tabla de hechos Avance

#### 4. Caso práctico: Implementación de un Data mart de datos operativos del FONAES para medir avances físicos y financieros.

##### 4.1. Proceso ETL

Una vez que se ha terminado el diseño del data mart en su fase arquitectónica y de modelado de datos, se continúa entonces con la fase de Implementación. Para ello, es necesario crear un esquema con el plan de carga de la información en el nivel más alto, en dónde se definen las fuentes de datos de dónde se extraerá la información y el orden de carga de cada una de las dimensiones y las tablas de hechos. Se anota la carga aproximada de registros a procesar por cada una de las dimensiones, que servirá también para definir el diseño físico de las tablas. En nuestro caso práctico sólo se tiene una única fuente de información que es la base de datos SIEL que es alimentada por el SIEL en línea y el sistema Cliente-Servidor de Aportaciones. El proceso ETL se hará de manera manual en la carga inicial y posteriormente el data mart será actualizado de manera automática todas las noches, para mantener la información lo más consistente posible con la que se encuentra en operación.

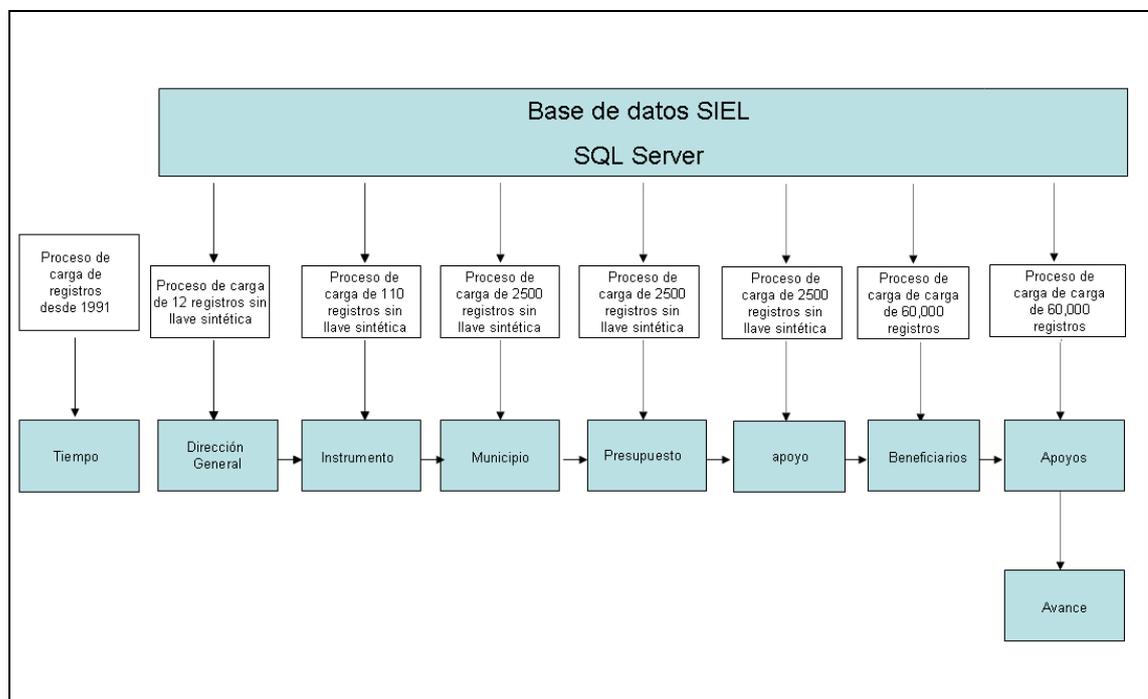


Figura 4.1 Plan ETL de alto nivel

Para continuar con la implementación del data mart es necesario crear un documento en el cual se detalle cada uno de los campos que son necesarios para cada dimensión, el origen de los mismos y el proceso de carga que será el encargado de obtener, modificar --en su caso-- y subir la información al modelo de datos del data mart.

## Dimensión tiempo

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fuente	Tabla	Campo	Proceso de carga
tiempo	cveTiempo	NA	NA	NA	sp_CargaTiempo
tiempo	dia	NA	NA	NA	sp_CargaTiempo
tiempo	Semana	NA	NA	NA	sp_CargaTiempo
tiempo	mes	NA	NA	NA	sp_CargaTiempo
tiempo	nombreMes	NA	NA	NA	sp_CargaTiempo
tiempo	trimestre	NA	NA	NA	sp_CargaTiempo
tiempo	semestre	NA	NA	NA	sp_CargaTiempo
tiempo	anio	NA	NA	NA	sp_CargaTiempo

NA= No aplica

**Tabla 4.1 Definición de campos a cargar Dimensión tiempo**

. En la figura 4.1 se puede observar – de izquierda a derecha -- que la primera dimensión a cargar será la dimensión tiempo, la cual no será cargada desde la base de datos sino con un proceso que insertará los registros de tiempo desde la primera aportación que se tiene en el sistema, en el año de 1991 hasta la fecha actual. Estos registros serán necesarios para ingresar los registros históricos de la Institución hasta los actuales.

A continuación se muestra el código del proceso de carga **sp\_cargaTiempo**, que es el que carga los registros en la dimensión tiempo. Este proceso segmenta cada fecha en año, semestre, trimestre, mes y día para insertar la información en cada uno de los campos de la dimensión.

--Se limita el registro inicial a la primera fecha de cheques en producción:

```

DECLARE @FechaInicio As smalldatetime,
        @FechaFinal As smalldatetime,
        @FechaCursor As smalldatetime,
        @cursorClave AS int,
        @DescripcionMes AS Varchar(10),
        @Semestre AS int

```

```

SET @FechaInicio='30/11/1991 '
SET @FechaFinal=getdate()

```

```

SET @FechaCursor ='30/11/1991'
SET @cursorClave=1

WHILE @FechaCursor <= @FechaFinal
BEGIN

    SELECT @DescripcionMes=Descripcion from Siel.dbo.Catmeses
    WHERE clave= datepart(month,@FechaCursor)
    IF (datepart(quarter,@FechaCursor) = 1 OR datepart(quarter,@FechaCursor)=2)

        SET @Semestre=1
    ELSE
        SET @Semestre=2

    INSERT INTO tiempo
    VALUES (@cursorClave,
            datepart(day,@FechaCursor),
            datepart(week,@FechaCursor),
            datepart(month,@FechaCursor),
            @DescripcionMes,
            datepart(quarter,@FechaCursor),
            @Semestre,
            datepart(year,@FechaCursor))

    SET @FechaCursor =Dateadd(day,1, @FechaCursor)
    SET @cursorClave=@cursorClave+1
END

```

## Dimensión Dirección General

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fuente	Tabla	Campo	Proceso de carga
direccionGeneral	cveDireccionGeneral	SIEL	Dir_gral	clave	sp_cargaDireccion
direccionGeneral	nombreDireccionGeneral	SIEL	Dir_gral	descri3	sp_cargaDireccion

**Tabla 4.2 Definición de campos a cargar Dimensión Dirección General**

Como se puede ver en la tabla 4.2 de definición de carga de esta dimensión, los campos a cargar se pasan directamente de la base de datos SIEL sin ningún proceso de transformación. De esta manera se simplifica el proceso de carga y el mantenimiento del Data mart.

## Dimensión presupuesto

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
presupuesto	cvePresupuesto	SIEL	kpresupuesto	cvePresupuesto	sp_cargaPresupuesto
presupuesto	descPresupuesto	SIEL	kpresupuesto	descPresupuesto	sp_cargaPresupuesto

**Tabla 4.3 Definición de campos a cargar dimensión presupuesto**

## Dimensión reasignación

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
reasignacion	cveReasignacion	SIEL	kreasignacion	cveReasignacion	sp_cargaDireccion
reasignacion	descReasignacion	SIEL	kreasignacion	descReasignacion	sp_cargaDireccion

**Tabla 4.4 Definición de campos a cargar dimensión reasignación**

Al igual que la dimensión Dirección General, la carga de presupuesto (tabla 4.3) y asignación (tabla 4.4) se harán directo de la base de datos sin necesidad de generar llaves sintéticas para facilitar su mantenimiento.

## Dimensión modalidad

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
metricafisica	cveMetricaFisica	SIEL	kUnidadMetrica	cveUnidadMetrica	sp_cargaUnidad
metricafisica	nombreMetricaFisica	SIEL	kUnidadMetrica	descUnidadMetrica	sp_cargaUnidad
instrumento	cveInstrumento	SIEL	Cat_tipInstrumento	clave	sp_cargaInstrumento
instrumento	nombreInstrumento	SIEL	Cat_tipInstrumento	descri3	sp_cargaInstrumento
modalidad	cveModalidad	SIEL	Cat_modalidad	ldModalidad	sp_cargaModalidad
modalidad	cveInstrumento	SIEL	Cat_modalidad	ldTipInstrumento,	sp_cargaModalidad
modalidad	cveMetricaFisica	SIEL	Cat_modalidad	Descripcion	sp_cargaModalidad
modalidad	nombreModalidad	SIEL	Cat_tipInstrumento	cveUnidadMetrica	sp_cargaModalidad

**Tabla 4.5 Definición de campos a cargar dimensión modalidad**

En el caso de la dimensión modalidad, es importante respetar el orden de carga, ya que esta dimensión tiene llaves foráneas, una de las cuales es parte de la llave primaria. Es por esto, que debe iniciarse la carga de la información con la tabla de “metricaFisica”, instrumento y por último la tabla de modalidad.

## Dimensión municipio

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
estado	cveEstado	SIEL	estados	clave	sp_cargaEstado
estado	nombreEstado	SIEL	estados	Descri1	sp_cargaEstado
municipio	cveMunicipio	SIEL	municipi	mun	sp_cargaMunicipio
municipio	cveEstado	SIEL	municipi	edo	sp_cargaMunicipio
municipio	nombreMunicipio	SIEL	municipi	municipio	sp_cargaMunicipio

**Tabla 4.6 Definición de campos a cargar Dimensión municipio**

Como en el caso de la dimensión modalidad, se debe de cargar la información en el orden en que se muestra en la tabla 4.6, debido a que la llave de esta dimensión esta conformada por la clave de estado y municipio, es necesario entonces poblar en primera instancia la tabla de estados y posteriormente la tabla de municipio.

### Dimensión apoyo

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
apoyo	cveApoyo	SIEL	maestro	Clave_Unica	sp_cargaApoyo
apoyo	nombreOrganizacion	SIEL	maestro	Organizacion	sp_cargaApoyo
apoyo	indigena	SIEL	maestro	Indigena	sp_cargaApoyo
apoyo	sociosMujeres	SIEL	maestro	Socios_Mujeres	sp_cargaApoyo
apoyo	sociosHombres	SIEL	maestro	Socios_Hombres	sp_cargaApoyo
apoyo	empleosDirectosMujeres	SIEL	maestro	Empl_Mujeres	sp_cargaApoyo
apoyo	empleosDirectosHombres	SIEL	maestro	Empl_Hombres	sp_cargaApoyo
apoyo	empleosIndirectosMujeres	SIEL	maestro	Empl_Mujeres *.50	sp_cargaApoyo
apoyo	empleosIndirectosHombres	SIEL	maestro	Empl_Hombres *.50	sp_cargaApoyo
apoyo	estratégico	SIEL	maestro	estrategico	sp_cargaApoyo
apoyo	impactoRegional	SIEL	maestro	impactoRegional	sp_cargaApoyo
beneficiario	cveApoyo	SIEL	tParticipante	CveUnica	sp_cargaBenef
beneficiario	cveBeneficiario	SIEL	tParticipante	cveParticipante	sp_cargaBenef
beneficiario	nombre	SIEL	tParticipante	Nombre	sp_cargaBenef
beneficiario	apellidoPaterno	SIEL	tParticipante	ApPaterno	sp_cargaBenef
beneficiario	apellidoMaterno	SIEL	tParticipante	ApMaterno	sp_cargaBenef
beneficiario	genero	SIEL	tParticipante	sexo	sp_cargaBenef
beneficiario	edad	SIEL	tParticipante	edad	sp_cargaBenef
Beneficiario	tipoBeneficiario	SIEL	kTipoParticipante	descTipoParticipante	sp_cargaBenef

**Tabla 4.7 Definición de campos a cargar dimensión apoyo**

Para la carga de la dimensión apoyo (tabla 4.7) es necesario iniciar con la tabla de apoyo en la cual el cálculo para los empleos indirectos se hará con un cálculo a los empleos directos que genera dicho apoyo, esto, porque en la mayoría de las oficinas regionales no capturan esta información.

## Tabla de hechos “Apoyos”

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
apoyos	cveTiempo	SIEL	cheques	fechaElab	sp_cargaApoyos
apoyos	cveReasignacion	SIEL	maestro	cveReasignacion	sp_cargaApoyos
apoyos	cveDireccionGeneral	SIEL	maestro	dirgral	sp_cargaApoyos
apoyos	cveModalidad	SIEL	maestro	tipo_modalidad	sp_cargaApoyos
apoyos	cveMunicipio	SIEL	maestro	municipio	sp_cargaApoyos
apoyos	cveApoyo	SIEL	cheques	chClaveUnica	sp_cargaApoyos
apoyos	cveInstrumento	SIEL	maestro	tipo_instrumento	sp_cargaApoyos
apoyos	cveEstado	SIEL	maestro	estado	sp_cargaApoyos
apoyos	monto	SIEL	cheques	monto	sp_cargaApoyos
apoyos	cantidad	NA	NA	Constante 1	sp_cargaApoyos

**Tabla 4.8 Definición de campos a cargar tabla de hechos “Apoyos”**

Gracias a que en ningún caso se generaron llaves sintéticas, la carga de la tabla de hechos “Apoyos” se facilita, como se puede ver en la tabla 4.8 el campo de cantidad se inserta como constante, ya que se genera un cheque por cada uno de los apoyos.

## Tabla de hechos “Avance”

En el caso de la tabla de hechos avance, se requiere solamente crear un proceso que agrupe la información según los campos definidos en el modelo de datos desde la tabla de hechos Apoyos como se define en la tabla 4.9.

Destino Data mart		Origen Fuente de datos			
Tabla	Campo	Fte	Tabla	Campo	Proceso de carga
avance	cveTiempo	POA	Apoyos	cveTiempo	sp_cargaAvance
avance	cveReasignacion	POA	Apoyos	cveReasignacion	sp_cargaAvance
avance	cveDireccionGeneral	POA	Apoyos	cveDireccionGeneral	sp_cargaAvance
avance	cveModalidad	POA	Apoyos	cveModalidad	sp_cargaAvance
avance	cveMunicipio	POA	Apoyos	cveMunicipio	sp_cargaAvance
avance	cveInstrumento	POA	Apoyos	cveInstrumento	sp_cargaAvance
avance	cveEstado	POA	Apoyos	cveEstado	sp_cargaAvance
avance	monto	POA	Apoyos	Sum(monto)	sp_cargaAvance
avance	cantidad	POA	Apoyos	Sum(cantidad)	sp_cargaAvance

**Tabla 4.9 Definición de campos a cargar tabla de hechos “Avance”**

## 4.2. Explotación de la información

Una vez que la información se encuentra ya cargada en el data mart con el proceso ETL y programados los procesos de actualización, se sigue con el proceso de explotación de la información. Para ello, se parte de la definición de los reportes definidos por el área usuaria.

### 4.2.1. Reporte de resultados por Entidad Federativa e Instrumento

Definición de reporte	
Nombre del reporte:	Reporte de resultados por Entidad Federativa e Instrumento
Descripción:	En este reporte se muestran la cantidad de apoyos e inversión que cada una de las Entidades Federativas recibió por tipo de Instrumento Financiero
Frecuencia de uso :	Mensualmente y cada vez que se requiera
Parámetros de entrada:	Un rango de fechas, entidad federativa e Instrumento financiero
Totales	Por Entidad Federativa
Grupos:	Por Entidad Federativa e Instrumento
Cálculos:	ninguno

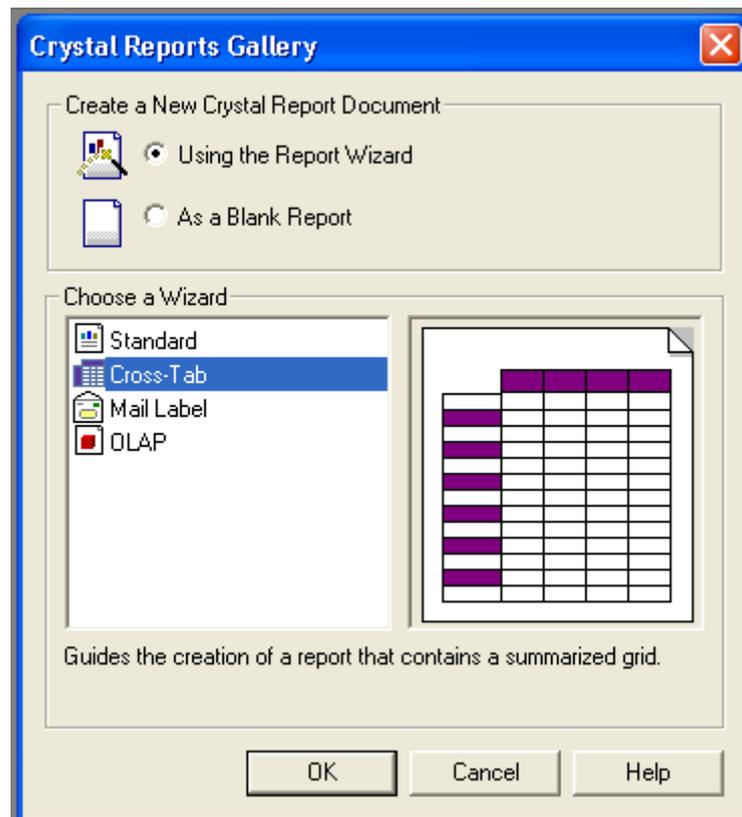
**Tabla 4.10 Definición de reporte de resultados por Entidad e Instrumento**

Para este reporte se usa la tabla de hechos “Avance” en donde el nivel de granularidad es a nivel de resumen de las dimensiones definidas para esta tabla. El resumen que se requiere es a nivel de Estado e Instrumento Financiero.

Este reporte será publicado en la Extranet institucional, en el sistema Siel, -- como se mencionó en el capítulo anterior--. Para esto, es necesario crear el reporte en Crystal Reports<sup>11</sup>, que es la herramienta usada para el desarrollo de los reportes en el SIEL. A continuación se detalla el proceso de desarrollo de este reporte , usando como fuente principal de información el data mart, ya creado en SQL Server en una base de datos llamada POA.

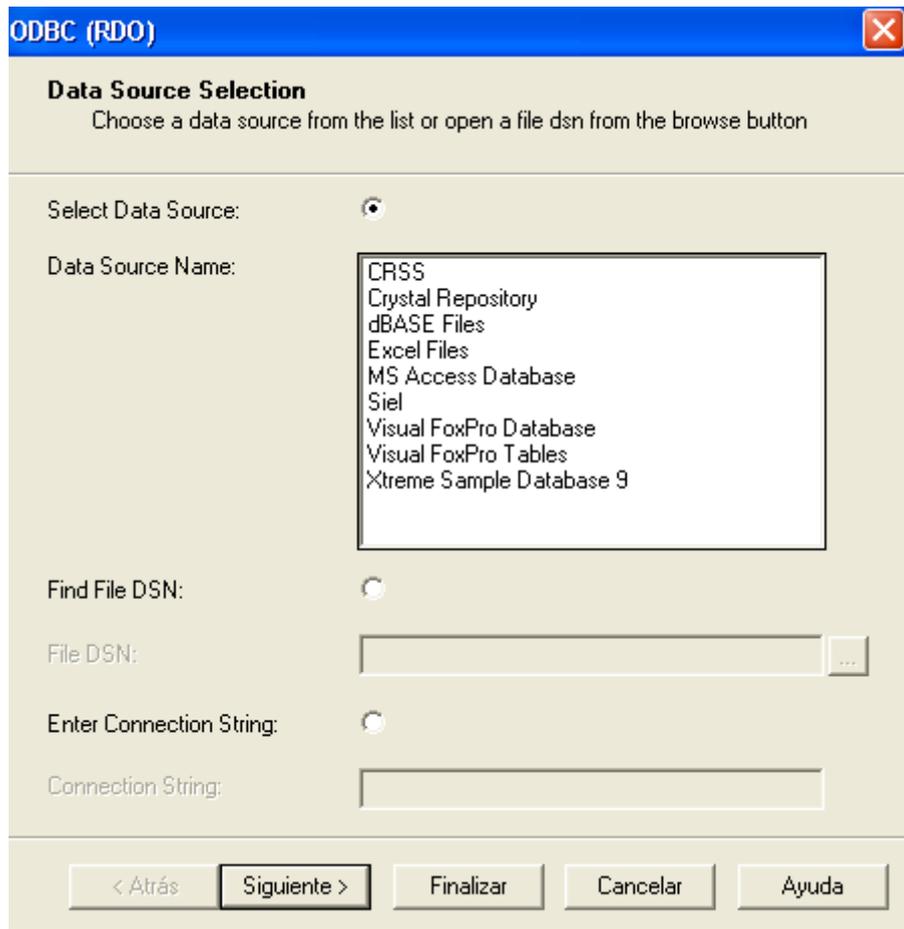
El formato necesario para este reporte es del tipo tabla cruzada, en donde los renglones serán los estados de la República mexicana y las columnas los Instrumentos financieros. De las opciones de formatos de reportes, se elige el de tipo Cross-Tab (Ver figura 4.1).

<sup>11</sup> Crystal Reports.- Reporteador propiedad de Business Objects.



**Figura 4.1 Selección del tipo de reporte**

Después se selecciona el nombre del DNS definido para el sistema, en este caso Siel, como se puede observar en la figura 4.2.



**Figura 4.2 Selección del DNS para el reporte**

Después de que se ha establecido la conexión a la base de datos, se seleccionan las tablas que se usarán en el reporte. En este caso, la tabla de hechos a usar será la de Avance, se agregarán las dimensiones de tiempo, estado e Instrumento (ver figura 4.3). La dimensión tiempo servirá para establecer el rango de fechas que se requiere en el filtro de la información, la dimensión estado para los renglones y la dimensión instrumento para las columnas.

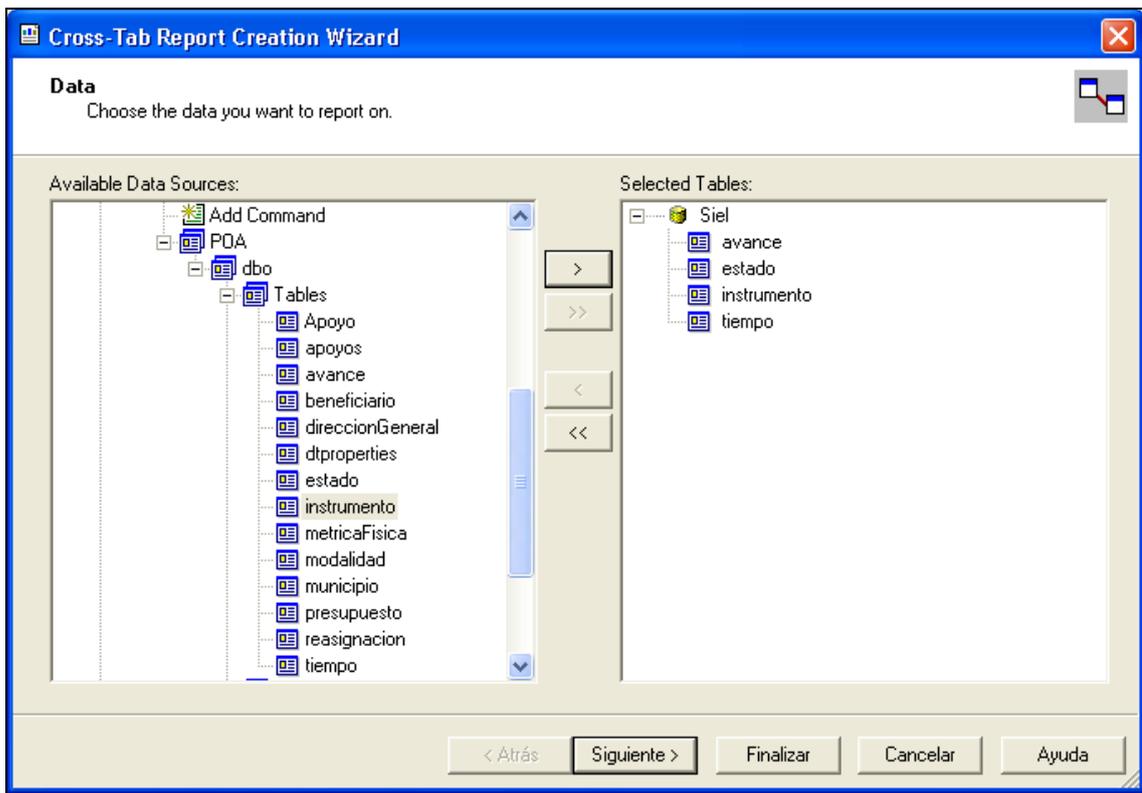
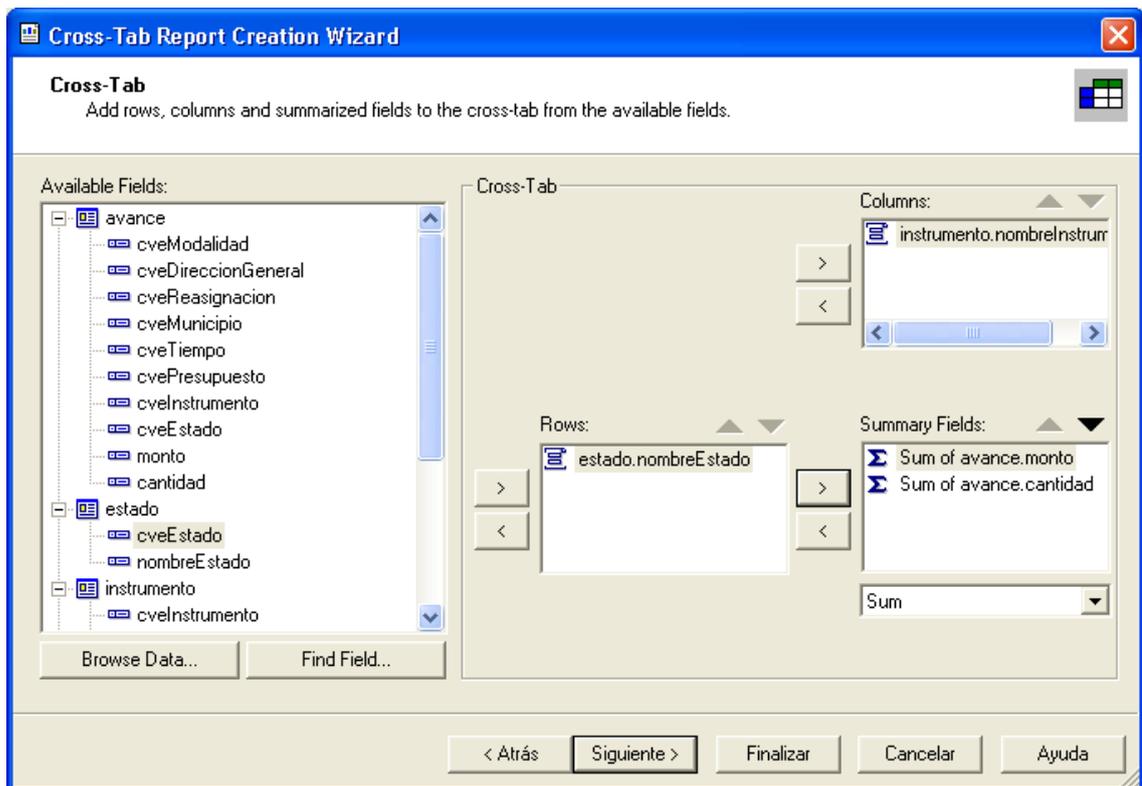


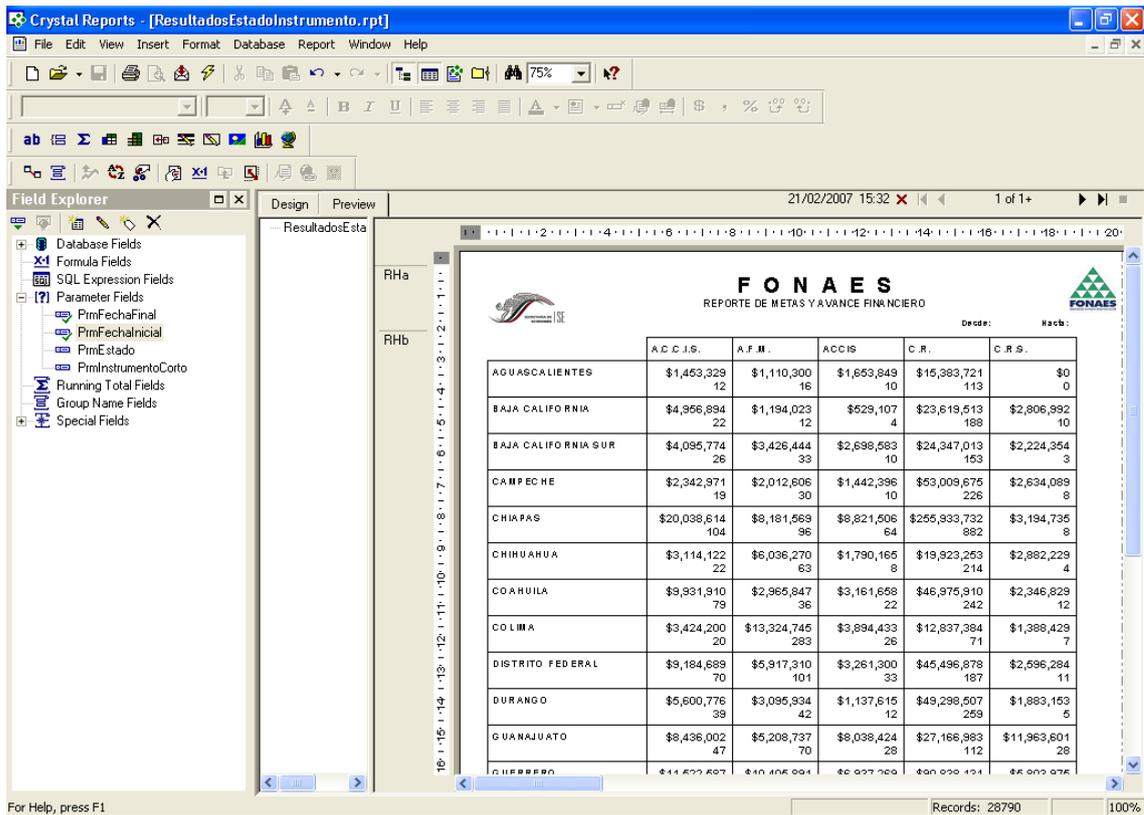
Figura 4.3 Selección de tablas para el reporte

La herramienta permite entonces definir los campos que formarán las columnas, renglones y campos a agrupar como se puede observar en la figura 4.4. Se define entonces el campo de **nombreEstado** de la dimensión **estado** para los renglones y el campo **nombreInstrumento** de la dimensión **instrumento**, y como campos de agrupación el campo de cantidad e inversión de la tabla de hechos Avance.



**Figura 4.4 Selección de campos para el reporte**

Para terminar con el desarrollo del reporte, se ajusta el encabezado con los logos de la Institución y campos especiales, se agregan los parámetros de entrada y los filtros del mismo. El reporte terminado en la herramienta de desarrollo se puede ver en la figura 4.5.



**Figura 4.5 Reporte de avance financiera desarrollado**

Una vez que se ha concluido con el desarrollo del reporte, se sigue con la publicación del mismo en el Siel. Para esto, se define un nuevo grupo de reportes que llamaremos Programa Operativo Anual (POA) en el módulo de reportes. Gracias a que el sistema está preparado para la publicación de reportes de manera parametrizable, se agrega el grupo en el catálogo de módulos en la base de datos como se puede ver en la opción once de la figura 4.6.

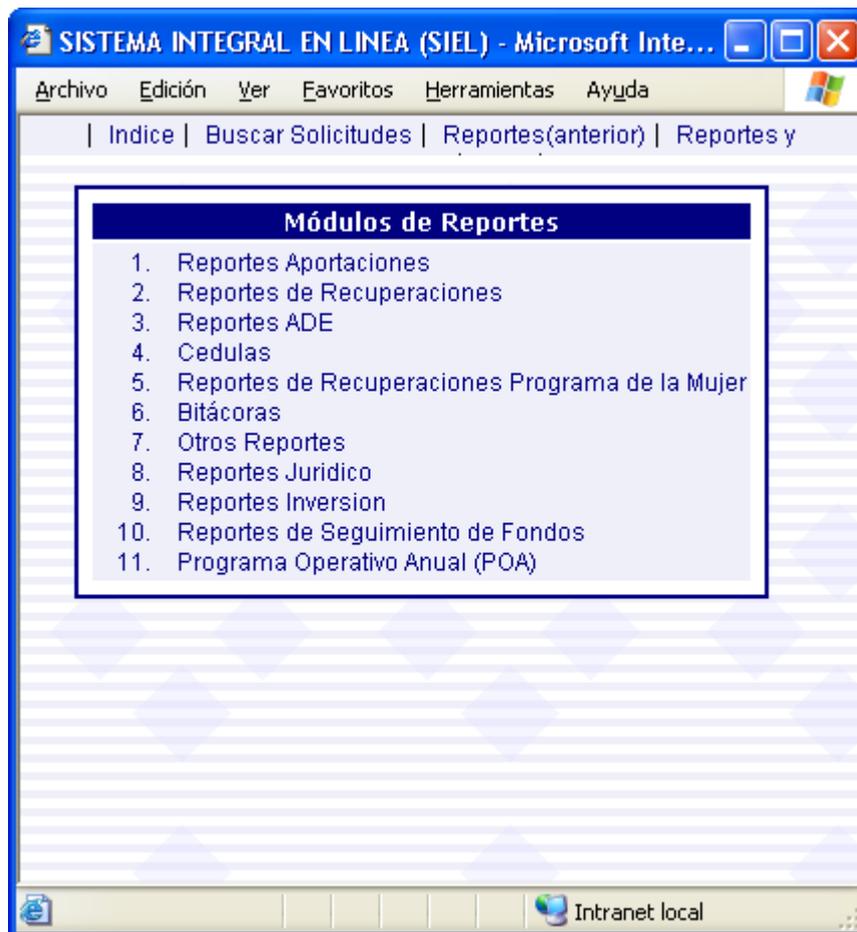


Figura 4.6 Grupos de reportes en el módulo de Reportes

Para publicar el reporte en el sistema, se da de alta en la base de datos, en donde también se asigna en el grupo de reportes correspondientes, como se puede ver en la figura 4.7.



Figura 4.7 Liga del reporte de avance financiero en el Siel

Así cómo en el catálogo de reportes se definió el nombre físico de reporte, también se definió el título del mismo y los parámetros de entrada. Esto facilita el trabajo de la publicación, ya que no es necesario programar ya nada en el sistema. Como se puede ver en la siguiente figura (4.8), los parámetros de entrada se despliegan en la pantalla de filtros del reporte.

SISTEMA INTEGRAL EN LINEA (SIEL) - Microsoft Internet Explorer

### Opciones de filtrado para el reporte de Reporte de metas y avance financiero

Estado:

Instrumento financiero:

Fecha Inicio:

Fecha Final:

Figura 4.8 Parámetros de entrada para el reporte de avance financiero

Después de haber dado de alta el reporte en el sistema, al igual que el nuevo grupo y los parámetros de entrada, el reporte desarrollado se coloca en el servidor del IIS<sup>12</sup> en donde se encuentra el Siel. Es entonces que el Report Application Server <sup>©</sup><sup>13</sup> se encarga de desplegar el reporte en el sistema. La parte de programación para la publicación ya se encuentra en el módulo, esto permite que cada vez que se integre un nuevo reporte sólo sea necesario integrarlo en la base de datos del sistema. Ver figura 4.9.

<sup>12</sup> IIS.- Internet Information Server. Servidor web desarrollado por Microsoft®.

<sup>13</sup> Report Application Server.- Servidor de publicación de reportes en web de Bussiness Object®.

SISTEMA INTEGRAL EN LINEA (SIEL) - Microsoft Internet Explorer

powered by crystal

Vista previa

**FONAES**  
**REPORTE DE METAS Y AVANCE FINANCIERO**

	A.C.C.I.S.	A.F.M.	C.R.	C.R.S.	C.S.
AGUASCALIENTES	\$1,453,329 12	\$1,110,300 16	\$4,309,800 17	\$0 0	
BAJA CALIFORNIA	\$4,956,894 22	\$1,194,023 12	\$6,218,099 20	\$2,806,992 10	
BAJA CALIFORNIA SUR	\$3,895,774 25	\$3,426,444 33	\$8,123,741 43	\$2,224,354 3	
CAMPECHE	\$2,042,971 18	\$2,012,606 30	\$15,586,707 49	\$2,634,089 8	\$6,900
CHIAPAS	\$20,038,614 104	\$8,181,569 96	\$52,006,251 173	\$3,194,735 8	\$28,450
CHIHUAHUA	\$3,064,122 19	\$6,036,270 63	\$2,952,105 45	\$2,882,229 4	\$5,300
COAHUILA	\$9,821,346	\$2,965,847	\$12,938,729	\$2,346,829	

**Figura 4.9** Reporte ya publicado en el Siel.

#### 4.2.2. Reporte de metas y avance financiero

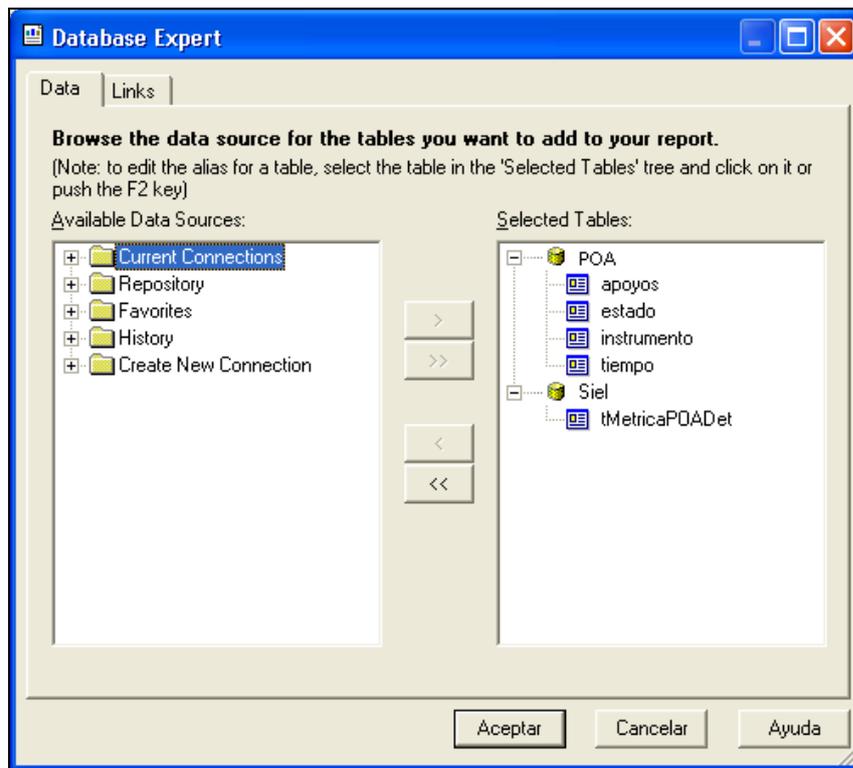
Este reporte le permitirá al usuario final comparar los avances que se tienen en la asignación de recursos por estado e instrumento financiero. Le dará la información necesaria para ir monitoreando el trabajo que en cierto período ha realizado la Institución.

Definición de reporte	
Nombre del reporte:	Reporte de metas y avance financiero
Descripción:	Este reporte servirá como base de medición de la productividad de la Institución, estas metas serán definidas por la DGPAT <sup>14</sup> .
Frecuencia de uso	Mensualmente y cada vez que se requiera
Parámetros de entrada:	Un rango de fechas, entidad Federativa e Instrumento
Cálculos:	Las metas serán capturadas desde el sistema en línea SIEL.
Notas:	Se mostrará en pantalla la meta de inversión y la inversión que hasta la fecha haya asignado la Institución

**Tabla 4.9 Definición de reporte de metas y avance financiero.**

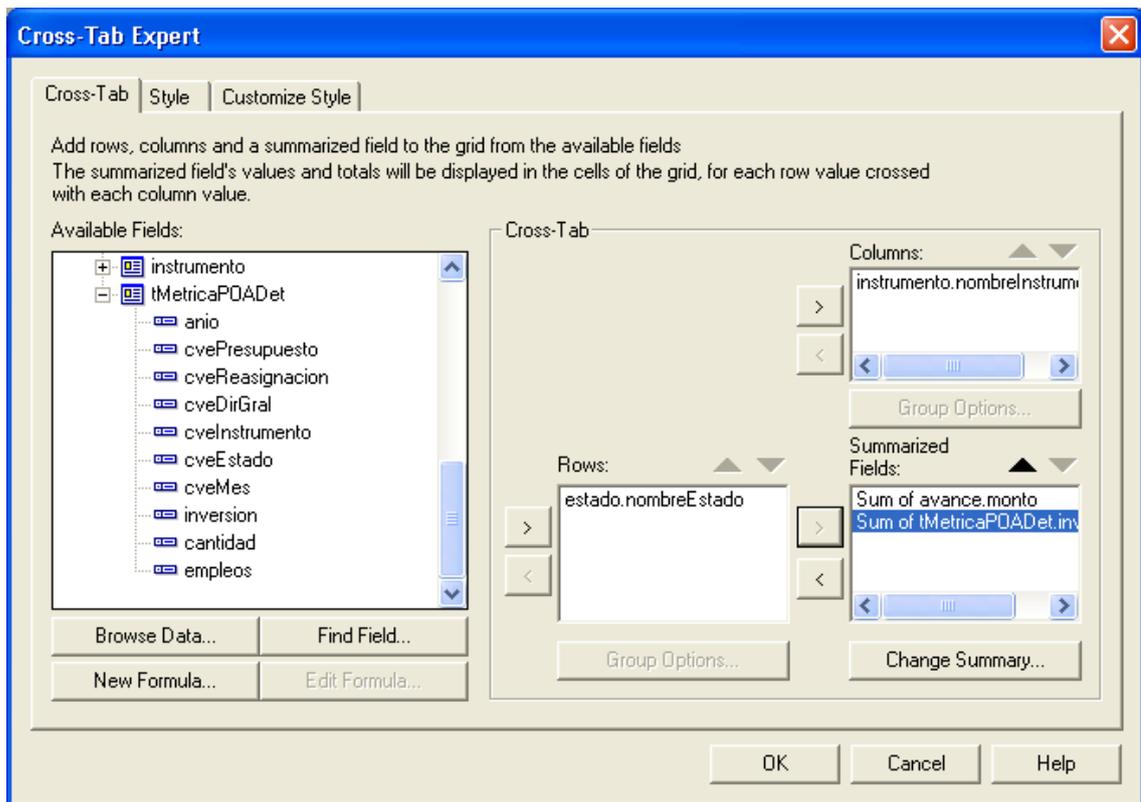
Para este reporte se utilizó la Base de datos SIEL de donde toma las metas financieras capturadas por la DGPAT (métricas) y los avances reales del Data mart . Ambas conexiones se pueden ver en la figura 4.10. Para este reporte se tomaron las tablas de Avance, tiempo, instrumento y estado de la base de datos POA y la tabla de tMetricaPOADet de la base de datos Siel.

<sup>14</sup> DGPAT.- Dirección General de Planeación y Asistencia Técnica, área usuaria principal.



**Figura 4.10 Definición de conexiones a Bases de Datos**

En la figura 4.11 se puede observar la definición del reporte, en donde se definen como renglones los estados y los instrumentos como columnas, en este caso los campos a agrupar serán la meta de inversión, el reporte ya con la comparación entre la meta definida por la DGPAT y los avances reales de la Institución. De esta manera se podría evaluar los avances cada vez que sea requerido, para que la Institución tome medidas correctivas de ser necesario y llegar así, a las metas establecidas.



**Figura 4.11 Definición de campos a incluir en el reporte.**

Una vez que se ha terminado con el desarrollo del reporte, con los logotipos de la institución, los parámetros de entrada, el filtro de información y los títulos, se sigue entonces con el proceso de publicación. Como vimos en el reporte anterior, sólo es necesario agregarlo en las tablas parametrizables del sistema. En la figura 4.12 se pueden observar los parámetros de entrada que son requeridos al usuario final en el sistema Siel.

**Opciones de filtrado para el reporte de Reporte de metas y avance financiero**

Estado:   
 TODOS  
 AGUASCALIENTES  
 BAJA CALIFORNIA  
 BAJA CALIFORNIA SUR  
 CAMPECHE  
 CHIAPAS

Instrumento financiero:   
 TODOS  
 A.C.C.I.S.  
 A.C.P.  
 A.F.M.  
 A.I.M.  
 ACCIS

Fecha Inicio:

Fecha Final:

Generar Reporte    Cerrar Ventana

Figura 4.12 Parámetros de entrada para el reporte de metas y avance financiero

El reporte final ya publicado en el Siel se puede ver en la figura 4.13, en donde el componente de publicación permitirá al usuario navegar por la información desplegada y también le da la posibilidad de exportarlo a varios formatos como pueden ser Word, PDF, texto, etc.

**FONAES**  
 REPORTE DE AVANCE Y METAS FINANCIERAS

		C.R.	C.T.S. 2003	FDO INV Y REIN	PROC
AGUASCALIENTES	inversión meta	\$0.00	\$751,299.00	\$0.00	
		\$0.00	\$500,000.00	\$0.00	
BAJA CALIFORNIA	inversión meta	\$12,436,198.00	\$5,148,566.00	\$0.00	
		\$6,570,000.00	\$200,000.00	\$0.00	
BAJA CALIFORNIA SUR	inversión meta	\$8,123,741.00	\$0.00	\$0.00	
		\$600,000.00	\$0.00	\$0.00	
CAMPECHE	inversión meta	\$15,586,707.00	\$0.00	\$0.00	
		\$9,000,000.00	\$0.00	\$0.00	
SINALOA	inversión meta	\$0.00	\$0.00	\$11,363,416.00	
		\$0.00	\$0.00	\$0.00	
VERACRUZ	inversión meta	\$0.00	\$0.00	\$0.00	
		\$0.00	\$0.00	\$0.00	
Total	inversión meta	36,146,646.00	5,899,865.00	11,363,416.00	
		\$16,170,000.00	\$700,000.00	\$0.00	

Figura 4.13 Reporte de metas y avance financiero ya publicado en el Siel

## Conclusiones

Business Intelligence (BI) o Inteligencia de Negocios es el conjunto de tecnologías que le sirven a una organización para utilizar la información que posee con el objetivo de hacer análisis, buscar oportunidades de mejora y tomar mejores decisiones. Los almacenes de información son un concepto invaluable de esta disciplina, creando información estructurada y de calidad, que permite a las empresas la mejora de sus procesos mediante su análisis.

La parte de diseño de un almacén de datos difiere de los sistemas operacionales, así como también del levantamiento de requerimientos del área usuaria, que en la mayoría de los casos son altos ejecutivos. El proyecto me permitió involucrarme en este tipo de sistemas, en un proceso de investigación que fue más allá de lo que conocía, permitiéndome también continuar con mi actualización profesional. Es así, como implementando los conocimientos que adquirí en mi carrera profesional, experiencia profesional y trabajo de investigación, llegué a la realización de este proyecto. Propuse una solución práctica minimizando los costos en software, arquitectura y recursos humanos. El software que se usó fue el mismo con el que ya se contaba en la Organización, usando también el sistema en línea Siel como base para la publicación y explotación de la información.

La solución creada para las necesidades del área usuaria tuvo las siguientes ventajas:

- Tiempo considerablemente corto en el desarrollo e implementación. Esto, debido a que se usó el módulo ya desarrollado para la publicación de reportes.
- Flexibilidad para modificar las métricas de comparación, gracias a que éstas se encuentran en el sistema Siel en un módulo para el área usuaria.
- Se comprobó también que la extracción de la información del Data mart no significó ningún retraso en la respuesta del sistema operacional, ya que los reportes consultan de la base de datos operativa sólo los indicadores y métricas POA.

Como es bien sabido, lo más importante en cualquier empresa, su activo principal, es la información que ésta genera y de la que se vale para la operación de

sus procesos. A medida que el data mart vaya siendo afinado en cuestión de calidad de la información, a través de revisiones iterativas por parte del área usuaria y se vaya ingresando la información que está rezagada por las oficinas regionales, será entonces una gran herramienta que servirá para identificar muchos otros indicadores además de los definidos en los requerimientos iniciales. Algunos de ellos podrían ser organizaciones con problemas de cartera vencida, con personas de la tercera edad, con deficientes procesos de recuperación, etc. La explotación eficiente de esta información dependerá entonces de involucrar al área usuaria con el área de informática, que deberá ayudarle a sacar el mayor provecho de la solución generada.

Partiendo como inicio, este proyecto de almacén de datos, se podría mejorar con otros data marts o creando una solución mixta de Data warehouse en dónde se podría incluir información de diferentes áreas de la Organización como recursos humanos, contabilidad, finanzas, empresas sociales, jurídico, etc. Si se logrará incluir la información de los empleados por oficina regional, se podría medir la productividad por empleado, o medir el desempeño en relación al salario o al tiempo trabajado. Manejar algún programa de estímulos para la mejora del desempeño y productividad. Si se agregará la información de recursos materiales, se podrían identificar también las áreas de la empresa en donde se hace un mayor gasto en recursos y significar un notable ahorro en la operación de la Organización.

Las herramientas y la información están ahí, el usarlas de la mejor manera depende de las personas que están involucradas en el área tecnológica y de las que toman decisiones para crear ventajas competitivas en esta y cualquier otra Organización. El desaprovechar la información que esta disponible significa perder infinidad de oportunidades de mejora en un mundo cada vez más cambiante, en donde sólo sobrevive quien se adelanta a su competidor. Este nuevo siglo XXI será el siglo de la Información, y como profesionales de la informática es crucial que aprendamos a aprovecharla.

## Referencias Bibliográficas

KIMBALL Ralph, Reeves Laura, Ross Margy, Thornthwaite Warren  
“The Data Warehouse Lifecycle Toolkit”  
Wiley Computer Publishing, 1998

BAIN Tony, Benkivich Mike, Dewson Robin, Ferguson Sam, Graves Christopher,  
Joubert Terrence J., et al.  
“Professional SQL Server 2000 Data Warehousing with Analysis Services”  
Wrox Press Ltd., 2000

CASARES, Claudio  
“Data Warehousing”  
<http://www.programacion.net/autor/28/>

ALEJANDRE González, Bruno  
“Minería de Datos y Datawarehouse”  
Centro Tecnológico Aragón

THOMAS J. Kelly  
“Dimensional Data Modeling”  
Sybase

INMON, W.H.  
“Building the Data Warehouse”  
PaperBack

WOLFF Carmen Gloria  
“Modelamiento multidimensional”  
<http://www.inf.udec.cl/revista/edicion4/cwolff.htm>

WOLFF Carmen Gloria  
“La Tecnología Datawarehousing”  
<http://www.inf.udec.cl/revista/edicion3/cwolff.htm>

SAS® Institute Inc.  
SAS® Rapad Warehousing Methodology  
2001

GUISAO Cartagena Carlos Mauricio  
“Bodega de Datos”  
<http://www.monografias.com/trabajos24/bodega-de-datos/bodega-de-datos.shtml#uso>

BORJA, Ronda  
“La empresa Multidimensional:OLAP”  
<http://www.idg.es/iworld/articulo.asp?id=143456>

PADAMPREET Singh Wadhwa, Prakash Kamalapur  
“Customized Metadata Solution for a Data Warehouse – A Success Story”  
WIPRO

FIGUEROA Fernando, Shanton Luciano, Turco Pablo  
"Data Warehouse"  
Universidad Tecnológica Nacional

SAS®, Institute Inc.  
"The SAS Methodology for Solutions Development"  
2003