

## GENERALIDADES Y PATRONES DE DESARROLLO DE DATA MARTS

*Kenneth Sánchez Sánchez\**  
*Rafael Martínez V.\*\**

### RESUMEN

El desarrollo de los *Data marts* en el ambiente global de implementación de *Data warehouse* en una empresa, está relacionado a distintos patrones de desarrollo generales ó alternativos que existen. En este documento se describe las características generales propias de los *Data mart*, los distintos modelos de desarrollo tanto generales como alternativos y además se presenta una explicación sobre las posibilidades de desarrollar *Data mart* versus *Data warehouse* o viceversa.

**Palabras clave:** *Data mart*, *Data warehouse*, modelos de desarrollo

### ABSTRACT

The development of Data marts in the global environment of implementation of Data Warehouse in a company is related to different general or alternative development existing patterns. In this document the general characteristics typical of Data marts as well as the different general and alternative development models are described. Also, an explanation about the possibilities of developing Data marts versus Data warehouse or vice versa is presented.

**Key words:** *Data mart*, *Data warehouse*, Development Models

---

\* Universidad de Costa Rica - Sede de Guanacaste.  
kenneths@cariari.ucr.ac.cr

\*\* Universidad de Costa Rica - Sede de Guanacaste.  
ramartin@cariari.ucr.ac.cr

## Introducción

La manera de desarrollar e implementar *Data marts* (bodegas de datos a nivel departamental) en una empresa, está relacionada con los requerimientos específicos de los distintos análisis que se quiere realizar en cada departamento, así como la estructura general que ésta posea. Existen tres modelos de desarrollo generales, *Top Down*, *Bottom Up* y el modelo Paralelo además se cuenta con los modelos de desarrollo alternativos.

En este artículo se describirán las generalidades de los *Data marts*, los diversos patrones o modelos de desarrollo que existen tanto generales como alternativos que utilizan la retroalimentación, y por último se detallan algunas razones que se deben de tomar en cuenta al momento de desarrollar e implementar los *Data marts* ante la *Data warehouse* o en forma contraria.

## Generalidades de los *Data marts*

Para poder acercarse al tema, se necesita establecer claramente los objetivos y fines que se persiguen con la construcción y la puesta en marcha, no solo de los *Data marts* sino de la *Data warehouse*, de la cual es necesario mencionar que no es un producto construido con estándares que se puede aplicar indistintamente del tipo de empresa que desea adquirirlo a través de una suma de dinero, sino más bien consiste en un proceso que debe ser obligatoriamente construido, evolutivo y duradero en el tiempo para que la empresa pueda obtener los resultados o metas que se propuso con su implementación [Map1].

Unido al trabajo que se pretende alcanzar con la *Data warehouse*, debe mencionarse que los *Data marts* son bodegas de datos con información de interés particular para un determinado sector de la empresa y aunque su enfoque sea para una sola perspectiva departamental, esto no lo exime de tener que seguir los lineamientos generales de implementación que posee la *Data warehouse*; éstos son: [Kimball]

- Recolección y análisis de requerimientos.
- Creación del modelo de datos y su diseño físico.
- Definición de los orígenes de los datos.
- Selección de la tecnología de base de datos y hardware a utilizar.
- Extracción de los datos desde sistemas operacionales, su limpieza, transformación y carga a la *Data warehouse*, en este caso específico al *Data mart*.
- Selección de las formas de acceso, herramientas de análisis, reporte y presentación.
- Desarrollo de los reportes y aplicaciones necesarias.
- Actualización del *Data mart*.

Todos estos lineamientos se derivan de la implementación de las *Data warehouse* pero debido a la relación estructural que existe con el diseño de los *Data marts*, permite que se empleen en ellos también.

Pese a esta relación, los *Data mart* no se pueden considerar como una *Data warehouse* en escala inferior, ya que ellos están diseñados para satisfacer las necesidades específicas de los departamentos o divisiones en las empresas esto permite asegurar que sin lugar a duda los *Data marts* utilizan un planteamiento de

“divide y vencerás”, que a menudo es la solución, cuando la *Data warehouse* crece desmedidamente, a tal punto que se hace incontrolable su operación [Gartner].

Ante esta característica los *Data marts* tienen ventaja en comparación con las *Data warehouse*, en puntos como el campo de acción, la perspectiva de los datos, fuentes de los datos, tiempo de implantación, espacio de almacenamiento, cantidad de datos, tiempo de consultas, esto tomándolo desde una perspectiva global. En términos específicos, se puede mencionar entre otras, que los *Data marts* requieren de una aplicación específica al estar enfocados a una sola área de la empresa, esto permite que la estructura de su información esté altamente detallada, las consultas se realizan en forma rápida, su proceso al estar orientado a una sola área puede usarse con medidas específicas de la empresa como mercadeo de un producto específico, proyección de ventas o promociones, y por último su tiempo de implantación requiere entre 4 y 12 meses [Gartner]. Estas ventajas permiten que el trabajo con los *Data marts* desde un punto de vista general sea muy provechoso, pero ante este panorama se debe conocer que para disfrutar de estas ventajas el *Data mart* por si solo no se constituyó en la empresa, sino que debió pasar por un proceso de diseño e implementación, en el cual no se puede dejar por fuera como punto de referencia la *Data warehouse*.

En este proceso de diseño y construcción de los *Data marts*, existen diversos patrones de desarrollo, entre los cuales podemos mencionar tres: el primero conocido como *Top Down* el cual tiene como base un sistema de *Data warehouse* para toda la empresa y a partir de este se desarrollan los *Data marts* para las divisiones o departamentos. Un segundo patrón denominado *Bottom - up*, el cual

se fundamenta en la construcción de los *Data marts* de cada departamento o división de la empresa, a partir de los cuales se construye la *Data warehouse*; y un tercer modelo, el cual consiste en llevar una construcción paralela entre ambas estructuras, sin definir como padre la *Data warehouse* sobre el *Data mart*, ni viceversa.

Dentro de esta categorización de los modelos, existen tres variaciones, una para cada uno de ellos, dicha variación consiste en incluir en cada patrón de desarrollo el uso de retroalimentación de datos.

Los *Data marts* por las consideraciones mencionadas, vienen a ser una excelente herramienta de análisis de datos y soporte para la toma de decisiones para las pequeñas y medianas empresas, por su versatilidad, corto tiempo de desarrollo y bajo costo económico, así como la obtención de los resultados esperados a un corto plazo. Para tener un panorama general sobre los *Data marts*, es preciso pasar a especificar de forma clara y detallada cada uno de los tres modelos de patrones de desarrollo así como sus variaciones.

## Modelos generales de desarrollo

En esta categoría de modelos no es tomada en consideración la retroalimentación, la cual permite agregar datos que no estén presentes en la *Data warehouse*.

Los tipos de modelos son los siguientes: el modelo *Top Down*, el *Bottom Up* y el Paralelo.

### Modelo *Top Down*

El modelo *Top Down* está basado en la estructura de la *Data warehouse*, la cual se construye a partir de los datos que se puedan obtener de los diferentes sistemas

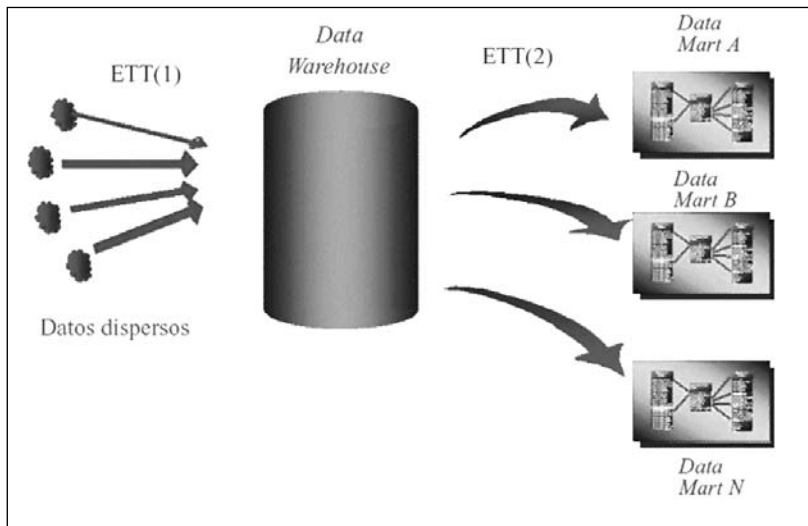
operacionales o externos (datos aislados) a través de un proceso de extracción, transformación y transportación (ETT) [Firestone].

Todos los datos necesarios para el apoyo a la toma de decisiones se encuentran en la *Data warehouse* después de que la *Data warehouse* está implementada con

la información procesada, solamente es necesario distribuir los datos según las necesidades de información por departamento, generando los *Data marts* como subgrupos de datos específicos para los requerimientos particulares de cada departamento, La representación gráfica se presenta en la figura 1.

Figura 1

**Modelo Top Down [Firestone]**



Es importante hacer notar que los *Data marts* en este modelo son derivados de la *Data warehouse* y debido a esta estrecha relación se da una inconsistencia al querer definir una nueva necesidad de información para un departamento, ya que en este caso es necesario modificar primero la *Data warehouse* para que ocurra el respectivo cambio en el *Data mart*.

En contraparte al proceso de construcción de este modelo se presenta el modelo de desarrollo *Bottom Up*.

**Modelo Bottom up**

En este modelo los *Data marts* se construyen a partir de los datos dispersos y la

*Data warehouse* se construye a partir de los *Data marts* existentes, esta construcción se realiza a través de dos procesos diferentes de extracción, transformación y transportación.

En el primer proceso cada *Data mart* se construye con los datos aislados que son necesarios para satisfacer las funciones del departamento por medio de los procesos de extracción, transformación y transportación. Es necesario enfatizar que por la naturaleza de su diseño no existe ninguna relación entre un *Data mart* y cualquier otro.

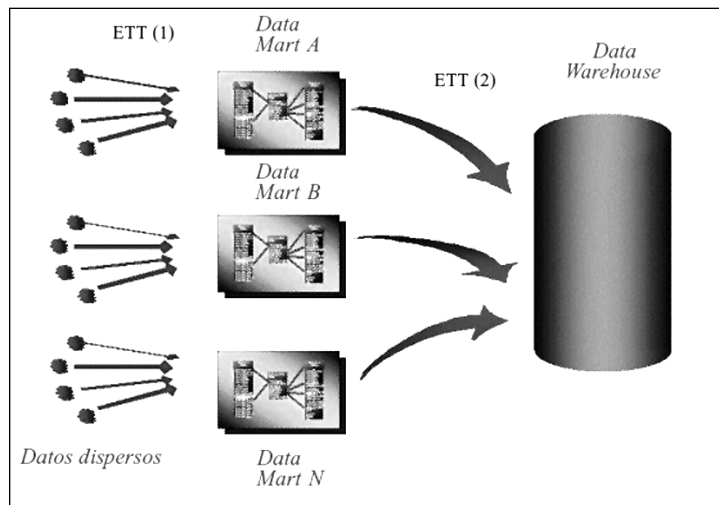
Un segundo proceso de tratamiento de datos (ETT) ocurre en el sentido de los *Data marts* hacia la *Data warehouse*,

característica representada en la Figura 2. Esta *Data warehouse* que ha sido construida a través de los diversos *Data marts*, contendrá toda la información que la empresa requiera de acuerdo a la

necesidad o análisis que quiera realizar [Gartner]. A parte de este modelo existe el modelo Paralelo, cuyo diseño está enfocado en la construcción de los *Data marts* y la *Data warehouse* de forma simultanea.

Figura 2

Modelo Bottom up [Firestone]



**Modelo Paralel**

El diseño del modelo paralelo se basa en dos alternativas, en la primera se tratan los *Data marts* con entidades independientes de los *Data waehouse* y en la segunda, esta independencia se trata de forma temporal.

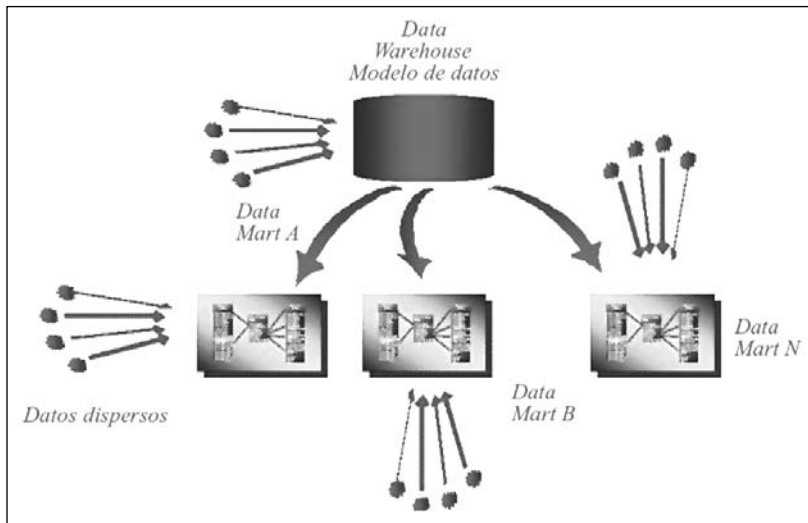
En el primer caso se enfoca la construcción de los *Data marts* como entidades independientes de la *Data warehouse*, utilizando el modelo de construcción de esta. Esto consiste en utilizar el mismo modelos de datos que se emplea en la construcción de la *Data warehouse* propuesto por la empresa. La ventaja de que este modelo funcione de esta forma es

que permite detectar y controlar problemas, como la falta y la redundancia de información presentes en el modelo de la *Data warehouse*. Esto permite el mejoramiento del modelo de construcción en los futuros *Data marts* y en la misma *Data Warehouse*.

Por otra parte este modelo puede verse desde otro punto de vista donde los *Data marts* se construyen con una independencia temporal de la *Data warehouse*, lo que indica que una vez que están implementados pasan a ser parte de ella, como un subconjunto de datos que conforma la *Data warehouse* que existe en la empresa [Firestone]. El diseño de este modelo está representado en la figura 3.

Figura 3

## Modelo paralelo [Firestone]



Es preciso recordar que dentro de la clasificación de los modelos ya descritos, existen tres modelos alternativos, que incluyen la retroalimentación de datos.

### Modelos de desarrollo con retroalimentación

Estos modelos de desarrollo son los mismos expuestos con anterioridad, su diferencia radica en la incorporación de la retroalimentación de datos en cada uno de ellos, a continuación se describe en detalle cada uno de estos modelos: Modelo *Top down* con Retroalimentación, Modelo *Bottom up* con Retroalimentación y el Modelo Paralelo con Retroalimentación.

#### Modelo *Top down* con retroalimentación

Este modelo incluye la posibilidad de agregar datos que no estén presentes en la *Data warehouse* y que se requieran para cumplir con los requerimientos de

un análisis específico. El uso de la retroalimentación consiste en incluir nuevos datos a los *Data marts* de la empresa, como una primera fase (ver figura 4) y consecuentemente estos cambios ocurridos en el *Data mart* sean integrados a la *Data warehouse* como la segunda fase.

El segundo modelo cuyo rol es diferente se presenta en el modelo *Bottom up* con retroalimentación.

#### Modelo *Bottom up* con retroalimentación

El flujo de retroalimentación de datos en el modelo *Bottom up*, afecta inicialmente a la *Data warehouse*, ya que la retroalimentación se da en los *Data marts* (ver figura 5), presentándose el problema entre los requerimientos que tengan los *Data marts* y lo que en realidad contiene la *Data warehouse*. Aunque esta situación puede abrir una puerta para que se lleve a cabo un trabajo parecido al que se realiza en el modelo *Top down*, no significa que éste deba de emplearse, sino más bien el modelo

Figura 4

Modelo *Top down* con retroalimentación [Firestone]

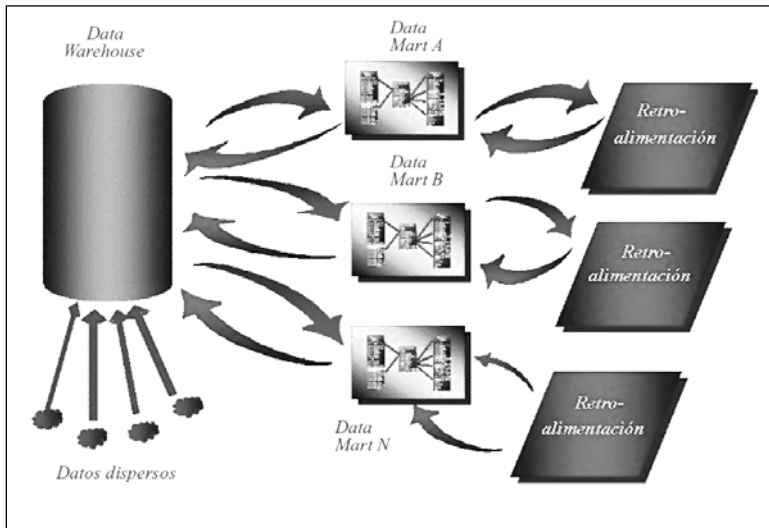
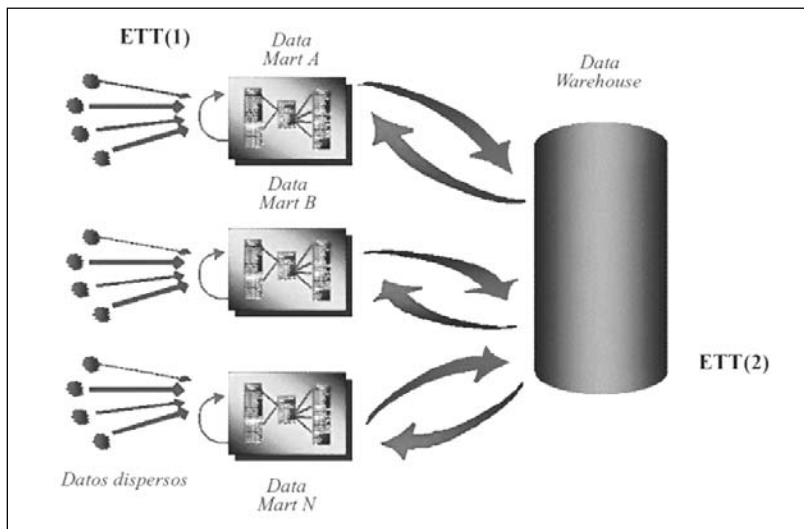


Figura 5

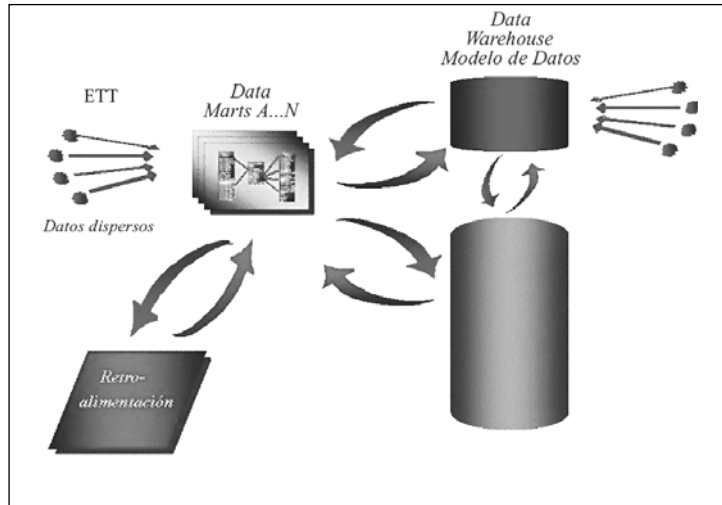
Modelo *Bottom up* con retroalimentación [Firestone]



*Bottom up* con retroalimentación puede obviar esta característica y permitir que la integración de los datos e información dada en los *Data marts* se pueda implementar durante la construcción de la *Data warehouse*.

Es necesario notar que aunque estos modelos alternativos con retroalimentación presentan un trabajo similar dado a sus características, existe el tercer modelo alternativo paralelo con retroalimentación, que se presenta en la figura 6.

Figura 6

**Modelo Paralelo con Retroalimentación [Firestone]****Modelo Paralelo con retroalimentación**

En este modelo el proceso de implementación es un poco más complicado, debido a las relaciones directas y paralelas que existen entre los *Data marts*, la *Data warehouse* y el modelo de datos de la *Data warehouse*, ya que este patrón de desarrollo asume que el modelo de datos de la *Data warehouse* está construido antes de que un *Data mart* inicie su implementación.

El desarrollo principalmente trabaja con la retroalimentación que tenga el *Data mart*, ya que su inicio depende de un período de ajuste entre éste y el modelo de datos de la *Data warehouse*. Mientras las entradas y salidas se estén dando en los *Data marts*, en la *Data warehouse* estos mismos se están realizando, dado a las características de paralelismo que existe entre ellos. Es importante denotar, que el modelo de datos de la *Data warehouse* evidentemente se actualiza de los *Data marts*, puesto que este modelo de datos es el que contiene todas las perspectivas y

funciones que se pretendan realizar con este patrón.

En la figura 6 se muestra el proceso de desarrollo que se realiza cuando se implementa el modelo paralelo con retroalimentación.

Con los distintos patrones de desarrollo expuestos hasta aquí, se debe notar la relación tan estrecha que existe, en cuanto a diseño se refiere, entre la *Data Warehouse* y los *Data marts*, aunque exista esta relación, la decisión de desarrollar una antes que la otra está ligada a diversos factores dependientes de la empresa en donde se quieran implementar [Firestone].

**Desarrollo de Data Marts vrs Data Warehouse**

Las razones para desarrollar en una empresa *Data warehouse* antes de *Data mart* o viceversa, dependen de factores tales como la naturaleza de la organización,



tamaño, requerimientos, estructuración y distribución geográfica de la misma. Por este motivo no existe un planteamiento estandarizado, que permita aplicarlo, indistintamente del tipo de empresa que lo requiera [Inmon1].

El diseñador de los *Data marts* o *Data warehouse* debe tener presente algunos aspectos que hacen la diferencia entre estos dos planteamientos:

- Los *Data marts* son creados para satisfacer las necesidades específicas de un departamento de acuerdo a los objetivos de los mismos. La *Data warehouse* se crean para satisfacer las necesidades globales de una corporación, pero no se puede satisfacer las necesidades de ambos.
- La granularidad entre un planteamiento y otro es muy diferente; los *Data marts* están constituidos por datos resumidos o agregados, por otro lado los datos en la *Data warehouse* mantienen

más detalle sobre los datos de la empresa, por esto resulta muy complejo pasar datos de los

- *Data marts* hacia la *Data warehouse*.
- Los datos contenidos en los *Data marts* son más recientes históricamente que los contenidos en la *Data warehouse*.
- Las relaciones en cuanto a temas en los *Data marts* no son las mismas que las que se pueden encontrar en la *Data warehouse*.
- Los tipos de consultas realizadas en los *Data marts* son muy diferentes a las realizadas en la *Data warehouse*.
- En los *Data marts* los usuarios son recolectores de información, mientras que en la *Data warehouse* los usuarios son exploradores de información.

Para hacer más claro el panorama general de las diferencias que existen entre los dos planteamientos, se presenta la siguiente tabla.

**Tabla 1**

**Comparación, *Data Warehouse Vs. Data Marts* [Gartner]**

Data Warehouse Vs. Data Marts		
Alcance	Data Warehouse	Data Marts
	<ul style="list-style-type: none"> <li>• Aplicación neutral</li> <li>• Centralizado, compartido</li> <li>• Estructurado</li> </ul>	<ul style="list-style-type: none"> <li>• Aplicación específica</li> <li>• Orientado a procesos</li> <li>• Bases de datos múltiples con redundancia de datos</li> </ul>
Perspectivas de Datos	<ul style="list-style-type: none"> <li>• Datos históricos detallados</li> <li>• Algo resumidos</li> <li>• Ligeramente desnormalizados</li> </ul>	<ul style="list-style-type: none"> <li>• Detallado (algo histórico)</li> <li>• Resumido</li> <li>• Altamente desnormalizados</li> </ul>
Temas	<ul style="list-style-type: none"> <li>• Múltiples temas</li> </ul>	<ul style="list-style-type: none"> <li>• Único tema</li> <li>• Múltiples temas</li> <li>• Instantáneos de fuente operacional</li> </ul>
Fuentes de Datos	<ul style="list-style-type: none"> <li>• Muchas</li> <li>• Operacional, datos externos</li> </ul>	<ul style="list-style-type: none"> <li>• Poco</li> <li>• Operacionjal, datos externos</li> <li>• OLTP instantáneas de BDs</li> </ul>
Tiempo de Implementación Características	<ul style="list-style-type: none"> <li>• De 9 a 18 meses, dependiendo de los temas</li> <li>• Flexible, extensible</li> <li>• Duradero</li> <li>• Orientado a datos</li> </ul>	<ul style="list-style-type: none"> <li>• De 4 a 12 meses</li> <li>• Restrictivo, no extensible</li> <li>• Corta Vida</li> <li>• Orientado a proyectos</li> </ul>

Dado lo expuesto anteriormente, el tomar una decisión sobre cual planteamiento seleccionar, es necesario evaluar detenidamente las características presentes en la empresa, relacionarlas con las propiedades de cada estructura y a partir de esta relación decidir cual de ellas utilizar [Datateam].

## Conclusiones

Los *Data marts* son bodegas de datos a nivel departamental que están íntimamente relacionadas con el diseño y construcción de la *Data warehouse*. Para el diseño de los *Data marts* es necesario mantener los mismos lineamientos utilizados en el proceso de diseño y construcción de una *Data warehouse*.

La relación entre los *Data marts* y la *Data warehouse* se da también en los diferentes patrones de desarrollo, tanto en los generales como en los alternativos que presentan los *Data marts*. En los modelos generales se definen: El modelo *Top down*, el modelo *Bottom up*, el modelo Paralelo en relación a los modelos alternativos se describen: El modelo *Top down* con retroalimentación, el modelo *Bottom up* con retroalimentación y modelo Paralelo con retroalimentación.

El modelo *Top down* parte de la visión de construir los *Data marts* de la construcción de la *Data warehouse*.

El proceso de desarrollo del modelo *Bottom up* se fundamenta en la construcción inicial de los *Data marts* para después integrarlos en el diseño y construcción de la bodega de datos.

El modelo Paralelo toma propiedades tanto del modelo *Top down* como el modelo *Bottom up*, ya que su proceso de diseño permite construir paralelamente

los *Data mart* como la *Data warehouse*, tomando en cuenta que el modelo de datos de la *Data warehouse* es utilizado por los *Data marts* en se construcción.

La diferenciación en los modelos alternativos se da con la inclusión del proceso de la retroalimentación en los modelos generales. Este proceso consiste en la posibilidad de agregar nuevos datos, tanto a los *Data marts* como a la *Data warehouse*, según el modelo utilizado.

La decisión a cerca de usar *Data marts* o *Data warehouse* como herramienta estructural para el análisis de los requerimientos en la empresa, no depende de una guía específica, sino más bien de las necesidades y características particulares de cada empresa.

## Bibliografía

- [Datateam] Consulting Data Team. "*Data Mart*". Home page del Consulting Data Team. URL: [http://www.datateam.com.mx/datateam/soluciones/data\\_m/index.html](http://www.datateam.com.mx/datateam/soluciones/data_m/index.html). Mexico 1999.
- [Firestone] Firestone J. "*Data Warehouses and Data Marts: A Dynamic*". Home page de Executive Information Systems, Inc. URL: <http://www.dkms.com/DWDDV.html>. EEUU. Marzo, 1997
- [Gartner] Strange Kevin. "*Data Warehouse Vs. Data Marts*". Home page de Gartner Group. URL: <http://www.gartner.com/webletter/ibmbusint/article4/article4.html>. San Diego, California, 1998.
- [Inmon1] Inmon, B. "*Data Marts "Data Warehouse"*". En *DM Review*, Mayo 1998. Pp. 38- 41.
- [Kimball] Kimball R. "*The Data Warehouse Toolkit*". John Wiley & Sons, 1996.
- [Map1] "*Fases de implantación de un data warehouse*". Home page del ministerio de administración pública. URL: <http://www.map.es/csi/silice/DW223.html>. Madrid, España, 1998.