



Universidad de San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

**Arquitectura técnica del *Data Warehouse*.**

**Evelyn Roxana Torres Catalán**  
Asesorado por Inga. Floriza Ávila

Guatemala, marzo de 2006



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA



FACULTAD DE INGENIERIA

**ARQUITECTURA TÉCNICA DEL *DATA WAREHOUSE***

TRABAJO DE GRADUACION

PRESENTADO A LA JUNTA DIRECTIVA DE LA  
FACULTAD DE INGENIERIA

POR

EVELYN ROXANA TORRES CATALAN

ASESORADO POR INGA. FLORIZA AVILA

AL CONFERIRSELE EL TITULO DE  
**INGENIERO EN CIENCIAS Y SISTEMAS**

GUATEMALA, MARZO DE 2006



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA  
FACULTAD DE INGENIERIA



**NÓMINADE JUNTA DIRECTIVA**

|                   |   |
|-------------------|---|
| <b>DECANO</b>     | <b>Ing. Murphy Olympo Paiz Recinos</b>  |
| <b>VOCAL I</b>    |   |
| <b>VOCAL II</b>   | <b>Lic. Amahán Sánchez Álvarez</b>      |
| <b>VOCAL III</b>  | <b>Ing. Julio David Galicia Celada</b>  |
| <b>VOCAL IV</b>   | <b>Br. Kenneth Issur Estrada Ruiz</b>   |
| <b>VOCAL V</b>    | <b>Br. Elisa Yazminda Vides Leiva</b>   |
| <b>SECRETARIO</b> | <b>Inga. Marcia Ivonne Véliz Vargas</b> |

**TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO**

|                   |  |
|-------------------|--|
| <b>DECANO</b>     | <b>Ing. Murphy Olympo Paiz Recinos</b>       |
| <b>EXAMINADOR</b> | <b>Ing. Marlon Antonio Pérez Turk</b>        |
| <b>EXAMINADOR</b> | <b>ing. Jorge Armín Mazariegos Rabanales</b> |
| <b>EXAMINADOR</b> | <b>Ing. Luis Alberto Vettorazzi España</b>   |
| <b>SECRETARIO</b> | <b>Inga. Marcia Ivonne Véliz Vargas</b>      |



## **HONORABLE TRIBUNAL EXAMINADOR**

Cumpliendo con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

### **ARQUITECTURA TÉCNICA DEL DATA WAREHOUSE,**

tema que me fuera Asignado en la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería con fecha Agosto de 2004.



Evelyn Roxana Torres Catalán



Guatemala, octubre 2005

Ing. Carlos Alfredo Azurdia Morales  
Coordinador Comisión de Trabajos de Graduación  
Escuela de Ciencia y Sistemas  
Facultad de Ingeniería USAC

Estimado Ingeniero:

Por medio de la presente hago de su conocimiento, que he procedido a revisar el trabajo final de graduación titulado: **Arquitectura t écnica d el D ata Warehouse**, elaborado por el estudiante Evelyn Roxana Torres Catalán y de acuerdo a mi criterio, se encuentra concluido y cumple con los objetivos propuestos para su desarrollo.

Agradeciendo de antemano la atención prestada, me suscribo a usted,

Atentamente.

  
Inga. Floriza Avila.  
Asesor





Universidad San Carlos de Guatemala  
Facultad de Ingeniería  
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 23 de Noviembre de 2005

Ingeniero  
**Jorge Armin Mazariegos Rabanales**  
Director de la Escuela de Ingeniería  
En Ciencias y Sistemas

Respetable Ingeniero Mazariegos:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación de la estudiante **EVELYN ROXANA TORRES CATALAN**, titulado: "**ARQUITECTURA TECNICA DEL DATA WAREHOUSE**", y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,

  
**Ing. Carlos Alfredo Azurdia**  
Coordinador de Privados  
y Revisión de Trabajos de Graduación





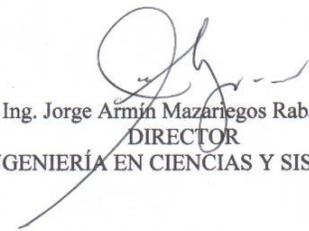
Universidad de San Carlos  
De Guatemala



Facultad de Ingeniería

El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del Revisor y del Licenciado en Letras, del trabajo de graduación titulado **“ARQUITECTURA TÉCNICA DEL DATA WAREHOUSE”**, presentado por el estudiante **EVELYN ROXANA TORRES CATALÁN**, aprueba el presente trabajo y solicita la autorización del mismo.

ID Y ENSEÑAD A TODOS



Ing. Jorge Armin Mazariegos Rabanales  
DIRECTOR  
INGENIERÍA EN CIENCIAS Y SISTEMAS

Guatemala, 24 de febrero de 2005





Universidad de San Carlos  
de Guatemala



Facultad de Ingeniería  
Decanato

Ref. DTG. 056-2006.

El Decano de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela en Ciencias y Sistemas, al trabajo de graduación titulado: **ARQUITECTURA TÉCNICA DEL DATA WAREHOUSE**, presentado por la estudiante universitaria **Evelyn Roxana Torres Catalán** procede a la autorización para la impresión del mismo.

IMPRÍMASE.

Ing. Murphy Olimpo Paiz Recinos  
DECANO



Guatemala, marzo 7 de 2,006

/gdech



## **AGRADECIMIENTOS**

A:

Dios, por darme la sabiduría necesaria para alcanzar esta meta.

Mis padres, Juan Torres y Patricia Catalán de Torres, por su esfuerzo y dedicación, durante todos estos años, no me alcanzara la vida para agradecerse los.

Mis hermanos, Alberto, Lucia y Giovanni, gracias por apoyarme en cada momento

Mi Asesora, Floriza Ávila por la paciencia y su tiempo dedicado, para concluir este trabajo.

Mis Amigos, a todos éxitos en su vida, y en especial a Oscar por compartir todos los momentos de estudio.



# ÍNDICE GENERAL

|  |             |
|--|-------------|
| <b>ÍNDICE DE ILUSTRACIONES .....</b>                                   | <b>III</b>  |
| <b>GLOSARIO .....</b>  | <b>V</b>    |
| <b>RESUMEN .....</b>   | <b>IX</b>   |
| <b>OBJETIVOS.....</b>  | <b>XI</b>   |
| <b>INTRODUCCIÓN .....</b>  | <b>XIII</b> |
| <b>1. EI DATA WAREHOUSE .....</b>                                      | <b>1</b>    |
| 1.1    Introducción .....  | 1           |
| 1.2    Características del <i>Data Warehouse</i> .....                 | 2           |
| 1.3    Diferencias de los ambientes OLTP y <i>Data Warehouse</i> ..... | 4           |
| 1.4    Arquitectura del <i>Data Warehouse</i> .....                    | 6           |
| 1.4.1    Arquitectura Técnica .....                                    | 7           |
| 1.4.2 <i>Back Room</i> .....   | 8           |
| 1.4.3 <i>Front Room</i> .....  | 9           |
| 1.4.4    Infraestructura.....  | 9           |
| 1.4.5 <i>Metadata</i> .....  | 10          |
| 1.5    Arquitecturas del <i>Data Warehouse</i> .....                   | 11          |
| 1.5.1    Básica .....  | 11          |
| 1.5.2    Con un <i>Staging Area</i> .....                              | 12          |
| 1.5.3    Con un <i>Staging Area</i> y <i>Data Marts</i> .....          | 12          |
| <b>2. ARQUITECTURA TÉCNICA.....</b>                                    | <b>15</b>   |
| 2.1 <i>Back Room</i> (Adquisición De La Información).....              | 15          |
| 2.1.1    Sistemas orígenes .....                                       | 17          |
| 2.1.2    Área de Presentación de Datos .....                           | 20          |
| 2.1.3 <i>Metadata</i> del <i>Back Room</i> .....                       | 21          |
| 2.1.4    Servicios del <i>Back Room</i> .....                          | 24          |

|           |  |           |
|-----------|--|-----------|
| 2.1.4.1   | Los Servicios de Extracción .....                                | 24        |
| 2.1.4.2   | Servicios de Transformación .....                                | 26        |
| 2.1.4.3   | Servicios de Carga .....   | 28        |
| 2.1.4.4   | Tareas de Presentación de Datos y Control de los Servicios ..... | 29        |
| 2.1.5     | Procesos de Recuperación del <i>Back Room</i> .....              | 29        |
| 2.2       | <i>Front Room</i> (Acceso a la Información) .....                | 31        |
| 2.2.1     | La <i>Metadata</i> del <i>Front Room</i> .....                   | 32        |
| 2.2.2     | Características Básicas .....                                    | 33        |
| 2.2.3     | Mineo de Datos.....  | 35        |
| 2.2.4     | Servicios .....  | 38        |
| 2.2.4.1   | Servicios para Accesos de Datos .....                            | 38        |
| 2.2.5     | Servicios de Escritorio .....                                    | 41        |
| <b>3.</b> | <b>EL CICLO DE VIDA.....</b>                                     | <b>43</b> |
| 3.1       | Introducción .....   | 43        |
| 3.2       | Plan de Arquitectura .....                                       | 44        |
| 3.3       | Como Iniciamos un Plan de Arquitectura.....                      | 46        |
| 3.4       | Que es la Arquitectura del <i>Data Warehouse</i> .....           | 47        |
| 3.4.1     | Las Estructuras del <i>Data Warehouse</i> .....                  | 50        |
| 3.4.2     | Componentes del <i>Data Warehouse</i> .....                      | 51        |
| <b>4.</b> | <b>HERRAMIENTAS .....</b>  | <b>55</b> |
| 4.1       | Introducción .....   | 55        |
| 4.2       | Warehouse Builder .....  | 56        |
| 4.3       | Cognos .....   | 59        |
| 4.4       | Características principales de una herramienta .....             | 61        |
| 4.5       | Diferencias .....  | 62        |
|           | <b>CONCLUSIONES .....</b>  | <b>67</b> |
|           | <b>RECOMENDACIONES.....</b>                                      | <b>69</b> |
|           | <b>BIBLIOGRAFÍA ELECTRÓNICA.....</b>                             | <b>71</b> |

## ÍNDICE DE ILUSTRACIONES

### FIGURAS

|  |    |
|--|----|
| 1. Nivel General de la Arquitectura Técnica                                    | 6  |
| 2. Arquitectura del Data Warehouse   | 11 |
| 3. Arquitectura del Data Warehouse con <i>Staging Area</i>                     | 12 |
| 4. Arquitectura del Data Warehouse con <i>Staging Area</i> y <i>Data Marts</i> | 13 |
| 5. Arquitectura técnica del <i>Back Room</i>                                   | 16 |
| 6. <i>El Front Room</i> del <i>Data Warehouse</i>                              | 32 |
| 7. Integración de la arquitectura del <i>Data Warehouse</i>                    | 37 |

### TABLAS

|  |    |
|--|----|
| I Diferencias de los ambientes OLTP y <i>Data Warehouse</i>                            | 4  |
| II Necesidades de la información y atributos de los usuario                            | 42 |
| III Criterios a considerar cuando se crea un <i>Data Warehouse</i>                     | 64 |
| IV Características de <i>Warehouse Builder</i> de <i>Oracle</i> vrs. <i>Cognos 8Bi</i> | 65 |



## GLOSARIO

### **B**

#### ***Back Room***

Cuarto de Atrás, parte de la arquitectura técnica del *Data Warehouse* que se encarga del proceso de adquirir los datos

#### **Bases de Datos Multidimensional**

Base de datos diseñada para el proceso analítico en línea. Estructurado como un cubo con un conjunto de tuplas por dimensión.

### **D**

#### ***Data Mart***

Subconjunto del *Data Warehouse* que se extrae para satisfacer las necesidades de una clase particular de usuarios.

#### ***Data Mining***

Es el proceso de extracción de tendencias ocultas en bases de datos de gran tamaño

#### **Data staging Area**

Área del *Data Warehouse*, que almacena los datos que han sido extraídos de los sistemas orígenes, generalmente se encuentran tal y como se han obtenido.

#### ***Data Staging Process***

Proceso que se refiere a la extracción de datos de los sistemas orígenes, para colocarlos en almacenes de datos generalmente temporales.

**Data Staging Services**

Servicios que proveen las herramientas de *Data Warehouse*, para el *Data Staging Process*.

**Data Warehouse**

Sistema de almacenamiento y distribución de cantidades masivas de información.

**Denormalización**

Se refiere al proceso de obviar las reglas de normalización tradicionales definidas para los modelos relacionales

**E  
EIS**

*Enterprise Information System* , Sistemas empresariales de información

**ERP**

*Enterprise Resource Planning*, Planeación de los recursos de la empresa.

**ETL**

*Extraction, Transformation, Loading*, conjunto de métodos que se encuentran involucrados de la extracción, transformación y carga de los datos en un *Data Warehouse*. El orden en que se ejecutan es variante.

**F  
Front Room**

Cuarto de Enfrente, parte de la arquitectura técnica del *Data Warehouse* que se encarga del proceso de presentación de datos al usuario

**G  
Granularidad**

Nivel de detalle de los datos que posee un *Data Warehouse* o un *Data Mart*.

**I  
IS**

*Information System*, Sistemas de información

**M**

***Metadata***

Es la información que se almacena de los datos de un sistema, por ejemplo la descripción de los tipos de una tabla relacional, representa la *Metadata* de la tabla.

**N**  
**Normalización**

Es el proceso en el que se aplican las reglas definidas para la normalización de las bases de datos relacionales

**O**  
**ODS**

*Operational Data Store*, Almacén de Datos operacionales.

**OLAP**

*On-line Analytical processing*. Se refiere a un conjunto de bases de datos multidimensionales que le permiten al usuario ver, navegar y analizar la información contenida en ellas.

**R**  
**Renormalización**

Consiste en reagrupar un modelo que se encuentra denormalizado.

**S**  
**Source Systems**

Sistemas orígenes, lugar de donde se obtienen los datos que alimentan el *Data Warehouse*. Generalmente se califican así a los sistemas operacionales.



## RESUMEN

La arquitectura del *Data Warehouse* está compuesta por infraestructura, arquitectura técnica y *Metadata*. La arquitectura técnica se divide en *Back Room* y *Front Room*, cada uno posee una serie de características como los servicios y sus propios tipos de *Metadata*, ya que, sin ella la arquitectura técnica no se llevaría a cabo, juntas las partes que conforman la arquitectura técnica forman el equilibrio necesario para poder prestar los servicios que el usuario necesita para retraer la información que solicita, acerca de los datos que se encuentran almacenados en el *Data Warehouse*.

Entre los servicios mas notorios del *Front Room* se encuentran las herramientas de *Data Mining* las cuales se encargan de la retracción de datos en base a consultas y la identificación de posibles tendencias ocultas de las empresas, esta herramienta se utiliza como soporte para la toma de decisiones en las empresas.

La parte importante del *Data Warehouse*, se ve representada por el equilibrio que guardan el *Back Room* y el *Front Room*. La diferencia que existe dentro de la teoría y la práctica se observa en la implementación del ciclo de vida que posee cada empresa, dentro de las características mas notoria se encuentra las distintas cargas de datos a la que se puede someter cada herramienta, es por ello necesario identificar las necesidades que se necesitan cubrir, para realizar una correcta evaluación de la misma antes de ser elegida.



# OBJETIVOS

## General

Definir la importancia que tiene la arquitectura del *Data Warehouse* y en especial la arquitectura Técnica.

## Específicos

- Dar un entendimiento general del plan de Arquitectura.
- Describir los niveles de la arquitectura, que posee los *Data Warehouse*.
- Mejorar el entendimiento de la arquitectura técnica, informando y entendiendo para que son necesarios el *Back Room* y el *Front Room*.
- Obtener el conocimiento necesario para realizar un plan de arquitectura técnica.
- Conocer la diferencia entre Arquitectura Técnica e Infraestructura.
- Describir la aplicabilidad de la teoría de la arquitectura técnica, tomando como referencia la herramienta *Warehouse Builder* y *Cognos 8 BI Data Integration*.



## INTRODUCCIÓN

Los *Data Warehouse* son herramientas tecnológicas de gran alcance en el medio, ellos permiten el manejo de grandes volúmenes de información y la consistencia de la información que almacenan. El *Data Warehouse* cuenta con una serie de técnicas procesos y herramientas que lo hacen apto para la presentación de información rápida y consistente. Una de las partes importantes de un *Data Warehouse* es la arquitectura, para poder realizar, efectivamente, la obtención de información, es necesario conocer a profundidad las necesidades que se deben de satisfacer con el diseño y la implementación del sistema.

La arquitectura del *Data Warehouse* cuenta con una serie de eventos que se cumplen al momento de ejecutar la creación del sistema; entre estas divisiones encontramos el *Back Room* o como se le conoce también *área de adquisición de datos*, aquí se lleva a cabo la obtención de la información de manera consistente, otra de sus divisiones es el *Front Room* o *Adquisición de la información* esta área es la encargada de la presentación y manejo de la información al usuario. Además, cuenta con el área de infraestructura, la cual complementa la arquitectura técnica y, a su vez, se encarga del manejo de dispositivos de hardware, funciones a bajo nivel, etc.

El área más compleja del *Data Warehouse*, se ve representada por la *Metadata* debido a que, el término representa la información que es almacenada de los datos, resulta confuso, sin embargo sin ella la magia que muchas veces realiza el *Data Warehouse* no sería posible. Toda la teoría se ve representada en las herramientas que ayudan al usuario a crear los sistemas de *Data Warehouse*, entre ella tenemos *Oracle Warehouse Builder* y *Cognos 8 BI Data Integration*.



# 1. EI *DATA WAREHOUSE*

## 1.1 Introducción

A medida que los sistemas de información fueron evolucionando, estos crearon una cantidad de recursos que se acoplan a sus necesidades. Dentro de los recursos que han sido creados para la eficiencia de los sistemas se encuentran las bases de datos transaccionales, estas cuentan con un complejo funcionamiento cuyo fin es almacenar la información importante que los sistemas de información reciben y transforman. Una vez transformados los datos en información se almacenan en las bases de datos, sin embargo este hecho, es demasiado costoso para las empresas, ya que el espacio que ocupa la información de un cliente de hace 20 años en un sistema de información, representa un costo demasiado alto para las empresas. Para solucionar este problema la información es extraída de las bases de datos es puesta en medio como cintas magnéticas, arreglos de discos duros, etc., a este proceso se le conoce como *backup*, seguidamente los sistemas son limpiados de los datos antiguos. Una vez realizado este proceso resulta imposible tener los datos disponibles de un cliente de hace 20 años ya que la secuencia correcta debe de ser, ir a buscar la cinta en la que se encuentra el *backup* correcto y luego cargarlo a la base de datos. El proceso descrito anteriormente resulta demasiado costoso para las empresas ya que representa el tiempo perdido del operador que buscará el medio en el que fue almacenada la información, cargarlo y hacer uso de la misma.

Como solución a los problemas de las bases de datos transaccionales, surgieron los *Data Warehouse*, este es una colección de datos integrados,

orientado a un tema, no volátil y de tiempo variante. Un *Data Warehouse* es una base de datos relacional que esta diseñada para consultas y análisis más que para soportar ambientes transaccionales. Generalmente poseen los datos históricos de las bases de datos transaccionales que se encuentran en los ambientes operacionales, estos datos pueden ser extraídos de más de un lugar y presentados como uno solo en un ambiente definido por el *Data Warehouse*.

## 1.2 Características del *Data Warehouse*

Para la introducción del concepto de *Data Warehouse* es necesario tomar en cuenta ciertas características. Estas características fueron publicadas por William Inmon y son las siguientes:

- **Orientado al tema ( *subject oriented* )**

Un *Data Warehouse*, se dice orientado al tema, ya que a diferencia de los sistemas transaccionales que generalmente se orientan a las aplicaciones, estos se enfocan en las actividades que le son importantes a las empresas, por ejemplo se puede mencionar que un *Data Warehouse* se encuentra enfocado en las ventas, este permite el análisis de la información y ayuda a obtener información importante sobre las ventas, como lo es quien fue el mejor vendedor el año pasado, o quizá cuanto las ventas han aumentado en los seis meses anteriores. A esta característica se le conoce como orientado al tema.

- **Integrado (*integrated*)**

Un sistema integrado es definido, como aquel sistema cuya información es nombrada de la misma manera durante todo el tiempo que este se encuentre funcionando. Esto se refiere a que la información que se almacena dentro del *Data Warehouse* proviene de distintos ambientes operacionales, en los que cada sistema transaccional posee su definición de los datos que almacena, por ejemplo supongamos que en uno de los sistemas el campo sexo se almacena de la forma F, M, en otro se almacena 1, 0 y en otro se almacena Femenino, Masculino, esta inconsistencia de información no puede ser almacenada en el *Data Warehouse*, así que la información es transformada y colocada de una sola manera, a esta característica de los *Data Warehouse* se le denomina Integrado.

- **No volátil (*nonvolatile*)**

La información que se almacena en el *Data Warehouse* se dice no volátil, ya que no se encuentra expuesta al nivel de transacciones que soportan los ambientes operacionales, como actualizaciones, inserciones, eliminación, etc. La manera en que la información es cargada al sistema se puede dar de dos maneras, la primera es la carga inicial que se da cuando el sistema es creado, y la segunda es la actualización de los datos mensuales, semanales o diarios, este proceso se realiza durante un periodo en el que la carga al sistema es mínima, con esto se puede observar que la información que muestra en los análisis no siempre está actualizada, esto depende de las políticas que se hayan definido para las cargas de las actualizaciones de los datos.

- **De tiempo variante (*time variant*)**

Para poder analizar la información que se posee y descubrir las tendencias que estas poseen es necesario contar con los datos históricos de los ambientes operacionales. A diferencia del deposito de datos del *Data Warehouse* los sistemas transaccionales, tienen la desventaja de que la información es removida y puesta en archivos u otro medio luego de un cierto tiempo.

### 1.3 Diferencias de los ambientes OLTP y *Data Warehouse*

Entre estos dos ambientes existe una serie de características que los diferencian, la que es más notoria es que los modelos operacionales se encuentran estructurados en la tercera forma normal (3FN), no así los del *Data Warehouse*, para visualizar las diferencias se listan en la Tabla I.

Tabla I. Diferencias de los ambientes OLTP y *Data Warehouse*.

| <b>Actividad</b> | <b>OLTP</b>   | <b><i>Data Warehouse</i></b>   |
|------------------|---|--|
| Carga de Trabajo | Estos ambientes se han hecho para optimizar y responder a una serie de tareas definidas que serán soportadas por los sistemas | Los fines específicos son los de acomodarse a las consultas, este puede optimizar el rendimiento para una amplia gama de operaciones sobre consultas |

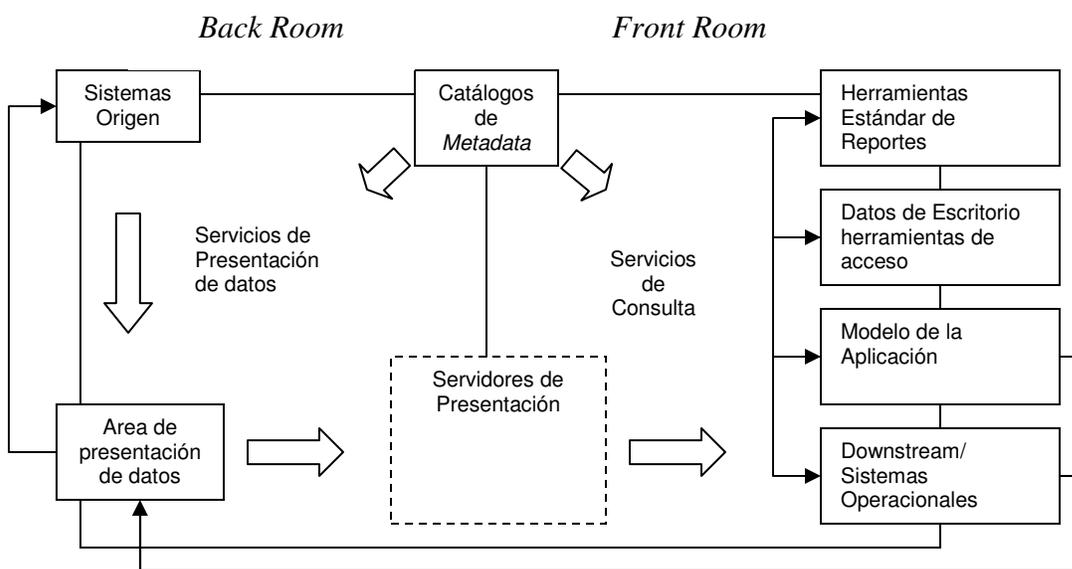
|                           |  |   |
|---------------------------|--|---|
| Modificación de los Datos | Aquí los datos generalmente son modificados por los usuarios finales de los sistemas operacionales. Los datos de estos sistemas reflejan el estado de las transacciones del sistema.               | Aquí los datos son modificados por medio del proceso de ETL ( <i>extract, transform and load</i> ), este se encarga de las actualizaciones de los utilizando técnicas de manejo de volúmenes grandes de datos. Aquí el usuario final no interfiere en las actualizaciones de información. |
| Diseño de Esquemas        | Para la optimización las inserciones, actualizaciones y modificaciones los datos se encuentran normalizados, esto garantiza la consistencia de la información que se almacena                      | Los esquemas utilizados son los de estrella, estos se encuentran denormalizados o parcialmente denormalizados, esto para optimizar el rendimiento de las consultas.   |
| Operaciones Típicas       | Se utilizan cantidades manejables de información, por ejemplo los datos de un determinado empleado   | Se exploran miles de filas al momento de ejecutar una consulta por ejemplo, para saber el total de las ventas de los clientes del mes pasado.   |
| Datos Históricos          | Los datos que se almacenan en los ambientes operacionales son los de 1 o 2 semanas, quizás 1 mes de transacciones, luego las bases de datos son limpiadas y la información puesta en <i>Backup</i> | Aquí se almacena la información de meses, años, esto da soporte al análisis histórico.  |

## 1.4 Arquitectura del *Data Warehouse*

El *Data Warehouse* al igual que cualquier sistema operacional posee una arquitectura. A continuación se muestra la división que se hace a lo que se conoce como arquitectura según *Ralph Kimball* (Ver figura 1):

- Arquitectura técnica
- Infraestructura
- *Metadata*

Figura 1. Nivel General de la Arquitectura Técnica



### 1.4.1 Arquitectura Técnica

Se refiere a la forma de representar las estructuras de datos, comunicación, procesos y presentación que existe para los usuarios finales dentro de una empresa<sup>1</sup>. La arquitectura técnica también incorpora las herramientas que se le aplicaran a la información, de esta forma ayuda a identificar como se realizara el cambio de la información de una base de datos a un almacenamiento que se ha definido según los requerimientos de los usuarios. la arquitectura técnica representa el lugar en donde el *Data Warehouse* se apoya, indicando como debe fluir la información a través de el.

Para una mejor comprensión de las funciones de la arquitectura técnica se ha dividido en:

- *Back Room* según se traducción, Cuarto de atrás, es también conocido como Adquisición de la información
- *Front Room* según se traducción, Cuarto de enfrente, a este también se le conoce como Acceso a la información

Estas dos áreas interactúan juntas en múltiples actividades, y ayudan a responder una serie de preguntas, tales como, ¿Qué es lo que necesitamos hacer? y ¿Cómo haremos que funcione lo que necesitamos?. Una de las actividades que no se puede pasar por alto al momento de construir la arquitectura técnica es que se debe tener un amplio conocimiento de los requerimientos de los usuarios, para definir y priorizar las actividades a desarrollar.

---

<sup>1</sup> ©Copyright 1996 by The Ken Orr Institute; revised edition, 2000, **A Data Warehouse Architecture**

Desde este punto de vista se puede identificar que la arquitectura técnica de los *Data Warehouse* es una combinación de código personalizado, del uso de herramientas desarrolladas en la empresa y herramientas estándar. Es importante recordar que la arquitectura técnica no se basa en la interacción independiente de cada uno de los componentes, si no en el equilibrio que guardan ambos haciendo del *Data Warehouse* un lugar de Almacenamiento de datos consistentes para su uso.

#### **1.4.2 Back Room**

La traducción de este termino es Cuarto de Atrás de aquí en adelante lo denominaremos *Back Room*. Este es el lugar en donde el proceso de presentación de información tiene lugar, aquí es en donde las preocupaciones de los Administradores de Bases de Datos de un sistema obtiene las respuesta para saber como es que la información se puede transferir de un punto A hacia un punto B en un estado consistentes.

A la parte del sistema *Back Room* y al proceso de presentación de información se les conoce con el nombre de Adquisición de Información. El proceso de representación de información da inicio cuando los desarrolladores codifican un sistema, sin embargo el *Back Room* provee las herramientas para automatizar las actividades que los desarrolladores realizan, dándoles con esto oportunidad de obtener características sensibles a la tecnología cambiante.

### **1.4.3 Front Room**

La traducción de este termino al español es Cuarto de Enfrente de aquí en adelante nos referiremos a el por su termino en inglés para una mejor comprensión. Esta área de la arquitectura se puede ver como la parte publica del *Data Warehouse*, esta es la interfaz con la que interactúa el usuario todos los días. Los usuarios del *Data Warehouse* no necesitan saber cómo es el proceso de extracción de la información o que tipos de estructura interna posee la información, ellos necesitan respuestas concretas a las preguntas que formulan.

Otro término común para el *Front Room* es el Acceso a la información, se puede considerar que es la parte del *Data Warehouse* responsable de proporcionar información consistente a los usuarios.

### **1.4.4 Infraestructura**

Es la base física del *Data Warehouse* aquí se definen las plataformas que darán soporte a la arquitectura técnica. Aquí se debe describir el hardware y el nivel de operación del software. Mucha de la infraestructura que se utiliza ya existe y forma parte de la infraestructura del Sistema de Información por sus siglas en inglés conocido como IS.

Como siempre antes de tomar una decisión de infraestructura se deben tener claros los requerimientos que el cliente y usuario definan. La línea que divide la arquitectura técnica de la infraestructura resulta confusa, así la infraestructura comprende para esta definición el Hardware, redes y funciones de bajo nivel, tales como seguridad. Cuando se consideran aspectos de la infraestructura se debe ser conciente que el *Data Warehouse* crecerá rápidamente. También se debe de considerar que la infraestructura no es independiente de la arquitectura Técnica(*Back Room, Front Room*) se encuentra ligada fuertemente una con la otra y da equilibrio a cada una de las áreas de la Arquitectura Técnica.

#### **1.4.5 Metadata**

Describir la *Metadata* es algo confuso, ya que un *Data Warehouse* debe contar con *Metadata* en el *Front Room* y en el *Back Room* la diferencia entre estas dos conceptos, es que la primera se encuentra en la extracción, limpieza y carga de los procesos y la segunda la distinguimos porque hace que las herramientas de consultas, funciones que realizan los reportes sean agradables.

A la definición anterior se puede agregar que la *Metadata* posee las siguientes características:

- Es un directorio que provee ayuda al analista, para una ubicación eficiente de los contenidos del *Data Warehouse*.
- Provee una guía para el mapeo de datos. Esto indica que ayudara a identificar como serán transformados los datos del ambiente operacional al del *Data Warehouse*.

- Muestra una guía de los algoritmos, que se encargan del resumen de los datos, en sus distintas fases (datos orígenes, ligeramente resumidos, resumidos en su totalidad, etc.)

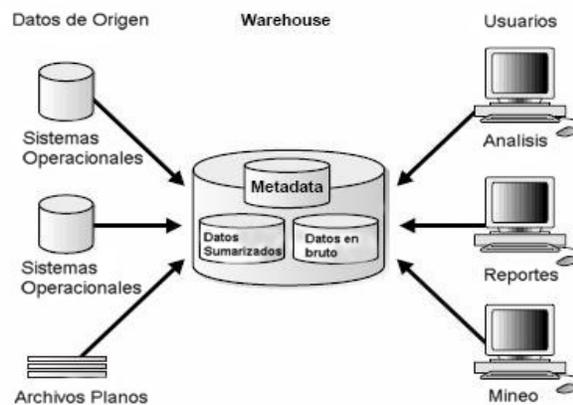
## 1.5 Arquitecturas del *Data Warehouse*

Existen muchas formas de mostrar la arquitectura del *Data Warehouse*, aquí se presenta tres estructuras que se identifican comúnmente.

### 1.5.1 Básica

La arquitectura básica como se muestra en la Figura 2, cuenta con la *Metadata*, los datos en bruto y los datos sumariados, se debe de notar que en esta arquitectura la limpieza de los datos del sistema operacional no se lleva a cabo, esta es una característica de este tipo de arquitectura, los datos de los sistemas operacionales son copiados tal y como se encuentran.

Figura 2. Arquitectura del *Data Warehouse*

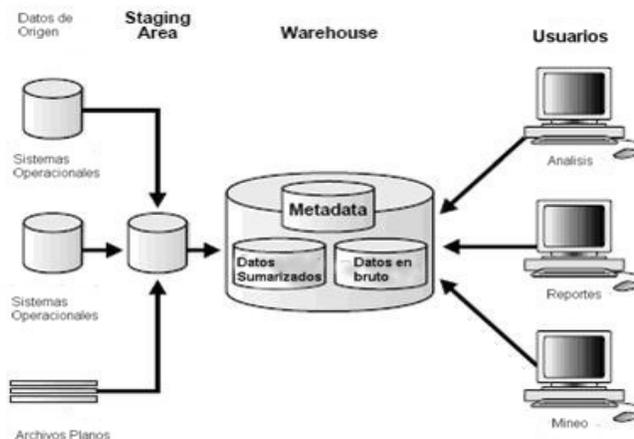


Fuente: Oracle9i Data Warehousing Guide, Release 2 (9.2) Part No. A96520-01, octubre 2005

### 1.5.2 Con un *Staging Area*.

Se puede notar que este tipo de arquitectura hace uso de la limpieza de los datos con ello elimina la información que sea innecesaria para el *Data Warehouse*. En el *Staging Area* es colocada la información que se transformara, generalmente se coloca en esta área una copia de la información de los sistemas operacionales, se debe de tomar en cuenta que no se copia toda la información si no solo la que se considere importante para los fines que se persiguen.

Figura 3. Arquitectura del *Data Warehouse* con *Staging Area*

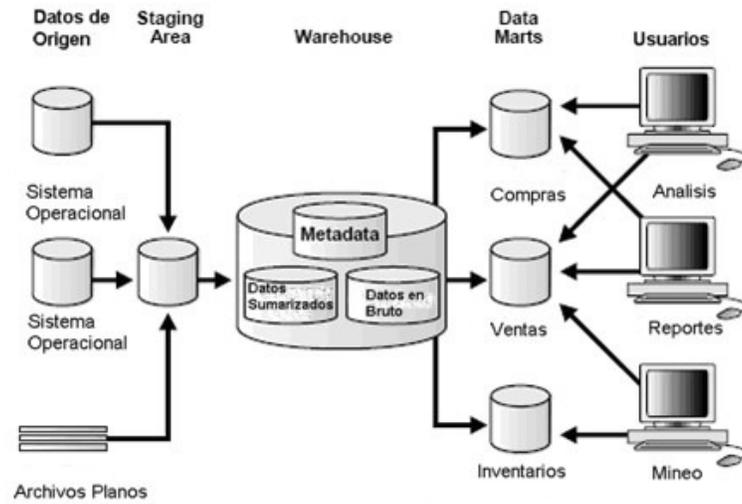


Fuente: Oracle9i Data Warehousing Guide, Release 2 (9.2) Part No. A96520-01, octubre 2005

### 1.5.3 Con un *Staging Area* y *Data Marts*

Como se observa en la figura 4, además de incluir una limpieza de los datos, esta arquitectura puede ser adaptada para diferentes grupos dentro de la organización, esto se logra por medio de los *Data Marts*, estos son diseñados para una línea específica del negocio, como las ventas, compras, etc..

Figura 4. Arquitectura del *Data Warehouse* con *Staging Area* y *Data Marts*



Fuente: Oracle9i Data Warehousing Guide, Release 2 (9.2) Part No. A96520-01, octubre 2005



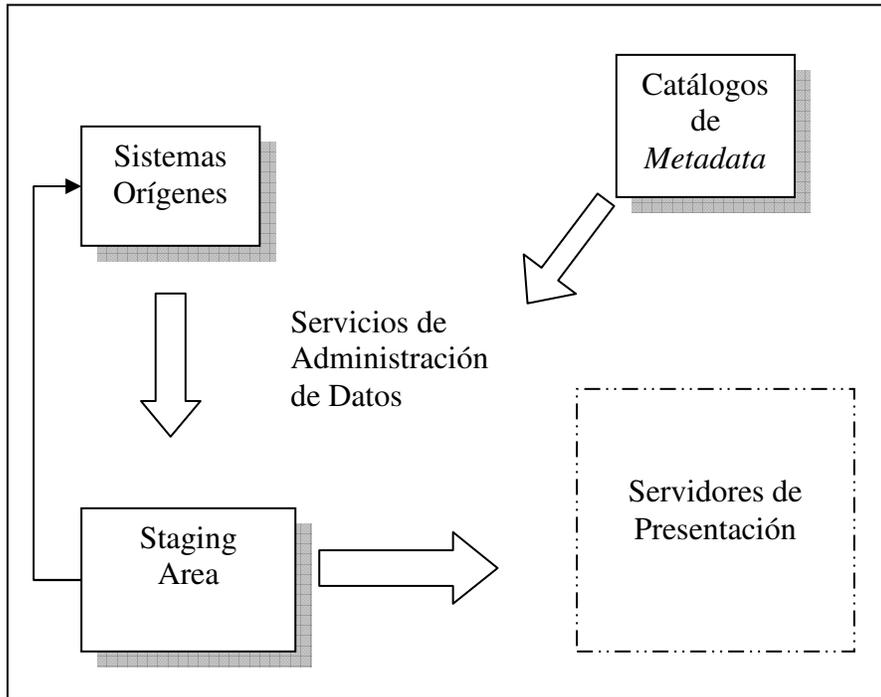
## 2. ARQUITECTURA TÉCNICA

### 2.1 *Back Room* (Adquisición De La Información)

Como se ha descrito en el capítulo anterior el *Back Room* es el lugar en donde da inicio el proceso de presentación de información, sin olvidar que es el encargado de realizar las transformaciones apropiadas a la información en el momento adecuado. Esta parte de la arquitectura del *Data Warehouse*, consta de varios segmentos importantes para el proceso de adquisición de la información, debido a los almacenamientos de información que maneja el *Back Room*, en algunas ocasiones estos son de tipo temporal y en otras permanentes, a ellos llega toda la información de todos los lugares que el *Data Warehouse* controla o controlará. Debemos de considerarse que el tipo de almacenamiento de datos que se necesite siempre depende de las necesidades que se describan en los requisitos planteados por el cliente y la complejidad del proceso de extracción de datos.

El *Back Room* posee una serie de áreas importantes, que son necesarias para que este ejecute sus funciones debidamente, la figura 5 muestra las partes que lo conforman.

Figura 5. Arquitectura técnica del *Back Room*



El proceso de presentación de datos involucra tres operaciones importantes: *extracción de los datos desde el origen, hacer que los datos pasen a través de los procesos de transformación, carga de los datos en el sistema específico*. Debe de recordarse que estos tres procesos no se dan de forma secuencial, en algunas ocasiones uno sigue al otro sin embargo en otras se pueden dar en cualquier orden.

### 2.1.1 Sistemas orígenes

El *Back Room* posee entre sus componentes un área denominada sistemas orígenes, estos sistemas son los que proveen de la información que alimenta al *Data Warehouse* de ellos se obtiene la información necesaria para que el usuario realice las funciones que solicita. Es importante entender la criticidad de este punto ya que define las herramientas de conexiones, servicios, etc. y todos dependen de la consistencia de la información y el lugar hacia donde se deposita finalmente toda la información. Entender los sistemas de origen es importante para la creación del *Data Warehouse*, entre estos encontramos los sistemas ERP, la instancia de reporte y los ODS (*operational Data Store por sus siglas en inglés*). La razón principal de combinar información de distintos sistema es proporciona referencias cruzadas de datos para las aplicaciones<sup>2</sup>

Cliente/Servidor Sistemas ERP(Enterprise Resource Planning), Los sistemas cliente servidor es de una generación de sistemas operacionales, fueron creados para resolver los problemas que se daban en los distintos departamentos de las empresas, cuentan con una gran cantidad de información dividida en muchas tablas que casi siempre son imposibles de consultar rápidamente. Uno de los problemas que se identifica en estos sistemas es que no proporcionan un soporte para la toma de decisiones, es por eso que este tipo de sistemas se convierte en el alimento del *Data Warehouse*.

---

<sup>2</sup> *Vivek R. Gupta, Senior Consultant*, Integrating data from more than one operational system, <http://system-services.com/dwintro.asp>

Se Debe recordar la diferencia que existe entre un *Data Warehouse* y un sistema Relacional, el sistema relacional fue creado para tener un alto rendimiento y tiempos de respuesta cortos<sup>3</sup> y se esperaba con ellos poder manejar volúmenes altos de transacciones. Debido a las exigencias descritas estos sistemas se volvieron complejos e inútiles para la toma de decisiones ya que el costo de una respuesta lenta o un tiempo muerto en el sistema llegaría a incurrir en pérdidas económicas importantes para la empresa.

Instancia de Reporte, Se obtiene también del sistema ERP y generalmente se encuentra separado del sistema base, debido a que el sistema central no posee la capacidad para atender las transacciones y consultas a la información al mismo tiempo. Con este tipo de sistemas orígenes podemos observar que desde los sistemas relacionales se encuentra la necesidad de obtener una retroalimentación de lo que sucede en el sistema, esta es la razón por la que existen los reportes sin embargo estos no proporcionan un buen soporte para la toma de decisiones. *Una instancia de reportes debe ser manejada como un componente central del sistema operativo*<sup>4</sup>.

---

<sup>3</sup> **Vivek R. Gupta, Senior Consultant**, Differences between transaction and analysis proceses, <http://system-services.com/dwintro.asp>

<sup>4</sup> **Ralph Kimball**, *The Data Warehouse The Life Cycle Toolkit*, pag 339

Almacén de datos Operacional (ODS por sus siglas en inglés), Según la definición de Ralph Kimball<sup>5</sup>, los almacenes de datos operacionales pueden definir de dos maneras, en el primer caso los servidores de almacenes de datos operacionales son un punto de integración para las operaciones del sistema. La segunda definición es que los almacenes de datos operacionales son suministradores, de detalle de datos para dar soporte a las decisiones. La primera representación de los almacenes de datos operacionales es un componente valioso para el ambiente operacional sin embargo el segundo solamente se una parte mas del *Data Warehouse*.

Hasta hoy día los ODS se han considerado sistemas independientes del *Data Warehouse*, debido a que estos soportan la carga operacional de los datos<sup>6</sup>. Sin embargo para los sistemas origen, es una parte importante porque en este tipo de almacenamientos se resguarda la información de las transacciones que se han realizado por una hora, un día o una semana, esto da la ventaja que no se satura el *Data Warehouse* con el almacenamiento de transacciones independientes a cada cierto tiempo. De esta manera podemos decir que ODS son sistemas que almacenan información volátil y estos pueden ser reescritos según se necesite. Esta clase de sistemas se pueden organizar en base al modelo estrella, que es la base del modelo dimensional.

---

<sup>5</sup> **Ralph Kimball, The Data Warehouse Lifecycle Toolkit, pag 340**

<sup>6</sup> **Ralph Kimball, Relocating the ODS, <http://www.dbmsmag.com/9712d05.html>**

## 2.1.2 Área de Presentación de Datos

En la arquitectura técnica también se desarrolla parte del Área de presentación de Datos (*Data Staging Area*), aquí son cargados, depurados combinados y almacenados los datos, además son exportados rápidamente a una o más plataformas de presentación. La meta de esta área es obtener la información y tenerla lista para que sea cargada dentro de los servidores de presentación, estos se refieren a los sistemas administradores de bases de datos Relacionales o a los de ingeniería OLAP. No consideramos al área de presentación de datos un servicio de consulta. Por ello que cada base de datos que se utilice como referencia de consulta, se asume que es parte física del área de presentación de datos<sup>7</sup>.

Podemos decir que cualquier base de datos que se encuentre disponible para realizar consultas se encuentra bajo el control del área de presentación de datos. Debemos hacer notar que los datos que son recogidos por esta se encuentran en las relaciones originales es decir de uno a muchos y de muchos a muchos, sin embargo para hacer el trabajo de obtener la información existen herramientas que realizan las traducciones según las especificaciones. Al momento de elegir la herramienta para adquirir la información de los almacenes originales se debe de considerar el volumen de los datos que se obtendrán.

---

<sup>7</sup>*Ralph Kimball*, Is Data Staging Relational?, <http://www.dbmsmag.com/9804d05.html>

### 2.1.3 *Metadata del Back Room*

En el capítulo uno se describe lo que es *Metadata* además se identifica que existe *Metadata* para cada área de la arquitectura técnica. Según Ralph Kimball <sup>8</sup>, existen diferentes formas de *Metadata* iniciamos con el origen de información( *su termino en inglés data source*), este es el primer tipo de *Metadata* que se debe de definir en un *Data Warehouse*:

- Especificación de repositorio
- Esquemas orígenes
- Especificación de copias de libros
- Especificaciones propias o de terceros
- Especificación de URL
- Especificación de presentación grafica
- Especificación de impresión de archivos en bobinas
- Especificación de viejos formatos para archivar información en un *mainframe*.
- Notas de especificación de hojas de calculo
- Descripciones propias de cada recurso
- Descripción del negocio de cada recurso
- Actualizaciones periódicas de cada recurso original
- Limitaciones legales de cada recurso original
- Métodos de acceso, correcciones de acceso, privilegios y contraseñas para el uso de recursos.

---

<sup>8</sup> *Raph Kimball*, Meta Meta Data Data, <http://www.dbmsmag.com/9803d05.html>

Luego que hemos obtenido la información esta puede ser colocada en el área de presentación (*staging area*) y ser preparada para la carga dentro de un *Data Mart*, ahora la preocupación es por describir la *Metadata* y el proceso que esta tiene:

- Transmisión cronológica de información y resultados de transmisión específica.
- Uso de archivos en el área de presentación incluyendo, duración, volatilidad, y propiedad
- Definición de dimensión y tabla principal (*Fact*)
- Especificación de tareas para reunir recursos, mostrar los campos y búsqueda de atributos
- Políticas de cambios lentos en dimensiones, para cada entrante atributo descriptivo
- Asignación de claves sustitutas para cada producción clave, incluyendo una tabla rápida de consultas para representar el mapeo en memoria.
- Copias de ayer para una dimensión de producción.
- Especificación de limpieza de datos
- Incremento de datos y mapeo de transformaciones.
- Transformaciones requeridas para *Data Mining*
- Diseño de esquema destino, datos propios, y carga de scripts de DBMS
- Agregar definiciones.
- Agregar usos estáticos, tabla base uso estático, agregados potenciales.
- Agregar modificaciones a las bitácoras.
- Descendencia de datos y auditoria de registros.
- Transformación de datos de bitácoras en tiempo de ejecución
- Datos para transformar el software en números de versión
- Descripción del negocio de los procesos extraídos.
- Configuración de seguridad para archivos extraídos, software y *Metadata*

- Configuración de seguridad para transmisión de datos
- Archivos de bitácora del área de presentación de datos y recuperación de procedimientos
- Configuración de seguridad archivos de presentación de datos

Además de la lista descrita anteriormente debemos de incluir

- Tablas del sistema Administrador de Bases de Datos
- Configuración de partición
- Indices
- Indicar las especificaciones de disco
- Indicar los procedimientos
- Nivel de seguridad del Sistema Administrador de Bases de Datos, Privilegios y concesiones
- Definición de vistas
- Procedimientos almacenados y scripts administradores de SQL
- Estado de copias de respaldo, procedimientos y seguridad.

Esta lista describe lo que se entiende por *Metadata* del *Back Room* y resulta importante en un *Data Warehouse*, además se debe de tomar en cuenta cuando realizamos el plan de la arquitectura.

## 2.1.4 Servicios del *Back Room*

Llamados también servicios de presentación de datos (conocido en inglés como *Data Staging Services*) son las herramientas y técnicas empleadas en el proceso de presentación de datos (conocido en inglés como *Data Staging Process*). Recordemos que al inicio describimos que el *Back Room* era el lugar en donde toma lugar el proceso de presentación de datos. Entonces diremos que un servicio es una tarea o función elemental<sup>9</sup>. Los servicios que presta el *Back Room* se encuentran divididos en: *Servicios de Extracción*, *Servicios de Transformación*, *Servicios de Carga* y *Tareas de presentación de datos Control de los servicio*.

### 2.1.4.1 Los Servicios de Extracción

Son los encargados de obtener los datos de los sistemas orígenes, este es el proceso que lleva mas tiempo en un proyecto de *Data Warehouse*, especialmente si el sistema del que se esta obteniendo la información es demasiado antiguo. Debemos de considerar que el 60 por ciento de las horas de desarrollo de un *Data Warehouse* son usadas en el proceso de extracción. Es necesario entender los requerimientos del proceso de extracción así podremos determinar que servicio necesitamos para realizar la extracción, este tipo se acciones involucran herramientas automatizadas que nos permiten tener un mejor control de lo que se realiza. Entre los servicios de Extracción se encuentran:

---

<sup>9</sup> *Ralph Kimball*, The Data Warehouse Lifecycle Toolkit, pag. 350

- Múltiples Orígenes

Este término nace cuando se reúne la información que almacenan los diferentes sistemas operacionales que alimentan al *Data Warehouse*, debido a que se encuentran construidos en distintas plataformas. Sin olvidar que cada uno de los sistemas operacionales puede tener un manejador de base de datos diferente.

- Generación de código

Debido a la diferencia que puede llegar a existir entre los sistemas operacionales, en algunos casos se debe recurrir a la generación de código, esto significa que se debe crear código que se ejecuta en los sistemas orígenes para que desde allí se creen archivos de extracción de la información. Para algunas herramientas significa código interno que es ejecutado.

- Múltiples tipos de extracción

Existen varios tipos de servicio de extracción que se utilizan para distintos propósitos entre ellos tenemos:

- Incremental
- Transacciones
- Renovación completa

- Replicación

La replicación en el contexto del *Data Warehouse* se refiere a la actualización continua que se le da a una tabla, en la presentación de la base de datos durante el día. La replicación puede hacer un espejo o copia de las dimensiones a múltiples *staging areas* de datos.

- Compresión y Descompresión.

De utilidad cuando se desea transferir información hacia distancias muy grandes, esta acción reduce el tiempo de transferencia de datos dependiendo de la naturaleza de los datos en los archivos originales.

#### **2.1.4.2 Servicios de Transformación**

Una vez que los datos han sido cargados a los almacenes que se hayan destinado para esta tarea, los datos sufren transformaciones que son necesarias para el rendimiento del *Data Warehouse*, se debe tomar en cuenta que la transformación es una combinación de procesos automatizados que realizan las herramientas de transformación. De esta manera el objetivo debe ser identificar la herramienta que proporcionen el mejor soporte para estos servicios. Entre estos procesos de transformación se encuentran:

- Integración

Abarca la generación de claves sustitutas, mapeo de las claves de un sistema a otro y mapeo de códigos con su descripción completa. Esta acción representa la responsabilidad de tener claves Maestras que puedan ser encontradas en una tabla hecha explícitamente para buscarlas. Es como una forma de respaldo que se guarda al momento de ejecutar esta acción.

- Cambio lento mantenimiento de dimensiones

Se aplica con las herramientas de presentación de datos, estas cuentan con algoritmos para administrar el cambio lento de dimensiones que son construidas allí.

- Control de la Integridad Referencial

Este aspecto de los servicios se representa como en el modelo relacional, con la regla de Integridad Referencial “*La base de datos no debe contener valores de llave ajena sin concordancia*”<sup>10</sup>. Esto quiere decir que en el *Data Warehouse* se resguarda la concordancia entre las dimensiones y la tabla principal (*Fact*).

- Denormalización y Renormalización

La denormalización es un proceso que consiste en agrupar las tablas que han sido normalizadas por medio de las reglas de Code cuando se trata de un modelo relacional. Y la renormalización es cuando poseemos por ejemplo archivos planos que se encuentran denormalizados y necesitamos un cierto tipo de registro, entonces debemos de agrupar los registros a esto le llamamos renormalización.

- Conversión de los tipos de dato

Transformar los tipos de las Bases de datos a los tipos que estamos utilizando. Entre los tipos más comunes encontramos los de caracteres, fechas, números, etc.

---

<sup>10</sup> *C. J. Date*, Introducción a los sistemas de Bases de Datos, pag. 278

- Calcular, derivar, asignar

Esta transformación se aplica a las reglas del negocio que se hayan identificado durante el proceso de los requerimientos. Se debe estar seguro que la herramienta elegida contenga un conjunto de herramientas entre las que incluya, manejo de cadenas, fecha y aritmética de tiempo, sentencias condicionales y de matemáticas.

- Valores nulos

Los sistemas antiguos por lo general no manejan el valor nulo, esto representa un problema para la transformación de datos ya que se debe de tomar precaución e identificar como se ha definido el valor nulo en el sistema que se esta transformando.

### **2.1.4.3 Servicios de Carga**

El proceso de carga de datos, es un proceso que la mayor parte de sus funciones son controladas por la plataforma. Algunas de las capacidades que debe tener son un proceso de carga de datos:

- Soporte para múltiples Objetivos
- Optimización de la carga.
- Soporte para el proceso completo de Carga.

#### **2.1.4.4 Tareas de Presentación de Datos y Control de los Servicios**

Estas tareas son importantes debido a que presentan al usuario una vista mas entendible del sistema, además de darle al *Data Warehouse* herramientas que lo ayudan a controlar los errores y a manejar las excepciones, proveen seguridad de bitácoras, cuando los sistemas de tornan complejos, algunas de las tareas que son necesarias en estos servicios son:

- Definición de tareas
- Programación de tareas
- Monitoreo
- Anotar lo que ha sucedido en el sistema
- Manejo de Excepciones
- Manejo de errores
- Notificaciones

#### **2.1.5 Procesos de Recuperación del *Back Room***

El *Data Warehouse* como cualquier sistema de información necesita copias de respaldo que lo ayuden a recuperarse cuando ocurra algún percance con alguno de los discos, una falla de energía eléctrica o el truncamiento de las operaciones en algún momento. Las copias de respaldo deben de tener las siguientes características:

- Un alto rendimiento
- Administración Simple
- Automatización de la finalización de las operaciones.

Las copias de respaldo realizadas al *Data Warehouse* usualmente son copias de respaldo físicas. Se obtiene la imagen de la base de datos en un cierto punto del tiempo, incluyendo índices y esquemas físicos de información.

Un punto importante en las copias de seguridad del *Data Warehouse* es que no solo son copias que se hacen de vez en cuando, este proceso necesita un plan que debiera iniciar listando las tablas por categoría que existan en el sistema además de las áreas de presentación. Luego se debiera estimar el esfuerzo y tiempo que nos tomaría recrear estas tablas, este proceso no incluye tablas que contengan historia, estas podrían no ser recreadas. Para realizar un plan de copias de respaldo se debieran tomar las siguientes consideraciones.

- Determine un proceso apropiado de copias de respaldo.
- Implemente el proceso
- Practíquelo.

El último punto de la lista resulta vital en un plan de copias de respaldo ya que de que sirve un excelente plan si somos incapaces de ponerlo en práctica.

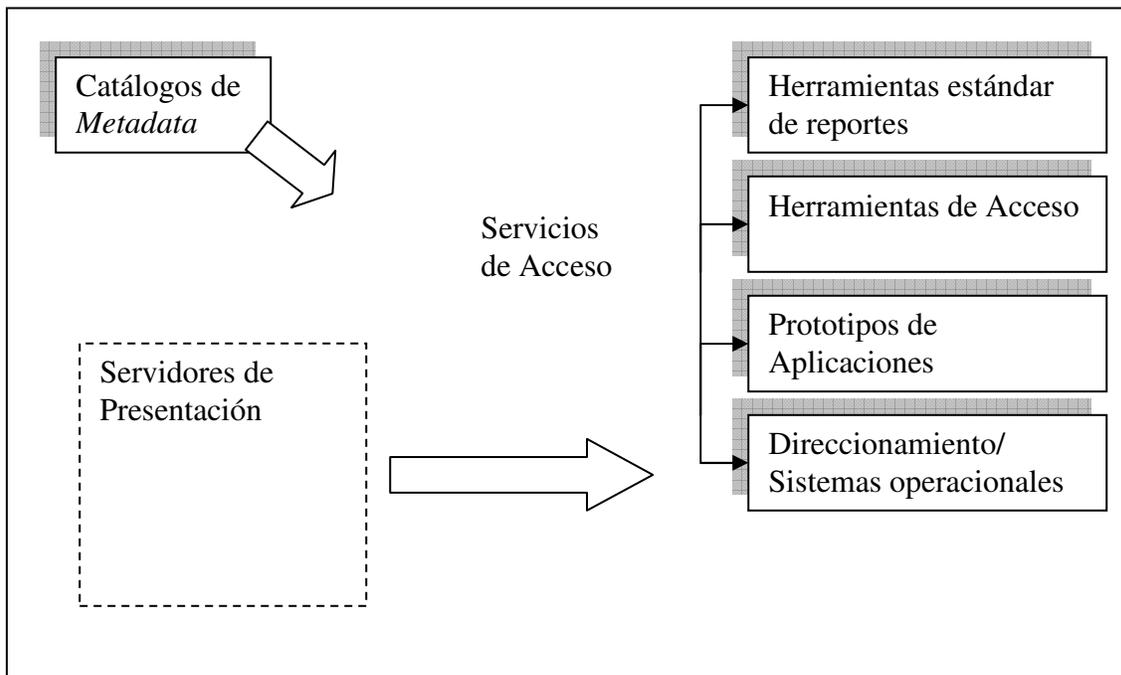
## **2.2 *Front Room* (Acceso a la Información)**

El *Front Room*, es el lugar en donde la presentación que los usuario ven se lleva a cabo, aquí al usuario no le importa que Sistema Administrador de bases de datos se utiliza o mejor aun cuales son los tipos de almacenamientos que se implementan en el sistema que esta utilizando, al usuario le interesa obtener la información que solicita lo mas rápido posible, pero además es importante que los resultado que obtenga sean precisos, para poder basar en ellos decisiones que pueden ser trascendentales para la empresa, si fuera el caso del Gerente de la empresa, o quizás para tomar medidas correctivas en cuanto a la estrategia de ventas que se esta utilizando.

El *Front Room* como se le llamara de ahora en adelante, al igual que el *Back Room* posee ciertas características, estas dos áreas juntas forman la fuerza que equilibran la arquitectura del *Data Warehouse*. Este también posee una serie de almacenes de datos, herramientas, servicios, catálogos de *Metadata* etc. Que le son importantes para darle al usuario final, la presentación amigable que este desea.

En la figura 6, se observan las áreas en las que se divide el *Front Room* del *Data Warehouse*.

Figura 6. El *Front Room* del *Data Warehouse*.



El *Front Room*, es por si mismo una herramienta, que utiliza una serie de herramientas que le permiten al usuario obtener la información que solicita.

### 2.2.1 La *Metadata* del *Front Room*

A continuación se describe una lista de almacenamientos de tipo *Metadata* que se le adjudican al *Front Room*<sup>11</sup>

- Pre empaquetar definición de consultas y reportes
- Asociar la especificación de la configuración de herramientas
- Especificación de la herramienta de una buena impresión

<sup>11</sup> **Raph Kimball**, Meta Meta Data Data, <http://www.dbmsmag.com/9803d05.html>

- Documentación para el usuario final y accesorios de entrenamiento, para abastecer a los vendedores y a la Tecnología de Información por sus siglas en inglés IT.
- Privilegios de perfiles para los usuarios en la seguridad de redes, certificados de autenticación, y uso de estadísticas.
- Perfiles individuales de usuario, con enlaces a recursos humanos para rastrear promociones.
- Enlaces de contratista y seguimiento de socios.
- Uso y acceso de mapas para los datos de elementos, tablas, vistas y reportes.
- Cambio de recursos tras las estadísticas.
- Sitios favoritos en la Web.

Al finalizar la descripción de lo que el autor del artículo *Meta Meta Data Data*, define como *Metadata*, podemos observar que esta área del *Front Room* y el *Back Room*, en conjunto es considerada como el área más importantes del *Data Warehouse* ya que unidas ambas partes representan todo el proceso que la información debe seguir para que el usuario final posea una aplicación transparente y fácil de utilizar.

### **2.2.2 Características Básicas**

Entre las características que posee el *Data Warehouse* es la flexibilidad y trata de ser un sistema abierto, claro considerando siempre lo más que se pueda esta característica. El sistema del *Data Warehouse* es accedido por muchas herramientas de usuarios y distintas plataformas.

El *Front Room* posee entre las bases que lo conforman, herramientas que le permiten acceso a grandes cantidades de información que se mueven por el medio de la infraestructura del *Front Room* y muy cerca del usuario final. Algunas veces estas herramientas trabajan con sus propios servidores de aplicaciones, en ocasiones estos servidores les proveen a sus herramientas un almacenamiento de datos temporal para ocultar los resultados de las consultas realizadas por los usuarios y los estándares de reportes.

Los almacenamientos de datos para los reportes estándares estos toman ventaja del tradicional sistema de reportes implementado en los sistemas Operacionales ya que poseen la capacidad de obtener la información de múltiples almacenamientos de datos, entre estos encontramos bases de datos dedicadas a reportes que forman parte del *Data Warehouse*. Esto da como ventaja la velocidad de respuesta en los reportes.

El *Data Warehouse*, maneja dos tipos de *Data Mart*. Los primeros llados *Data Mart Personales* permiten poseer un sistema en producción propio y uno que el usuario maneja, la ventaja de estos sistema es que proporcionan al equipo la facilidad de desarrollo descentralizado, sin embargo no se debe de olvidar que por muy independientes que puedan llegar a ser siempre tendrán necesidad de volver a actualizar los sistemas finales.

Los segundos son llamados *Data Mart* Desechables son utilizados para que el *Data Warehouse* obtenga el conocimiento de cómo es el ambiente en el que se desarrollara. Algunas veces por motivos de seguridad estos sistemas son creados fuera del sistema completo y solamente sirven para obtener un análisis de la situación real para luego adaptarla a la solución final. Estos sistemas al igual que los *Data Mart* personales son un conjunto de datos creador para soportar un escenario pequeño de la situación de un negocio en particular.

### 2.2.3 Mineo de Datos

También conocido por su término en inglés como *Data Mining* que es la manera a la que nos referiremos de ahora en adelante para una mejor comprensión. Este forma parte del modelo de aplicaciones, es un ejemplo de lo que esta área puede realizar, para entender a profundidad veremos cuales son las funciones y objetivos de este. Las consultas y reportes a los que el usuario esta acostumbrado, están adecuadas para responder las preguntas del negocio que responde al ¿Que?, la posibilidad de adentrarse dentro del detalle de los resultados responde a las preguntas de ¿Porque? Y ¿Como?.

El *Data Mining* es una ciencia que evoluciona. Este análisis inicia en los resúmenes que son mostrados a los usuarios finales y estos se adentran a los detalles buscando argumentos que aprueben o desapruében hipótesis<sup>12</sup>. Las herramientas del *Data Mining* han evolucionado rápidamente con el objetivo de entender el ambiente de las unidades negocios como los clientes y los productos.

---

<sup>12</sup> Vivek R. Gupta, Senior Consultant, ***An Introduction to Data Warehousing***

El *Data Mining* es una herramienta que ha evolucionado en características útiles para los negocios de hoy, este tipo de ayuda le permite al usuario predecir tendencia que no son obvias para los mejores expertos, al mismo tiempo que permite la consulta rápida sobre información almacenada durante mucho tiempo. Esta técnica tal como se le conoce, se encuentra soportada por tres tecnologías que en la actualidad se encuentran en un estado de madurez aceptable, las tecnologías que soportan el *Data Mining* son:

- Recolección masiva de datos
- Poderosas computadoras con múltiples procesadores
- Algoritmos de *Data Mining*.

Estas herramientas proveen una serie de capacidades como lo son:

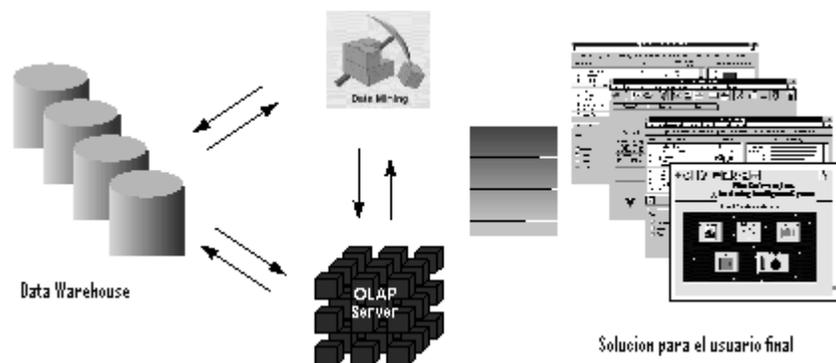
- Predicción automatizada de tendencias y ambientes, este tipo de características se obtienen cuando estas herramientas buscan en grandes bases de datos e identifican posibles tendencias que seria muy difícil identificar, por medio de largos análisis.
- Descubrimientos automatizados de patrones previamente conocidos, esto se refiere a la habilidad de este tipo de herramientas de identificar pasos ocultos para los sistemas, por ejemplo el ingreso incorrecto de los datos en una transacción de tarjetas de crédito o aun mas sencillo, la identificación de patrones para un supermercado que se habían obviado.

Entre las técnicas de *Data Mining* más conocidos encontramos:

- Redes neuronales artificiales
- Árboles de Decisiones
- Algoritmos genéticos
- Reglas de Inducción

Las herramientas del *Data Mining* en la mayoría de ocasiones se encuentran implementadas solas, sin embargo para obtener un mejor rendimiento de esta tecnología, debe estar completamente integrada con el *Data Warehouse*, ya que de no ser así estas necesitan un esfuerzo extra para extraer, importar y analizar los datos. A continuación en la figura 7, se describe una arquitectura para el *Data Mining*, puede ser implementada en una variedad de sistemas de bases de datos relacionales.

Figura 7. Integración de la arquitectura del *Data Warehouse*



Fuente: <http://www.thearling.com/text/dmwhite/dmwhite.htm>, agosto 2004

En la figura anterior se puede observar como interactúa el *Data Warehouse* con el servidor OLAP este le permitirá al usuario navegar a través de los datos contenidos en los almacenes del *Data Warehouse*. Esta estructura multidimensional de bases de datos le permite hacer uso de las herramientas de *Data Mining* para una mejor respuesta en el análisis de la información solicitado. Otorgando así al usuario la información que este ha solicitado.

## **2.2.4 Servicios**

Al igual que el *Back Room* el *Front Room* presta una serie de servicios entre los que encontramos:

### **2.2.4.1 Servicios para Accesos de Datos**

Este tipo de servicios le permite al usuario la navegación a través de los múltiples almacenes de información con los que cuenta un *Data Warehouse*. Este tipo de servicios deben de considerar la demanda que las herramientas Web hacen a los sistemas, ya que el *Data Warehouse* podría llegar a estar implementado en un ambiente Web. Este tipo de servicios cubre cinco tipos de actividades en el *Data Warehouse*:

- Navegación a través de almacenes o *Metadata*
- Acceso y Seguridad
- Actividades de monitoreo
- Administración de Consultas
- Reportes Estándar

El proceso de navegación a través de la información ha sido desde siempre un proceso del *Front Room*, debido a que el usuario inicia una herramienta del sistema y siempre lo que hace es buscar algún dato que le sea importante o necesite saber, las ventajas que presenta la navegación del *Front Room*, es que se implementa por medio de catálogos de *Metadata* que dan al usuario soporte en su afán de buscar y acceder a la información que necesitan.

La navegación a través de los almacenes de datos provee un enlace dinámico hacia los catálogos de *Metadata* que permite el despliegue de las áreas sujetos disponibles y de los elementos que contenga esos sujetos. Esto le permite al usuario localizar el valor de interés que se busque, además se podría proporcionar un enlace que permita asociar el recurso apropiado, como lo son: un reporte predefinido, alguna herramienta, o un cronograma de actividades.

Entre los servicios de acceso a los datos se encuentran los servicios de acceso y seguridad. Estos servicios también pueden ser llamados de Autorización y Autenticación, y ambos se refieren a las conexiones que los usuarios necesitan para acceder a las bases de datos. La autenticación se refiere a alguna técnica de verificar que la persona que dice ser es quien sea, esto se puede comprobar de distintas formas, esto dependen de la importancia que tengan los datos que los usuarios solicite. En el nivel básico se encuentran las contraseña almacenada en tablas que se localizan en la base de datos, y si le sumamos que se encuentre ligado a un sistema que obligue al usuario a cambiar su contraseña periódicamente, además de la contraseña podríamos requerir una evidencia física y podrían ser útiles las tarjetas magnéticas.

Sin embargo la autenticación es un servicio de infraestructura del cual los almacenamientos no son responsables y podría ser controlado por otro sistema. Una vez identificada la persona se debe saber a que tipo de información posee acceso. La autorización es un problema más complejo en los almacenamientos, porque limitan el acceso que puede tener un costo significativo para el mantenimiento, especialmente en un ambiente relacional

Las actividades de monitoreo están ligadas a cuatro áreas: de rendimiento, soporte a los usuarios, Mercadeo y planeación. El rendimiento proporciona información sobre el uso del sistema, dándole al *Data Warehouse* ventajas para poder proporcionar información con mayor eficiencia. Esta opción proporciona a los administradores de la base de datos información de cuales tablas son las de mayor consulta y que consultas están siendo utilizadas para realizar cambios que se hayan identificado y previamente analizado. El soporte a los usuarios se refiere a que el equipo del *Data Warehouse* se toma el trabajo de monitorear las acciones de los usuarios para identificar que es lo que ellos hacen, con esto están preparados para responder a las consultas o a los posibles problemas a los que se enfrenten los mismo, esto garantiza que nuestros usuarios tendrán una experiencia grata al utilizar el *Data Warehouse*. El mercadeo, utiliza estadísticas que proporcionan a los gerentes como se esta utilizando su inversión. La planeación es identificar el promedio de tiempo, el manejo de la concurrencia en base a la cantidad de usuarios, el tamaño de las bases de datos, y cuantificar el tiempo que es necesario para un incremento en la capacidad.

La administración de consultas le permite al usuario una interacción con las bases de datos, existen algunos tipos de servicios de administración de consultas que se deben de evaluar al momento de identificar una arquitectura entre los que encontramos:

- Simplificación de Contenido
- Re formulación de Consultas
- Agregar conocimiento
- Apreciación del tiempo
- Manejo de consultas

El servicio de reportes estándar le permite al usuario producir estilos con formatos estándar, reportes que tienen un uso limitado además de poder regular la ejecución de las agendas. Algunos requerimientos para que las herramientas estándar de reportes incluyen:

- Ambiente de Desarrollo de Reportes
- Ejecución iterativa
- Definición flexible de los reportes
- Puesta en marca flexible de los reportes
- Facilita a los usuarios la publicación y la suscripción.
- Enlace de los reportes
- Librerías de reportes con capacidad de navegación
- Herramientas administrativas para el ambiente de los reportes.

### **2.2.5 Servicios de Escritorio**

Le permiten al usuario manejar de una manera transparente la información que solicita a una base de datos. Estos servicios son los responsables de la aprobación que un usuario tenga sobre el sistema que maneja. Al momento de crearlos se debe de tomar en cuenta la información que necesita cada usuario sistema. La tabla 3.1 muestra las necesidades que presentan diferentes usuarios relacionados a un sistema, los tipos de acceso a los datos que necesita, y las herramientas que utiliza generalmente.

Tabla II. Necesidades de la información y atributos de los usuarios

| <b>Categorías de Necesidades de Información</b>   | <b>Roles de usuario</b>   | <b>Categorías de acceso a Datos</b>          | <b>Herramientas Comunes</b>  |
|---|---|--|--|
| <b>Alto nivel de monitoreo</b> , métricas claves, banderas  | Gerente Señor   | Presionando botones, "Presentación Elegante" | EIS; Algunas herramientas de con ambientes de consultas, como MS Acces, Visual Basic                 |
| <b>Rastreo de Comercio</b> , mercados, productos, clientes, etc. Cambio Rápido (drill down) para detalles | Gerentes medios, vendedores, gerentes de mercadeo, gerentes de comercio, etc. | Reportes estándares, manejo de parámetros    | Herramientas de reportes, administración de ambientes de consulta, presentación del estilo OLAP.     |
| <b>Investigativo</b> , excepciones, problemas nuevos u oportunidades, desarrollo de casos de negocios     | Los del nivel anterior mas analistas del negocio                              | Análisis ad hoc                              | Herramientas OLAP ambientes de administración de consultas, herramientas de alta calidad de análisis |
| <b>Análisis complejo</b> , mezcla de consultas, análisis estadístico, modelos de desarrollo               | Analistas del negocio y Expertos Analíticos                                   | <i>Data Mining</i> , análisis avanzados      | Herramientas de alta calidad de análisis, herramientas estadísticas, herramientas <i>Data Mining</i> |

## 3. EL CICLO DE VIDA

### 3.1 Introducción

El *Data Warehouse*, al igual que el software posee un ciclo de vida, que le permite identificar y regular las actividades importantes para el mantenimiento y desarrollo del mismo. Antes de definir la arquitectura de un *Data Warehouse* y las características propias de esta, es necesario recordar que se debe de identificar claramente las actividades del ciclo de vida. Luego debemos de considera la importancia del *Plan de Arquitectura*, ya que sin estas herramientas bases de la arquitectura la definición de la arquitectura técnica y la infraestructura resultaría inútil. Las actividades que se consideran parte del ciclo de vida del *Data Warehouse*, son:

- Planeación del proyecto
- Definición de requerimientos del negocio
- Diseño de arquitectura técnica
- Selección e instalación de productos
- Modelado dimensional
- Diseño físico
- Diseño y desarrollo del *Data Staging*
- Especificación de las aplicaciones del usuario final
- Desarrollo de las aplicaciones del usuario final
- Puesta en marcha
- Mantenimiento y crecimiento

Todas las actividades descritas son monitorizadas por la administración del proyecto.

### **3.2 Plan de Arquitectura**

Las actividades descritas anteriormente son las partes del ciclo de vida, sin embargo el plan de arquitectura debe de considerar los siguientes aspectos:

- Comunicación

En un plan de arquitectura es la fase del proyecto en la que se comprende la magnitud y complejidad del problema. La comunicación ayuda a comprender, como es que cada persona que conforma el equipo se ajusta en el proceso, además de definir que es lo que cada persona debe hacer.

- Planeación

El plan de arquitectura generalmente muestra requerimientos técnicos que no son incluidos en la planeación de los proyectos, aquí es el lugar en donde se presenta los detalles técnicos de la arquitectura.

- Flexibilidad del Sistema

La flexibilidad de un *Data Warehouse* se basa en que confía en una capa llamada Semántica. Esta capa describe el contenido del *Data Warehouse* y sus procesos, es usada para crear, navegar y mantener, esto lo hace flexible.

- Facilidad de Aprendizaje

Como en cualquier proyecto la facilidad de que otras personas ajenas al proyecto se enteren y aprendan el mismo es importante, para ello se utiliza la documentación que muestra la información escrita de lo que el *Data Warehouse* o *Data Mart* representara. Esto ayuda a los nuevos miembros a la creación de componentes, contenido y conexiones que sean más eficientes y eficaces.

- Mejorar la Productividad

Una mejora en la productividad se puede dar cuando la arquitectura nos ayuda a elegir herramientas que automatizan parte del proceso.

A lo largo de los capítulos anteriores se ha discutido la diferencia que existe cuando se retrae, almacena y consulta la información en un *Data Warehouse*, a esto se le ha denominado Arquitectura y cuando se hace uso de los almacenamientos de *Metadata* se le denomina Arquitectura Técnica del *Data Warehouse*.

Es importante entender la importancia de la arquitectura de un *Data Warehouse* ya que esta debe estar documentada claramente, puesto que representa la descripción de los elementos, servicios, además mostrara con detalle como se comportaran los componentes juntos y al mismo tiempo se debe de considerar el crecimiento que se espera tenga el sistema.

Es por ello que se puede decir que la arquitectura es un conjunto de documentos, planes, modelos, esquemas y especificaciones, con una definición en detalle para cada punto clave, esto les permite a las personas expertas desarrollar las especificaciones sin problema. Sin embargo la arquitectura no es únicamente un conjunto de documentos que dicen como se deben hacer cosas, la arquitectura representa Que es lo que se hará.

### **3.3 Como Iniciamos un Plan de Arquitectura**

No existe una receta para iniciar ningún plan, por lo tanto es importante recordar que cada decisión que se tome o cada requerimiento solicitado que se omita traerá consecuencias negativas o positivas a nuestro proyecto. Para evitar el riesgo de caer en situaciones comprometedoras se debe de considerar<sup>13</sup>:

- Como es el Negocio?

Se debe de reconocer como es que se realizan las actividades actualmente en el negocio, si se utiliza algún tipo de sistema gerencial, cual es y como funciona actualmente para los miembros de la empresa.

---

<sup>13</sup> **Peter Nolan**, Getting Started and Finishing Well,  
[http://www.intelligententerprise.com/010507/webhouse1\\_1.jhtml?\\_requestid=123231](http://www.intelligententerprise.com/010507/webhouse1_1.jhtml?_requestid=123231)

- Porque?

Para contestar esta pregunta, debemos de identificar porque se realizan las actividades actuales, no juzgar si están bien hecha o no. Es necesario conocer a detalle la información para identificar las causas de los problemas identificados, además a estos problemas se les debe de dar un análisis extenso para identificar las posibles soluciones al mismo.

- Que pasa si?

Una vez que se ha identificado el porque del problema, se puede formular la pregunta pasa observar si el negocio se comporta de una manera diferente. Este tipo de comportamientos se puede prevenir si se estima el modelo de la arquitectura en distintos escenarios. De esta manera se tendrá una certeza de que pasará si, el negocio cambia rápidamente.

### **3.4 Que es la Arquitectura del *Data Warehouse***

Cuando relacionamos información estratégica de la empresa, con la arquitectura de las aplicaciones, la arquitectura técnica, obtenemos como resultado Información de la arquitectura empresarial. La arquitectura del *Data Warehouse* es un subconjunto de esta clase de arquitectura. La arquitectura en general de una empresa, es un conjunto bien de documentado de los siguientes puntos:

- Metas estratégica, objetivos, estrategias
- Reglas del negocio y métricas
- Información de los requerimientos
- Sistemas de aplicación existentes

- Relaciones que existen entre aplicaciones, actividades, requerimientos del negocio y los elementos de datos.
- Y en la infraestructura de la tecnología.

La arquitectura y diseño del *Data Warehouse* de una empresa debe de reflejar los requerimientos de la empresa y su rendimiento continuo. La arquitectura no se refiere a elegir una serie de tecnología y utilizar tecnología de punta. Para la creación de la arquitectura es necesario que se tome en consideración los siguientes puntos:

- **El modelo de datos**

El modelo de datos describe cuidadosamente todos elementos de datos que tienen valor para el *Data Warehouse*, este también provee una definición clara y sin ambigüedades de cada entidad clave de datos, describiendo la forma en que se utilizaran, así como también define las formas de derivación que serán utilizadas, periodos de refrescamiento, etc. También se debe de incluir en el modelo de datos que datos serán retraídos de los almacenamientos orígenes para poblar el *Data Warehouse*, como será retraída y transportados los datos, y como los datos serán transformados para asegurar la integridad y la calidad de la información del *Data Warehouse*. Debemos recordar que existen dos características importantes del *Data Warehouse* La calidad de los datos y el Acceso a ellos.

## - **La *Metadata* del *Data Warehouse***

La *Metadata* del *Data Warehouse* también llamada datos de los datos, ejecuta un papel importante dentro del sistema esto no quiere decir que es lo principal pero si es necesaria. La *Metadata* es utilizada para mantener información permanente fuera del *Data Warehouse*, si se le observa como una analogía entonces se dice que es un repositorio digital par que los usuario accedan a la información con mayor rapidez. Para los usuarios representa el control de acceso y análisis. Ahora se divide la *Metadata* en dos tipos distintos de los descritos en los capítulos anteriores. Diremos que la *Metadata* es de tipo Estructural y de Acceso, esto representa que los nombres que se han trabajado durante el documento no son los únicos términos que se conocen.

### ***Metadata* Estructural**

Es la encargada de la creación y mantenimiento del *Data Warehouse*, esta describe completamente la estructura y el contenido, lo que necesitamos para construir la *Metadata* estructural es el conjunto de entidades, como se relacionan entre ellas y sus características. Además necesitamos saber que usuario tendrán que interactuar con el sistema, o cuales son sus usuarios potenciales. Este tipo de *Metadata* incluye usos estratégicos y operacionales que le permiten a la empresa saber la eficiencia del sistema.

## **Acceso a la *Metadata***

Este tipo le permite al usuario el acceso a la información, por medio de esta *Metadata* el usuario accede a la información del sistema. En ella se definen reglas de seguridad y acceso, al mismo tiempo que permite el acceso a los distintas jerarquías que se encuentren definidas, realiza el manejo de las consultas, y posee una descripción de los servidores, tablas, entidades, y campos que posee la información. Es capaz de manejo de grupos de trabajo, y en ocasiones es responsable del manejo creación y distribución de análisis estadísticos.

### **3.4.1 Las Estructuras del *Data Warehouse***

El *Data Warehouse* como cualquier sistema esta compuesto de una serie de estructuras que hacen que el sistema funcione correctamente, sin embargo se debe de recordar que existen unas cuantas estructuras que resultan importantes para su buen funcionamiento entre ellas encontramos:

#### **- Datos Físicos del *Data Warehouse***

Estos están representados por el conjunto de bases de datos que almacenaran los datos operacionales

#### **- Datos Lógicos del *Data Warehouse***

Ayudado este proceso por la *Metadata*, aquí se almacenan las reglas del negocio y el proceso lógico para retraer, organizar, empaquetar y almacenar los datos.

#### - **Data Mart**

Es una parte pequeña del *Data Warehouse*, en el se puede almacenar información de áreas específicas de la organización. Con la ayuda de este tipo de estructura es posible la redundancia de la información en el *Data Warehouse*, ya que la información se encuentra en el Data Mart y a la misma vez el *Data Warehouse* la puede acceder.

### **3.4.2 Componentes del *Data Warehouse***

La arquitectura del *Data Warehouse* contiene una serie de componentes necesarios para su funcionamiento, entre ellos encontramos.

#### - **Detalle Actual**

Aquí es donde encontramos toda la información proveniente de los sistemas operacionales, aquí se almacena la información como se obtiene, representa el nivel mas bajo de granularidad en el *Data Warehouse*.

En esta área se almacena la información histórica de las empresas, en ocasiones es necesario almacenar la información por periodos largos de tiempo, sin embargo se podría almacena por periodos de cinco años y si la información pasados los cinco años para la empresa no es indispensable se podría eliminar, esto último no es muy recomendable.

## - **Sumarización de los datos**

Existen dos tipos de sumarización de datos, el primero es un sumariado al que se le llama ligero, este proporciona información entre las distintas entidades que conformen el *Data Warehouse (Data Mart)*. El segundo es un sumariado mas detallado este les permite a los altos ejecutivos observar el comportamiento que tienen las cifras de los productos en determinado departamento.

## - **Archivos**

Es toda la información del detalle actual que tenga mas de dos años de antigüedad, esta es almacenada en archivos que poseen el mismo detalle de los datos originales y además contienen la *Metadata* que daba soporte a estos datos, para el mejor manejo de este tipo de información si fuera necesario por algún motivo consultarla.

## - **Sistema de grabado**

Estos son componentes creados una vez que los datos han superado las pruebas de integridad, exactitud y tiempo de respuesta. Son creados a decisión de los procesos.

## - Programas de integración/transformación

Estos componentes relacionan los sistemas operacionales con el *Data Warehouse*. En la mayoría de ocasiones la información proviene de distintos sistemas operativos, bases de datos, archivos planos, *mainframes*, etc., existen incompatibilidad entre los orígenes de la información y es por ello que los sistemas de integración y transformación son los encargados de la limpieza, edición y formateo de los datos una vez cargados al *Data Warehouse*.

Sin embargo estos sistemas poseen una serie de actividades que se debe de asegurar que realicen, entre ellas encontramos:

- Calcular, modificar o formatear nuevamente las estructuras claves y otros elementos de datos.
- Agregar elementos de tiempo
- Identificar valores predefinidos
- Sumarizar, etiquetar, reunir datos de múltiples orígenes.
- Compaginar datos de múltiples orígenes.



## 4. HERRAMIENTAS

### 4.1 Introducción

La arquitectura técnica del *Data Warehouse*, como concepto esta representada por el *Back Room* y el *Front Room*, ambas partes se encuentran relacionadas por medio de la *Metadata*. Cuando estos conceptos son llevados a la practica se incorporan a un ciclo de vida como cualquier software, sin embargo entre las herramientas que se identifican y se apollan en el ciclo de vida del *Data Warehouse* encontramos *Oracle 9i Warehouse Builder*, los productos de *Cognos*, etc.. La mayoría de estas herramientas de Business Intelligence proveen una solución integrada para diseñar y construir *Data Warehouse* empresariales, Data marts y aplicaciones de *Business Intelligence*, proporcionando todo lo necesarios para el mantenimiento del ciclo de vida del sistema que se construye.

A lo largo de los capítulos anteriores se ha descrito la teoría de la arquitectura técnica del *Data Warehouse*, aquí se podrá observar las diferencias que cada herramienta le proporciona a las diferentes fases del ciclo de vida del *Data Warehouse*.

## 4.2 *Warehouse Builder*

La arquitectura del *Warehouse Builder* se encuentra constituida de dos áreas

- El ambiente de diseño
- El ambiente de ejecución

Estos dos ambientes manejan partes importantes del sistema, el primero maneja la *Metadata*, mientras el segundo se encarga de los datos físicos.

- Componentes de diseño

Para la manipulación de la *Metadata*, se lleva a cabo por medio de los repositorios que se encuentran ubicados en la base de datos Oracle, además posee un conjunto de reportes escritos en java y HTML que proveen un ambiente amigable para el usuario.

La creación de *Metadata* es una actividad del diseño de nuevos sistemas de *Data Warehouse*, la herramienta provee herramientas a los clientes para que puedan crear objetos, procesos, etc. Para la creación de la *Metadata*. Proporciona al cliente la facilidad de diseñar los procesos de ETL, creación de modelos relacionales y multidimensionales.

Proporciona a los usuarios la capacidad de realizar un análisis de impacto que permite identificar la magnitud que tendrán los cambios luego de haber sido terminado el sistema.

- Componentes de ejecución

Una vez que se tiene el diseño del proceso ETL o parte lógica, es necesario mover la información a bases de datos físicas. Antes de realizar este proceso, cierta información de la base de datos es agregado al diseño lógico. Luego cuando la configuración es completada, el código es generado.

Debido a que la información que se extraerá de los ambientes operacionales, se encuentran en distintos manejadores de base de datos, *Warehouse Builder*, proporciona la facilidad de extraer la información de los ambientes que se le indiquen, además la información se genera como sentencia SQL DDL, pueden ser creadas en archivos o bien como objetos dentro de la base de datos.

- Características Principales de *Warehouse Builder*

La creación de un *Data Warehouse*, involucra una gran cantidad de sistemas operacionales, recursos y áreas funcionales, el correcto manejo de toda esta información resulta sumamente complejo. *Warehouse Builder*, proporciona ayuda sobre este problema proveyendo una interfaz integrada y fácil de utilizar, con esto asegura la escalabilidad, confiabilidad y flexibilidad.

Dentro de las características claves del *Warehouse Builder* encontramos:

- Importar la definición de los datos orígenes
- Diseñar y crear el esquema destino de la base de datos
- Definición de los datos a ser trasladados y transformación dentro de el origen y el destino

- Asignar dependencias dentro del proceso de ETL.
- Administrar y actualizar las definiciones origen
- Poner en marcha, actualizar y administrar los esquemas destinos
- Diseñar y crear ambientes de herramientas de consulta con fines específicos.
- Diseñar y crear un ambiente OLAP.

*Warehouse builder*, contiene un conjunto de objetos que pueden ser importados, diseñados y puestos en construcción. Entre los objetos que se destacan se encuentran:

- Colas Avanzadas
- Tablas externas
- Cubos
- Dimensiones
- Vistas materializadas
- Secuencias
- Tablas
- Vistas
- Flujos de procesos

Cuando se construyen sistemas destino en el *Data Warehouse*, es necesario considerar que existe una gran variedad de sistemas destino que pueden ser construidos. Depende de las necesidades del negocio, se debe de elegir que es lo que el diseño debe de ejecutar. El *Warehouse builder* puede construir sistemas para:

- Bases de datos relacionales
- Bases de datos OLAP

- *Oracle Discoverer*
- *Oracle Work Flow*

Dentro de los beneficios que proporciona *Warehouse builder* se pueden mencionar:

- Diseños rápidos, es proporcionada al usuario por medio de herramientas visuales fáciles de utilizar.
- Diseños centralizados, dado que toda sobre el diseño del sistema es almacenado en un repositorio de *Warehouse Builder*, que se encuentra en un solo lugar.
- Reducción del tiempo para incorporar cambios, una administración avanzada de las características del ciclo de vida basada en un simple repositorio, proporciona una suave incorporación de los cambios en los procesos.
- Código libre de errores, se considera libre de errores ya que este es generado de un repositorio único en donde es almacenada la información sobre el diseño, además de ser libre de errores, también es fácil de ser recreado, actualizado y administrado.

### **4.3 Cognos**

Provee un conjunto de herramientas que integran el ciclo de vida del *Data Warehouse*, entre ellas tenemos *Cognos 8 BI Data Integration*, esta herramienta apoya el proceso de unir, extraer, transformar y la administración de dimensiones, para la construcción de *Data Warehouse*, con altos rendimientos del *business intelligence*.

Esta herramienta posee dos tipos de *Framework* el primero que se nombra *Dimensional Framework*, que es en donde es llevado a cabo todo el proceso de creación de dimensiones, esto significa que la gran cantidad de sistemas orígenes son integrados y colocados en un ambiente consistente.

En el dimensional *framework*, se soporta una amplia gama de sistemas orígenes, estos incluyen información de sistemas relacionales (Oracle, SQL, IBM, Teradata, Sybase, and ODBC), así como también dimensionales como IBM DB2, OLAP, también sistemas de ERP, modernas fuentes como Web Services y XML, EXCEL y archivos planos, y main frames.

Entre las características principales del *Dimentional framework* encontramos:

- Garantiza que la información es estructurada de acuerdo a las dimensiones específicas del negocio, tales como clientes, tiempo y ubicación.
- Implementa una vista consistente de la información corporativa, a través del *Data Warehouse*.
- Provee una ubicación centralizada para actualizar las definiciones de las dimensiones.

Un aspecto importante es la centralización de la *Metadata* con esto se pueden aplicar un conjunto de reglas consistentes del negocio. Posee una integración con el *Cognos Query Framework*, que proporciona apoyo al modelado previo de la información existente así como la asociación de joins/cardinalidad, reglas del negocio, tales como los filtros y los cálculos predefinidos.

Para el manejo de la información del sistema esta herramienta proporciona una integración con *Cognos Planning Data*, permitiendo la integración de la información actual con los planes.

#### **4.4 Características principales de una herramienta**

Parte importante del proceso de creación de un *Data Warehouse*, es la elección de la herramienta a utilizar, un conjunto de características que se pueden considerar importantes al momento de decidir una herramienta son las siguientes:

- Paquetes de *Data Warehouse* y *Data Marts*.
- Almacenes de tiempo real y reportes de tiempo real
- Construcción de aplicaciones analíticas.
- Herramientas OLAP
- Herramientas de ETL.
- Sistema de reportes de funciones cruzadas.

Algunos criterios que se deben de cumplir en un *Data Warehouse*:

- La variedad de componente del *Data Warehouse* que pueden ser compartidos por las dimensiones, estas representan las dimensiones que tienen un significado, estructura e información idéntica. Para identificar si deben permanecer separadas físicamente, debe de decidirse basado en la complejidad de la implementación.
- Documentar e integrar las reglas del negocio, que serán utilizadas a través de los componentes del *Data Warehouse* en la arquitectura.

- Una herramienta simple de ETL, que permita el aseguramiento de la *Metadata* común.

#### 4.5 Diferencias

- *Warehouse Builder* es una herramienta que se encuentra integrada, provee procesos de ETL, modelación y creación de *Metadata*, al mismo tiempo funciones de transformación, creaciones y manejo de *Data Marts*, y generación de código para otros sistemas, etc.
- *Cognos 8 BI*, provee un conjunto de frame Works que se incorporan unos con otros para prestar una serie de servicios de ETL, modelación de *Metadata* e integrarse con herramientas de consulta.
- La estabilidad de ambos productos se ve reflejada en la amplitud de sus recursos. Por ejemplo el *Warehouse Builder* es una herramienta que ha sido evolucionada a la par de los productos *Oracle*, es por ello que posee una interconexión, entre su base de datos y productos tales como *Discoverer*, transparente, esto lo hace una herramienta, robusta que puede manejar la conectividad con distintos ambientes al momento de la extracción de la información.
- El caso de *Cognos* es totalmente diferente, puesto que es una empresa dedicada al BI. Se entiende que sus productos se encargan de que la información que se posee sea utilizada adecuadamente, para la toma de decisiones, en el análisis, acceso y descubrimiento de nuevas oportunidades.

- Las diferencias se enmarcan en el proceso de ETL, para vendedores de la talla de *Oracle*, este proceso resulta pequeño o quizá inexistente, sin embargo para *Cognos*, representa la definición de una serie de actividades que deben ser monitoreadas

Se han descrito dos herramientas, que poseen capacidades totalmente distintas, elaboradas por empresas que persiguen fines distintos. Esto refleja la importancia que debe darse a los pasos iniciales del ciclo de vida del *Data Warehouse*, entre ellos encontramos la definición de que y para que servirá el sistema, cuanto es que va a crecer, etc., este tipo de preguntas, son las que definen la herramienta que se adaptará mejor al sistema que se está construyendo.

Al tomar en cuenta estos aspectos, se debe comprender que no existen herramientas buenas o malas sino simplemente con capacidades distintas, el *Warehouse Builder* es considerado una herramienta para cargas extremas de información, no así el *Cognos 8 BI*, este último es considerado para cargas livianas, con una serie de procesos limitados. Y esta diferencia se enmarca en que el primero es una herramienta integrada casi en su totalidad al Ciclo de vida del *Data Warehouse*. Y el segundo es un conjunto de herramientas que integradas cubren los pasos del ciclo de vida.

Tabla III. Criterios a considerar cuando se crea un *Data Warehouse*

| <b>Criterio</b>                      | <b>Descripción</b>   |
|--------------------------------------|--|
| Conocer el negocio                   | Se debe de reconocer como es que se realizan las actividades actualmente en el negocio, si se utiliza algún tipo de sistema gerencial, cual es y como funciona actualmente para los miembros de la empresa.  |
| La razón de las actividades actuales | Es necesario conocer a detalle la información para identificar las causas de los problemas identificados, además a estos problemas se les debe de dar un análisis extenso para identificar las posibles soluciones al mismo.   |
| Escenarios                           | Una vez que se ha identificado el porque del problema, se puede formular la pregunta pasa observar si el negocio se comporta de una manera diferente. Este tipo de comportamientos se puede prevenir si se estima el modelo de la arquitectura en distintos escenarios. De esta manera se tendrá una certeza de que pasará si, el negocio cambia rápidamente |
| Estudio del modelo de datos origen   | describe cuidadosamente todos elementos de datos que tienen valor para el <i>Data Warehouse</i> , este también provee una definición clara y sin ambigüedades de cada entidad clave de datos, describiendo la forma en que se utilizaran   |
| Estudio del modelo de datos destino  | Describe las entidades que se tomaran en cuenta al momento de poblar el <i>Data Warehouse</i> .  |

|  |   |
|--|---|
| Proceso ETL                                | El diseño y planificación de este proceso muestra la manera en que serán extraídos, transportados para luego ser transformados para que estos aseguren la integridad y calidad de los datos.  |
| Diseño de los almacenes de <i>Metadata</i> | Estos son considerados de uso temporal, son diseñados para el proceso de carga, por medio de ellos se lleva a cabo el proceso de ETL.   |
| Definición de almacenes físicos.           | Se refiere al conjunto de bases de datos que contendrán la información del <i>Data Warehouse</i> . estos dependerán de las necesidades del negocio.   |
| Definición de Data Marts                   | Representan una parte del <i>Data Warehouse</i> . Son diseñados para dar soporte a una unidad de la empresa y para mantener un orden en el <i>Data Warehouse</i> .  |
| Crecimiento                                | Se debe evaluar cual es el crecimiento esperado, al momento de definir los almacenes físicos.   |
| Sistema de Backup y Recuperación           | Aunque el <i>Data Warehouse</i> por si mismo representa, la transformación de datos transaccionales, se debe de realizar un plan de backup y recuperación, debido a que este, también es un sistema de computación como cualquier otro. |

Tabla IV. Características de *Warehouse Builder* de Oracle vrs. *Cognos 8 Bi*.

| <b>Característica</b>                                  | <b>Oracle</b> | <b>Cognos</b> |
|--|---------------|---------------|
| Importar la definición de los datos orígenes           | ☑             | ✘             |
| Diseñar y crear el esquema destino de la base de datos | ☑             | ?             |

|   |                                     |   |
|---|-------------------------------------|---|
| Definición de los datos a ser trasladados y transformación dentro de el origen y el destino | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |
| Asignar dependencias dentro del proceso de ETL.   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |
| Administrar y actualizar las definiciones origen  | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |
| Poner en marcha, actualizar y administrar los esquemas destinos                             | <input checked="" type="checkbox"/> | ?   |
| Diseñar y crear ambientes de herramientas de consulta con fines específicos                 | <input checked="" type="checkbox"/> | Con otra herramienta distinta a la descrita |
| Diseñar y crear un ambiente OLAP.   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |
| Generación de Código DDL  | <input checked="" type="checkbox"/> | Algunos objetos                             |
| Herramientas amigables al usuario   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |
| Centralización de <i>Metadata</i>   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |
| Base de datos propietaria   | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/>         |

## CONCLUSIONES

1. El beneficio obtenido del *Data Warehouse* es proporcional al grado de relación que este posea con las reglas, normas, métricas y metodologías de la empresa. Esto quiere decir que cuanto mas definidos se encuentren los requerimientos y cuando mas estén relacionados con las necesidades de la empresa, se obtendrá un mejor desempeño de la arquitectura que se diseñe en base a los lineamientos definidos, en los requerimientos.
2. El éxito de un proyecto de *Data Warehouse* no es representado por la tecnología con la que se apoye, si no con la calidad del diseño que se elabore, ya que, la tecnología, solamente representa un eslabón en la cadena de la elaboración de este tipo de sistemas y no así el punto clave del sistema.
3. El uso de herramientas de tecnología tales como *Warehouse Builder* y *Cognos 8 BI*, proporcionan un aumento en el rendimiento de las actividades involucradas al momento de crear un *Data Warehouse*, debido a que proporcionan interfaces accesibles a los usuario finales, sobre acciones que presentan complejidad avanzada.

4. La elección de una herramienta esta ligada, proporcionalmente, a la carga y solicitud de datos al que el sistema se encuentra sometido, por lo que si un sistema posee cargas extremas de información, debe hacer uso de una herramientas robustas tales como *Warehouse Builder*. A diferencia de un sistema con una carga pequeña o liviana puede ser manejado con una herramienta como *Cognos 8 BI* que ofrece ventajas en el manejo de ambientes de *Business Intelligence*. Por tanto, cada herramienta cubre las necesidades para las que fue diseñada.

## RECOMENDACIONES

1. Antes de construir o diseñar la arquitectura debe de considerarse las necesidades de la empresa que debe cubrir los servicios del *Data Warehouse*
2. Se debe de identificar si las características del problema que se esta resolviendo encaja en las de un *Data Warehouse*, antes de iniciar un proyecto de la magnitud de este tipo de sistemas
3. Al momento de considerar este tipo de sistemas no se debe de perder de vista el objetivo que guardan la infraestructura, *Metadata*, el *Back Room* y el *Front Room*.
4. Tener una proyección de crecimiento del sistema, para poder realizar una buena planificación de infraestructura.
5. Al momento de elegir una herramienta, evaluar la escalabilidad que esta posee.
6. Se debe evaluar las necesidades que debe cubrir el sistema para identificar la herramienta que cubra adecuadamente los objetivos del sistema.



## BIBLIOGRAFÍA ELECTRÓNICA

1. Alan Perkins, *Data Warehouse Architecture, Blue Print* ,  
[www.Visible.com](http://www.Visible.com)
2. Alex Berson, Stephen Smith, and Kurt Thearling, *An Overview of Data Mining Techniques*,  
<http://www.thearling.com/text/dmtechniques/dmtechniques.htm>
3. *An introduction to Data Mining*,  
<http://www.thearling.com/text/dmwhite/dmwhite.htm>
4. Kurt Thearling, *Understanding Data Mining: It's All in the Interaction*,  
<http://www.thearling.com/text/dsstar/interaction.htm>
5. *Larry Greenfield, Different Aspects of Data Warehouse Architecture*,  
<http://www.dwinfocenter.org/architect.html>
6. *Laura Hadley, Developing a Data Warehouse Architecture*,  
<http://www.users.qwest.net/~lauramh/resume/thorn.htm>
7. Oracle9i Warehouse Builder, User's Guide, Release 9.2, Part No. B10996-01
8. Oracle9i Data Warehousing Guide, Release 2 (9.2), March 2002, Part No. A96520-01
9. Peter Nolan, Editado por Ralph Kimball, *Getting Started and Finishing Well*,  
[http://www.intelligententerprise.com/010507/webhouse1\\_1.jhtml?\\_requestid=123231](http://www.intelligententerprise.com/010507/webhouse1_1.jhtml?_requestid=123231)
10. Ralph Kimball Laura Reeves, Margy Ross, *The Data Warehouse Lifecycle Toolkit*, Wiley Computer Publishing, 1998

11. *Ralph Kimball, Relocating the ODS*,  
<http://www.dbmsmag.com/9712d05.html>
  
12. *Ralph Kimball, STIRRING THINGS UP*,  
[http://www.intelligententerprise.com/db\\_area/archives/1999/992206/warehouse.jhtml?\\_requestid=121169](http://www.intelligententerprise.com/db_area/archives/1999/992206/warehouse.jhtml?_requestid=121169)
  
13. *Ralph Kimball, The Anti-Architect*,  
[http://www.intelligententerprise.com/020114/502warehouse1\\_1.jhtml?\\_requestid=121447](http://www.intelligententerprise.com/020114/502warehouse1_1.jhtml?_requestid=121447)
  
14. *Ralph Kimball, The Data Webhouse Has No Center*,  
[http://www.iemagazine.com/db\\_area/archives/1999/991307/warehouse.jhtml?\\_requestid=130769](http://www.iemagazine.com/db_area/archives/1999/991307/warehouse.jhtml?_requestid=130769)
  
15. *Ralph Kimball, What Does The Central Team Do?*,  
<http://www.dbmsmag.com/9706d05.html>
  
16. *Raph Kimball, Is Hand-Coded ETL the Way to Go?*,  
[http://i.cmpnet.com/intelligententerprise/031118/618rca1\\_1.jhtml](http://i.cmpnet.com/intelligententerprise/031118/618rca1_1.jhtml)
  
17. *Raph Kimball, Meta Meta Data Data*,  
<http://www.dbmsmag.com/9803d05.html>
  
18. *Steve Tracy, Close the Loop*,  
[http://www.intelligententerprise.com/010327/webhouse1\\_1.jhtml](http://www.intelligententerprise.com/010327/webhouse1_1.jhtml)
  
19. The Ken Orr Institute; revised edition, 2000, *Data Warehousing Technology*, <http://www.kenorrinst.com/dwpaper.html>
  
20. Vivek R. Gupta, Senior Consultant, *An Introduction to Data Warehousing*, <http://system-services.com/dwintro.asp>

