



UNIVERSIDAD NACIONAL AUTÓNOMA  
DE MÉXICO

FACULTAD DE CONTADURIA Y ADMINISTRACION

SOPORTE A LA TOMA DE DECISIONES  
CON DATA WAREHOUSING

SEMINARIO DE INVESTIGACION  
I N F O R M A T I C A  
QUE PARA OBTENER EL TITULO DE  
LICENCIADO EN INFORMATICA  
P R E S E N T A  
OSCAR JOSE CONTRERAS HERNANDEZ

ASESOR DEL SEMINARIO: ING. MARIO RODRIGUEZ MANZANERA



MEXICO, D. F.

1997

TESIS CON  
FALLA DE ORIGEN



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



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

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

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

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

---

A mis padres Josefina y Rodolfo

a quienes todo debo e intento entregar un poco de lo  
mucho que sus amorosas manos me han dado siempre.

A mis hermanos Filiberto, Cristina, Juana y Rodolfo

que siempre han sido un brillante ejemplo a seguir,  
pero ante todo un gran apoyo.

A mi tía Guillermina

por ser otra mamá para mí.

A mis amigos, todos

por la confianza, paciencia y cariño brindados.

A todas aquellas personas

que con desinterés han colaborado para que este sueño se haga realidad.

---

---

**Gracias a la Universidad Nacional Autónoma de México  
por ofrecerme lo más valioso que un ser humano puede tener, el conocimiento.**

**Gracias a la Facultad de Contaduría y Administración  
por todas las cátedras recibidas.**

**Gracias a la Dirección General de Servicios de Cómputo Académico y  
especialmente a la Dirección de Cómputo para la Administración Académica  
por los conocimientos y experiencias adquiridas en sus aulas y proyectos, especialmente a  
las personas con las cuales me he honrado colaborar.**

**Gracias al Ing. Mario Rodríguez Manzanera  
por ser guía constante y brindarme su experiencia en todo momento.**

---

## Introducción

*Tomar decisiones es, por una parte, una de las manifestaciones más fascinantes de actividad biológica y, por otra, un asunto con implicaciones aterradoras para toda la raza humana*

C. West Churchman

Debido a la gran competitividad del mercado actual las organizaciones deben buscar todas las ventajas tecnológicas que la Informática pueda ofrecerles, durante mucho tiempo las soluciones informáticas han provisto a las organizaciones de datos que fundamentalmente soportaban la operación principal de la organización, en ocasiones estos datos eran resumidos por los sistemas mismos y la información tomaba algunas características tácticas, pero nunca se ofrecía como información estratégica para la toma de decisiones.

Enmarcada en nuevos paradigmas administrativos como el estudio de los procesos, mejora continua, calidad, grupos de trabajo, etc. la necesidad de información estratégica en línea para la toma de decisiones ha crecido exponencialmente, los sistemas que contienen la información no podrían en la mayoría de los casos soportar las consultas que requiere un análisis exhaustivo para la toma de decisiones, por lo que una solución es tener fuera del sistema en operación otro que permita realizar el proceso de toma de decisiones estratégicas sin afectar los sistemas en operación.

Data Warehousing es una innovación tecnológica relativamente reciente que ofrece todas estas posibilidades de soporte al proceso de toma de decisiones y específicamente la explotación de datos para el análisis de alternativas de decisión.

El desarrollo de la presente investigación presenta un panorama de la Toma de Decisiones como un elemento fundamental de cualquier administración, su importancia y el proceso por el cual el estudio de las alternativas puede hacer una diferencia entre el éxito y el fracaso. Se describe también la evolución de la informática en relación a las decisiones estratégicas y como ha auxiliado a la generación de información valiosa.

Dentro del proceso de toma de decisiones el análisis de alternativas y su evaluación son fundamentales por lo que el estudio de un método de análisis conocido como análisis multidimensional, se presenta como segunda parte de esta investigación, su desarrollo informático y las diferentes alternativas de aprovechamiento que existen en el mercado.

Una de estas alternativas es la utilización del Data Warehouse como gran repositorio de información lista para ser explotada multidimensionalmente para la toma de decisiones. la conceptualización de un repositorio de datos, va más allá y se convierte en un proceso de obtención, transformación, modelado, almacenamiento y explotación de información. Este proceso está descrito la tercer parte de la investigación y es conocido como Data Warehousing, los tipos y arquitecturas posibles, también se encuentran descritas allí.

Esta investigación propone en las partes cuarta y quinta una alternativa para el desarrollo de un proyecto, con los elementos implicados, como personal, definición de requerimientos, elección de arquitectura, entre otros. Dentro de la propuesta de desarrollo se ofrece una metodología, entendiéndola como el procesos necesarios de desarrollo, tales como: el análisis de requerimientos, diseño, construcción, pruebas e implantación. Todo ello específicamente de una arquitectura descrita en la tercer parte de esta investigación como mercado de datos.

Finalmente la presente investigación ante propone un proyecto de desarrollo de Data Warehousing atendiendo una necesidad específica de la U.N.A.M. como lo es la planeación de sus recursos de personal académico, esta necesidad se establece con base en el análisis

---

de una publicación anual de la Dirección General de Asuntos del Personal Académico dependiente de la Coordinación de Programas Académicos.

En conjunto la presente investigación ofrece al lector una panorámica del concepto Data Warehousing, su utilidad dentro de un proceso de Toma de Decisiones, los lineamientos de un proyecto de desarrollo, así como un metodología orientada a desarrollos muy específicos o de prueba y prototipo. También ofrece un ejemplo de utilización de lo antes mencionado, muy descriptivo aunque no desarrollado completamente, para conocer los ámbitos de aplicación tanto de los lineamientos de proyecto como de la metodología de desarrollo.

Finalmente unas breves conclusiones permitirán al lector formarse una opinión en conjunto del concepto y una de sus posibles utilizaciones obteniendo una serie de beneficios para una organización no lucrativa como lo es la U.N.A.M..

## Contenido

<b>Introducción</b>	<b>I</b>
<b>Contenido</b>	<b>IV</b>
<b>1 Toma de decisiones</b>	<b>1 - 1</b>
<b>1.1 La importancia de la toma de decisiones en las organizaciones</b>	<b>1 - 1</b>
1.1.1 ¿Qué significa la toma de decisiones?	1 - 1
1.1.2 Proceso para la toma de decisiones	1 - 2
1.1.3 Naturaleza de la toma de decisiones	1 - 4
1.1.4 Actualidad de las organizaciones en relación a la toma de decisiones	1 - 5
<b>1.2 La toma de decisiones auxiliada por sistemas</b>	<b>1 - 6</b>
1.2.1 Evolución histórica	1 - 6
1.2.2 Actualidad	1 - 9
<b>2 Análisis Multidimensional</b>	<b>2 - 1</b>
<b>2.1 Visión de los datos del tomador de decisiones</b>	<b>2 - 1</b>
<b>2.2 Visión multidimensional de la información</b>	<b>2 - 2</b>
<b>2.3 Procesamiento analítico en línea (OLAP)</b>	<b>2 - 6</b>
<b>2.4 Arquitectura del Análisis Multidimensional</b>	<b>2 - 7</b>
2.4.1 Depósito de datos multidimensional y relacional	2 - 7
2.4.2 Depósito de datos multidimensional	2 - 8
2.5.3 Depósito de datos relacional	2 - 10



<b>3 Data Warehousing</b>	<b>3- 1</b>
<b>3.1 Definición de Data Warehouse</b>	<b>3- 1</b>
<b>3.2 Definición de Data Warehousing</b>	<b>3- 3</b>
<b>3.3 Arquitectura del Data Warehousing</b>	<b>3- 5</b>
3.3.1. Data Warehousing ligero.	3- 8
3.3.2. Data Warehousing virtual.	3- 8
3.3.3. Data Warehousing corporativo.	3- 9
3.3.4. Data Warehousing con mercados de datos.	3- 10
<b>3.4 Aplicación de las arquitecturas.</b>	<b>3- 11</b>
<b>4 Proyecto para la Toma de Decisiones utilizando Data Warehousing</b>	<b>4 - 1</b>
<b>4.1 Selección de la estrategia.</b>	<b>4 - 1</b>
4.1.1 Enfoque de lo general a lo particular.	4 - 1
4.1.2 Enfoque de lo particular a lo general.	4 - 2
4.1.3 Enfoque combinado.	4 - 3
<b>4.2 Desarrollo de los objetivos organizacionales.</b>	<b>4 - 5</b>
<b>4.3 Estudio de los requerimientos.</b>	<b>4 - 7</b>
<b>4.4 Equipo de trabajo</b>	<b>4 - 11</b>
<b>4.5 Justificación de proyecto</b>	<b>4 - 12</b>

---

<b>5 Metodología para el desarrollo de Data Warehousing</b>	<b>5- 1</b>
Nota aclaratoria sobre la metodología	5- 1
<b>5.1 Análisis de Requerimientos</b>	<b>5- 2</b>
5.1.1 Requerimientos del usuario final.	5- 2
5.1.2 Requerimientos de la arquitectura.	5- 4
5.1.3 Requerimientos importantes del diseño	5- 8
<b>5.2 La arquitectura Tecnológica</b>	<b>5- 10</b>
5.2.1 Análisis general de Fabricantes	5- 11
5.2.2 Proveedores de soluciones versus proveedores de componentes	5- 12
5.2.3 Estructura de fabricante	5- 14
<b>5.3 Diseño</b>	<b>5- 15</b>
<b>5.4 Construcción.</b>	<b>5- 17</b>
<b>5.5 Implantación</b>	<b>5- 19</b>
<b>6 Ante- propuesta de Proyecto para Data Warehousing</b>	<b>6- 1</b>
6.1 Estrategia General	6- 1
6.2 Descripción del problema	6- 2
6.3 Objetivos de la organización (D.G.A.P.A.)	6- 4
6.4 Justificación del proyecto	6- 5
6.4.1 Beneficiados directos del proyecto	6- 6
6.5 Análisis de requerimientos	6- 7
6.5.1 Requerimientos de los procesos	6- 7
6.5.2 Requerimientos de los datos	6- 8
6.6 Elección de plataforma y Diseño	6- 10

**Conclusiones**-----**VIII**

**Anexo**

**Bibliografía**-----**X**

**Hemerografía**-----**XI**

## *1 Toma de decisiones*

### *1.1 La importancia de la toma de decisiones en las organizaciones*

#### *1.1.1 ¿Qué significa la toma de decisiones?*

La toma de decisiones es la selección de un curso de acción entre diferentes alternativas, los administradores consideran la toma de decisiones su trabajo más importante<sup>1</sup>, lo cierto es que toda persona que deba elegir entre dos o más mecanismos de acción para lograr objetivos esta tomando una decisión.

Necesariamente la toma de decisiones debe ser un proceso racional. Para que este proceso sea fundamentalmente racional los tomadores de decisiones deben comprender claramente los cursos de acción, es decir, las diferentes alternativas, para ello, deben considerar grandes volúmenes de información así como los mecanismos involucrados para analizarlos detalladamente y así evaluar dentro de estas alternativas la más viable en términos de los objetivos planteados.

Las decisiones planteadas con una completa racionalidad no existen, para cada decisión existe siempre asociada una cantidad de incertidumbre que puede ser valuada de muchas formas: costo, porcentaje, etc.; además de eso, existe el inconveniente de poder reconocer absolutamente todas las alternativas de acción que pueden existir para el logro de uno de los objetivos planteados. Esto se complica si el curso de acción a elegir implica que

---

<sup>1</sup> Harold Koontz y Heinz Weihrich, Administración, p. 189.

la meta sea algo jamás antes logrado por lo que tanto las técnicas de análisis como su aplicación pueden resultar insuficientes o poco prácticas.

La solución es conseguir tener la certeza más amplia que las decisiones son las correctas o las que tienen menos riesgo de fallo, este término se le conoce como racionalidad limitada<sup>2</sup> o racionalidad circunscrita, es decir las limitaciones de información, tiempo y certeza limitan la racionalidad del tomador de decisiones por lo que debe alcanzar la mejor de las soluciones con las limitantes que ya tiene.

### ***1.1.2 Proceso para la toma de decisiones***

El proceso de la toma de decisiones necesariamente implica varios pasos en donde cada uno de ellos puede significar el éxito o fracaso de las decisiones que resulten de él. A continuación se detallan estos pasos:

#### **Búsqueda de alternativas.**

El primer paso para tomar una decisión es buscar diferentes alternativas. Es importante tener la capacidad de desarrollar las diferentes alternativas, esto significa ser capaz de documentar con cifras, ventajas, desventajas, etc. las diferentes posibilidades de acción. El eliminar o descartar alternativas desde un principio puede ser tan riesgoso como evaluarlas equivocadamente, en muchas ocasiones la mejor alternativa es desechada por tener inconvenientes que no resultaban tan significativos, tales como el costo o el tiempo de operación.

Saber identificar estas alternativas requiere de muchos elementos de conocimiento, la evaluación o en ocasiones una pre - evaluación de estas determina con mayor certeza su grado de factibilidad.

---

<sup>2</sup> Ibid, p. 200.

### Evaluación de alternativas.

Una vez encontradas las alternativas de decisión el siguiente paso consiste en evaluarlas, es decir ponderar las variables específicas de cada una de las posibilidades. Es aquí donde altos volúmenes de información correctamente procesada pueden dar amplias ventajas competitivas a los tomadores de decisiones, donde los sistemas que permitan el cruzamiento de gran cantidad de variables en estadísticas históricas que generen comportamientos de las variables en cuestión son de gran utilidad.

En la evaluación de alternativas se utilizan, al menos uno, de los siguientes tres métodos:

- Análisis de factores cuantitativos y cualitativos
- Análisis marginal
- Análisis costo - beneficio

### Selección de una alternativa

Una vez evaluadas las alternativas utilizando los diferentes métodos de análisis mencionados, la decisión final necesita ser valorada en términos de la experiencia del tomador de decisiones, la experiencia es probablemente la fuente más rica de intuición para un tomador de decisiones, de esta experiencia es necesario rescatar las razones fundamentales de los errores y éxitos del pasado para aplicar con criterio estas en la toma de una nueva decisión sobre asuntos distintos o similares.

Aprender de la experimentación es hoy en día un criterio de selección de decisiones demasiado caro, aunque es por tanto mucho más acertado, los errores son corregidos necesariamente sobre nuevos elementos de investigación que son provistos de fuentes rápidas y de análisis exhaustivos.

La selección de una alternativa debe considerar tanto la experiencia como la confianza en la información obtenida de análisis y experimentación tales como prototipos y otras innovaciones.

### 1.1.3 Naturaleza de la toma de decisiones

La toma de decisiones responde a la necesidad de elegir uno o más cursos de acción para el logro de objetivos, por lo tanto, la toma de decisiones es parte integral de la planeación, ya que se compromete con la congruencia de otros planes con fines distintos, lo antes mencionado necesariamente implica el análisis de otras variables de aquellos otros planes y nuevos impactos que deben ser analizados por lo que el proceso de la toma de decisiones se puede convertir especialmente, en la etapa de evaluación de alternativas, en un muy complejo conjunto de análisis que se puede ir incrementando exponencialmente.

Hay que considerar que no siempre es posible tomar una decisión en un momento dado sobre la ejecución de un plan específico, también la toma de decisiones debe responder a los problemas que una organización enfrenta sin haberlos previsto en absoluto o a los que se pueden encontrar dentro de su plan de contingencia, estos problemas carecen de una estructura planificada y casi nunca conllevan a una decisión previamente programada. (Ver fig. 1.1)

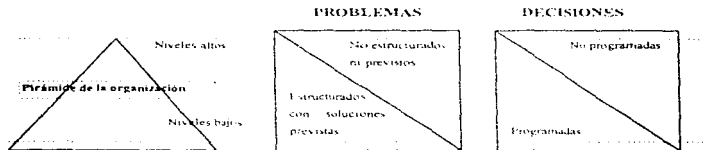


fig. 1.1

Como se muestra en la figura 1.1 dentro de la pirámide organizacional los niveles altos toman gran cantidad de decisiones, básicamente sobre problemas no estructurados ni previstos, por lo tanto sus decisiones no son previamente programadas, los niveles operativos por el contrario, requieren de una toma de decisiones basada en la operación rutinaria y sobre los problemas programados y claramente identificados de la organización.

#### ***1.1.4 Actualidad de las organizaciones en relación a la toma de decisiones***

Hoy por hoy las organizaciones viven en mercados donde la competencia es crecientemente global, hoy es probable encontrar numerosas empresas compitiendo en un mismo rubro, siendo algunas de ellas, no nacionales o probablemente filiales de una firma mundial. La competencia en general, parece mucho más agresiva, las repercusiones de una decisión pueden tomar tintes fatalistas y de alto riesgo con asuntos que en el pasado eran rutinarios.

El valor asociado a la correcta toma de decisiones es mucho más alto, el costo de oportunidad con la globalización es también mayor, la intuición y "la prueba y el error" son sustituidos necesariamente por métodos más efectivos donde el riesgo de una decisión sea menor, donde un cúmulo de información almacenada sobre diversos comportamientos de las variables críticas para la organización sean correctamente analizadas.<sup>1</sup>

---

<sup>1</sup> McDermontt, Karl. "Muchos datos, poca información útil para análisis y toma de decisiones". Soluciones Avanzadas, junio 1996, p. 62.



## ***1.2 La toma de decisiones auxiliada por sistemas***

### ***1.2.1 Evolución histórica***

Durante la primera mitad del siglo, no se soñaba con una computadora como hoy las conocemos. Si bien las máquinas de cálculo mecánico eran ya una sofisticación a la que pocas organizaciones tenían acceso, los negocios sustentaban sus datos en grandes libros contables que un ejército de laboriosos expertos en la materia, mantenían lo más al día posible, su eficiencia se medía en estas actualizaciones y en poder consolidar en una, dos o tres semanas, después de un corte requerido la información vaciada en estados financieros que al llegar a los escritorios de los tomadores de decisiones reflejaban, con poca certidumbre los datos reportados.

Las decisiones eran tomadas con algunos elementos de información y un mucho de intuición y experiencia en el negocio. La crisis de 1929 e inicio de los años treinta refleja un poco esta falta de certidumbre en la orientación general de los negocios, la toma de decisiones era en parte a base de "prueba y error", durante los años veinte al igual que en la posguerra del medio siglo esta forma de conducción no resultaba deficiente para los negocios norteamericanos.

A finales de los 40's y como producto de una necesidad militar, se desarrollaron las primeras computadoras que eran significativamente más poderosas que las calculadoras mecánicas, pero estaban destinadas a propósitos específicos y su funcionalidad era muy limitada. El costo tanto de equipo como de elementos asociados era probablemente el impedimento más importante para su uso comercial, las capacidades que se requerían para mantenerlas en operación implicaban recursos abundantes de todo tipo

Al inicio de la década de los 50's la principal característica de la computadora era el ser una máquina para computar, es decir, para obtener resultados en investigaciones específicas en diversos campos principalmente militares y científicos. Poco a poco, los precios se reducen, los requerimientos de mantenimiento bajan, las computadoras se hacen accesibles a las grandes empresas, la necesidad de tener el control de diversos aspectos del negocio llevan las computadoras de las universidades y de los grandes centros de investigación a los enormes corporativos transnacionales. Las computadoras comenzaron a ser las herramientas más útiles para administrar negocios, y lo son hasta nuestros días.

Al final de la década de los 50's se dió un avance muy importante en este sentido, conscientes de la importancia estratégica del uso de la computadora como herramienta administrativa, se propició la invención de un lenguaje que permitiera contener una variedad de tareas que podía resolver la máquina sin tener una modificación estructural a la arquitectura; surge la herencia que hasta nuestros días forma parte importante de muchos de los grandes sistemas administrativos el lenguaje: COBOL (Common Oriented Business Language)

Para los grandes directivos COBOL se convierte en un elemento que permite a sus colaboradores operativos generar los mismos estados financieros y algunos reportes relativamente específicos con mayor facilidad y mejores tiempos de entrega, pero la constante en la información era la gran dificultad para convertir los datos almacenados, en reportes útiles para al toma de decisiones.

Con la posibilidad de almacenar grandes volúmenes de información, la computadora que solo calculaba se vuelve además un gran almacén de datos, se mejora la forma de procesar, aparece la multitarea y los grandes sistemas operativos, la gente se especializa en el recién aparecido término francés: *Informática*. Se logran diversos avances en diferentes campos de la industria que siempre eran rápidamente consumidos por los requerimientos corporativos de información administrativa.

Aún con estas ventajas de información, su confiabilidad para analizar los resultados era aún muy limitada o nula, los sistemas eran herramientas de control y administración pero no de análisis de información, la toma de decisiones se sustentaba en gran parte en el personal mismo, existía un gran contacto personal entre el vendedor y su clientes, lo que generaba confianza en las decisiones de los directivos basados en este conocimiento y no tanto en los reportes programados en COBOL.

Durante los 70's las grandes computadoras procesan cada vez más volumen y más rápido, las bases de datos evolucionan al modelo relacional, aparecen los procesos de actualización de bases de datos al momento en que la información ingresa en la compañía gracias a avances en telecomunicación entre terminales y grandes computadoras y entonces sucedió que las grandes compañías empezaron a depender notablemente de la labor del personal de informática y de las grandes computadoras. Apareció entonces una computadora con capacidades reducidas, que no requería infraestructura asociada como temperatura, espacio, etc., que soportaba un usuario y de precio muy reducido, era la computadora personal.

Hace dos décadas la cantidad de información que las diversas organizaciones habían acumulado desde finales de los 50's era ya un impresionante almacén de datos que poco decía del comportamiento real del mercado en todos esos años y de cualquier variable que se pudiese cuestionar de la organización, el problema se agravó cuando algunos de los directivos optaron por incorporar computadoras personales para "esos reportes que el sistema no puede generar" y con ellos obtener un análisis de información útil para la toma de decisiones, fuera del sistema centralizado.

Al inicio de los 80's los MIS (Management Information Systems o Sistemas de Información para la Administración) representaron de alguna forma una colección más estructurada de la abundante información que en ese momento las bases de datos relacionales comenzaron a almacenar, estos sistemas se popularizaron rápidamente, pero se caracterizaron por la poca flexibilidad del enlace entre los datos y los resultados reportados

es decir los reportes programados no podían ser modificados sin una modificación estructural al sistema por sí mismo. Otras características interesantes de los MIS son:

- Se centra en tareas estructuradas y decisiones rutinarias
- Es importante el almacenamiento de la información
- Los tomadores de decisiones solo tiene acceso a las interfaces de los datos
- Se desarrollo con expertos en cómputo

### ***1.2.2 Actualidad***

Hoy en día, las organizaciones siguen almacenando grandes volúmenes de datos que en realidad les representan poca información, a esto es necesario añadir la creciente distribución de información en diversas plataformas que en ocasiones no interoperan completamente. Surge entonces la necesidad no solo de *agregar* valores a una o varias bases de datos sino además de *resumirlos* en elementos mucho más significativos y que no solo permitan, como hasta ahora, generar datos para los niveles operativos de la organización, por el contrario, elaborar los complejos reportes requeridos por los niveles táctico y estratégico, aún cuando estos nunca fueron planeados como tales.

El concepto de OLTP (On Line Transaction Processing, Procesamiento de Transacciones en Línea) se asocia a las diferentes aplicaciones que soportan la diaria operación del negocio, es decir los sistemas de nómina, facturación, inventarios, ventas, etc. que son consideradas de misión crítica, cada una de las transacciones deben completarse o tener claro el fin de la transacción. Los OLTP deben caracterizarse por:

- Tiempos muy cortos de respuesta por transacción.
- Los datos deben permanecer íntegros, correctos, consistentes, evitando redundancia, expedidos, etc.
- Se requiere una alta seguridad en datos y procesos.

- Permanecer en línea todo el tiempo.

Los OLTP son sistemas fundamentalmente listos para el funcionamiento operativo de la organización, en la actualidad estos sistemas continúan funcionando en sistemas mainframe o superservidores de alto costo pero de enormes capacidades; esto ha ocasionado que por ejemplo, un departamento con el suficiente presupuesto arme una pequeña red local, adquiera una serie de computadoras personales, un servidor y comience a elaborar sus aplicaciones específicas, el caso extremo es cuando un aplicación de interés corporativo reside en una sola computadora personal,

Los DSS, (Decision Support Systems, Sistemas de Soporte de Decisiones) permiten una exploración más rápida y a detalle utilizando hojas de cálculo y otras herramientas de graficación, un DSS completo soporta también, preguntas del tipo "¿qué tal si...?" y buscar por sí mismo algunas correlaciones de los datos almacenados.<sup>4</sup>

Los DSS son sistemas Cliente/Servidor que no se encuentran generalmente dentro del esquema de misión crítica de la organización por lo que cierta redundancia tiene que soportarse para su funcionamiento, así como un cierto margen de diferencia entre los datos que almacenan las bases en operación y el DSS por sí mismo. Un DSS es una prometedora evolución de los MIS, los DSS se caracterizan por:

- Elaborar como parte misma del sistema las estructuras de consulta
- Requieren criterio organizacional y no expertos en cómputo para su desarrollo
- Los tomadores de decisiones tiene acceso directo a los datos
- Es muy importante la manipulación de la información

Los DSS son sistemas que auxilian a la toma de decisiones en el segundo paso del proceso esto en la evaluación de alternativas de decisión, por ello deben ser sistemas que

---

<sup>4</sup> Orfali, Harkey y Edwards, The Essential Client/Server, p. 204

---

procesen rápidamente grandes cantidades de información que probablemente no esta estructurada dentro de un mismo sistema en operación.

Actualmente los EIS (Executive Information Systems, Sistemas de Información Ejecutiva) son considerados mucho más poderosos, fáciles de utilizar pero también son más caros, las diferencias son poco claras entre los DSS y los EIS, en muchas ocasiones son considerados lo mismo, lo cierto es que a ambos se les asocian otros conceptos tales como: análisis multidimensional, procesamiento analítico en línea, minería de datos y agentes inteligentes.<sup>5</sup>

---

<sup>5</sup> Ibid. p. 205

## ***2 Análisis Multidimensional***

### **2.1 Visión de los datos del tomador de decisiones**

El proceso de la toma de decisiones por parte de los altos mandos de una organización enfrenta dos retos fundamentales: operarla de manera eficiente en relación al logro de objetivos y tomar decisiones sobre la planeación del futuro de la organización. En este proceso, específicamente durante la evaluación de alternativas de acción y la creación de nuevas opciones que permitan horizontes innovadores de planeación: cualquier ventaja tecnológica puede significar adelantarse de manera significativa a los competidores o prevenir un posible problema tomando mejores decisiones en mejores tiempos.

Los buenos tomadores de decisiones saben que el futuro pertenece a quienes pueden verlo y llegar ahí primero. Por tal razón no sólo comprenden lo que está pasando en la organización, sino también qué va a suceder. Los tomadores de decisiones desean extraer los datos adecuados de manera intuitiva, con la mínima inversión de tiempo y frustración. Una vez que extraen los datos correctos del sistema, los analizan, sintetizan y consolidan en información, para después emplearla a fin de continuar con el proceso de la toma de decisiones.<sup>1</sup>

En el marco de referencia de la plataforma que de soporte para una toma de decisiones adecuada, el procesamiento analítico se utiliza principalmente como el uso de verificación de hipótesis organizacionales. Es decir, el usuario crea una hipótesis (una cuestión organizacional) y accede datos para verificarla. Por lo regular el análisis de los datos es iterativo, una hipótesis conduce a otra, hasta que se deduce un conjunto claro de

<sup>1</sup> Harjinder S. Gill y Prakash C. Rao, Data Warehousing, p. 111-115

alternativas y recomendaciones de acción potenciales, esto se identifica perfectamente dentro del proceso de toma de decisiones como la búsqueda de alternativas y la evaluación de estas.

Durante la actividad de análisis iterativo, los usuarios experimentados "descubren" en ocasiones relaciones insospechadas entre parámetros y variables organizacionales, por ejemplo, la relación de variables como la edad y el sexo de los clientes en relación a cierto producto y área de venta. De ahí que, aunque generalmente el procesamiento analítico no se considera limitado a la modalidad de descubrimiento del soporte de decisiones, se presta también al descubrimiento mediante hallazgos fortuitos y la observación del usuario.

Este proceso de análisis iterativo de las relaciones entre las diversas variables organizacionales mediante sistemas automatizados, tiene básicamente dos vertientes dentro de los DSS, (mencionados anteriormente) que son por un lado, permitir a los usuarios dirigir estas búsquedas de información explotando su intuición y experiencia, y por otro lado, la búsqueda de estas relaciones por métodos también automatizados donde sólo se reportan al usuario los resultados de los hallazgos, a este proceso se le conoce dentro de los DSS y los ESS como Minería de Datos y Agentes Inteligentes . (Estos temas no serán tratados como parte del presente estudio).<sup>2</sup>

## ***2.2 Visión multidimensional de la información***

Necesariamente la forma de ver la información por parte de los diferentes niveles jerárquicos de la organización implica una visión totalmente distinta de los datos de los sistemas operacionales, para los niveles de toma de decisión estratégica la información importante es una serie de hechos multidimensionales. Se encuentran relacionados y

---

<sup>2</sup> Ibid., p. 193-200.



regularmente son jerárquicos; por ejemplo los datos de ventas, los datos de inventarios y los pronósticos de presupuestos están interrelacionados y dependen entre sí.

En la práctica, para predecir el éxito y el comportamiento de una organización es necesario no solo analizar sus perspectivas desde un sólo punto de vista y desde la visión del momento del análisis, es decir, los datos históricos propios de la organización así como los datos del entorno principalmente económico cobran un valor impresionante en la toma de decisiones sobre la planeación de acciones futuras, y sobre la presupuestación de estas.

Para facilitar un análisis complejo, el análisis multidimensional presenta una visión organizacional sencilla de los datos. Un usuario puede acceder la información de las diferentes variables organizacionales importantes para el análisis por ejemplo: ingresos por departamento y región de los últimos tres trimestres, para un conjunto dado de productos o servicios, en un rango de precios y hasta horarios. Los resultados se pueden pivotear o girar para cambiar los ejes y la perspectiva. (Ver figura 2.1)

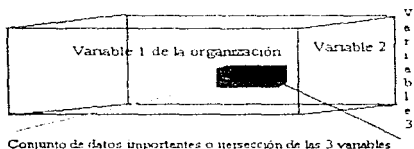


Figura 2.1

El procesamiento multidimensional se usa para entender lo que está sucediendo en la organización y también un análisis del tipo ¿qué tal si...? o ¿y ahora qué si...? El ámbito

del procesamiento es por lo regular un análisis más complejo que dos o tres dimensiones añadiendo datos históricos y estadísticas de todo tipo.<sup>3</sup>

Los usuarios organizacionales requieren navegar por las dimensiones profundizando y obteniendo resúmenes a lo largo de los elementos de una dimensión, o penetrar a través de las dimensiones para ver otras perspectivas. El modelo de estrella permite modelar muchas dimensiones que los usuarios desean intersectar en diferentes consultas. (Ver figura 2.2)

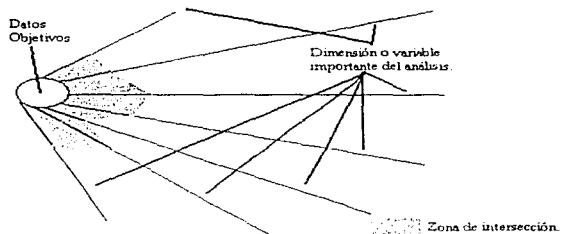


Figura 2.2

La elección del enfoque, procesamiento analítico multidimensional incorrecto podría provocar frustración en el usuario organizacional, un impacto negativo en la productividad y una pérdida de ventaja estratégica, un error en el procesamiento de la información no solo provocaría un impacto negativo en la moral del tomador de decisiones sino un muy grave riesgo.

<sup>3</sup> Livingston, Glen y Rumsby, Bob; Database Design for Data Warehouses: The Basic Requirements, Planning and Designing the Data Warehouse, p. 179-183.

Por lo que, tratar de elaborar reportes o consultas con los elementos de relación antes requeridos sin la ayuda de un sistema de cómputo que auxilie los cálculos y conteos sería materialmente imposible. Un sistema de cómputo debe ofrecer también capacidades más específicas que también se requieren como:

- Posiciones al principio o al final de la lista
- Promedios móviles
- Tasas de crecimiento
- Cálculos de interés financiero
- Tasas internas de recuperación y depreciación
- Conversiones monetarias
- Otras funciones estadísticas.

El tomador de decisiones le interesa la visión multidimensional de los datos en un nivel de desempeño aceptable. Al personal de tecnología de la información le interesa dónde, cómo se almacenan y acceden los datos, para asegurar un desempeño aceptable y una administración eficaz, sin embargo, coinciden en reconocer que el diseño inicial y la actividad de configuración son conducidas por el modelo de información requerida.

El modelo de información requerida se relaciona directamente al diseño de todo el proyecto de soporte a la toma de decisiones sustentado en los siguientes elementos:

- Seleccionar la función empresarial
- Identificar los valores numéricos
- Determinar las dimensiones
- Definir el modelo lógico
- Cargar el depósito de datos multidimensional

### **2.3 Procesamiento analítico en línea (OLAP)**

Al procesamiento multidimensional se le conoce también como OLAP (On Line Analytical Processing, Procesamiento Analítico en Línea). El cual se apoya en una visión multidimensional de los datos organizacionales con los siguientes elementos:

- Presenta una visión multidimensional lógica del análisis de datos que se hace en un repositorio de datos. La visión es independiente de como se almacenan los datos.
- Soporta la consulta interactiva y el análisis de los datos. Por lo regular la interacción es iterativa, lo cual incluye una profundización en niveles cada vez mas detallados o el acceso a niveles superiores de resumen y adición.
- Ofrece opciones de modelado analítico, incluyendo un motor de cálculo para obtener proporciones, desviaciones, etc., que comprende mediciones de datos numéricos a través de muchas dimensiones.
- Crea resúmenes y adiciones (también conocidas como consolidaciones), jerarquías, y cuestiona todos los niveles de adición y resumen en cada interacción de las dimensiones.
- Maneja modelos funcionales de pronóstico, análisis de tendencias y análisis estadístico.
- Recupera y exhibe datos tabulares en dos o tres dimensiones, cuadros y gráficas, con pivoteo fácil de los ejes.

El pivoteo es fundamental ya que los usuarios organizacionales necesitan analizar los datos desde perspectivas diferentes; y el análisis desde una perspectiva conduce a otra cuestión que se va a examinar desde alguna otra perspectiva.

El procesamiento multidimensional no puede estar completo si carece de una poderosa fuente de datos, por lo que requiere de una interminable y abundante fuente de

datos.<sup>4</sup> En el siguiente capítulo se describe las opciones para fuentes de datos conocidas como Data Warehouses.

## ***2.4 Arquitectura del Análisis Multidimensional***

OLAP es una opción clara para la implementación de un DSS, dentro del proceso de la toma de decisiones los aspectos de búsqueda y evaluación de alternativas están plenamente cubiertos. El componente OLAP dentro de un DSS implica acceder los datos directamente desde el repositorio de datos. La visión multidimensional requerida puede llegar de dos formas dependiendo de la arquitectura utilizada para el almacenamiento de la información. En ambos casos las consultas reflejan después del proceso necesario, una estructura multidimensional.

OLAP requiere la presentación multidimensional y lógica de los datos del repositorio de información o de la base de datos hacia el tomador de decisiones, sin importar como y donde están almacenados los datos, lo que implica un esquema Cliente/Servidor. Dentro de las opciones de tecnología para el depósito de datos, existen dos formas populares: el depósito de datos multidimensional y el depósito de datos relacional.

### ***2.4.1 Depósito de datos multidimensional y relacional***

El depósito de datos relacional se apega al modelo de datos relacional. Las bases de datos relacionales almacenan los datos en tablas con registros con llaves, y los datos de acceder mediante un lenguaje común, el SQL (Structured Query Language, Lenguaje de Consulta Estructurada). Por otra parte, los depósitos de datos multidimensionales, almacenan los datos de manera lógica en arreglos. Como no hay un modelo

---

<sup>4</sup> Harjinder S. Gill, Op. Cit p 217-222

multidimensional común o acordado, no hay un método común o estándar para el acceso a los datos. Hay algunos productos cuyo motor cuenta con API (Application Program Interface, Interfaz de los Programas de Aplicación) o con una hoja de cálculo como proceso frontal; aunque la mayoría tiene métodos de acceso de propietario como procesos frontales. Además, con motores multidimensionales incorporados y una gran funcionalidad OLAP.

Con un depósito de datos relacional, el repositorio de información requiere tener una dimensión muy grande. El tamaño del depósito aumenta por el uso de índices y técnicas de retiro de normas para lograr un desempeño aceptable en consultas multidimensionales. El gran tamaño pudiera señalar la necesidad de una tecnología para consulta y carga paralelas. Con el depósito multidimensional, por lo general el tamaño es limitado, pero utiliza tecnología de compresión, como la compresión de matriz dispersa para almacenar mas datos en menos espacio.<sup>5</sup>

#### ***2.4.2 Depósito de datos multidimensional***

Los datos se almacenan de manera lógica en arreglos, utilizando un enfoque conocido como de hiper-cubo. En este los objetos con tres o más dimensiones se describen como los lados planos y cada dimensión con ángulos rectos en relación con los demás. Una sola celda en el cubo puede almacenar, por ejemplo, valores de con referencia a tres dimensiones. Un celda puede almacenar una variable importante para la organización, como la de ventas, además de por tienda, en un periodo de tiempo específico dado para un producto o servicio seleccionado

Los datos se organizan de acuerdo con la visión empresarial de los mismos y, como regla, se almacenan en forma resumida o agregada, lo que en ocasiones se denomina

---

<sup>5</sup> Raden, Neil; Choosing the Right OLAP Technology, Planning and Designing The Data Warehouse; p.202-204.

consolidar o modelar datos organizacionales en rangos estadísticos útiles. El índice es más pequeño, dando por resultado una respuesta mucho más rápida a consultas complejas. Debido a que los valores se almacenan en arreglos, la actualización del valor no afecta al índice. Esta característica facilita la implementación de aplicaciones de actualización, como las de pronósticos y presupuestación.<sup>6</sup>

Las funciones y elementos principales que se ofrecen para auxiliar la toma de decisiones son las siguientes:

- **Rápida respuesta a consultas de cómputo intensivo** tales como escenarios "¿Qué pasa si...?". La respuesta rápida no interrumpe el proceso racional y de análisis.
- **Actualización interactiva de la base de datos multidimensional** para permitir aplicaciones de pronósticos, planeación a futuro y presupuestación.
- **Explotación de relaciones ricas** entre los elementos o valores de las dimensiones para describir relaciones insospechadas.
- **Un poderoso motor de cálculo y análisis comparativo:** posiciones, comparaciones, porcentajes de clase, máximo, mínimo, promedios, promedios móviles, comparaciones entre periodos, y otros.
- **Cálculos cruzados** entre dimensiones para aplicaciones orientadas a hojas de cálculo como las declaraciones de pérdidas y ganancias.
- **Ampliación de las funciones básicas de las funciones definidas por el usuario o exploración de funciones de modelado incorporadas.**
- **Potentes funciones estadísticas y financieras**
- **Inteligencia de manejo del tiempo** año calendario y periodos diversos.
- **Pivote, tabulación cruzada profundización, niveles de resumen para una o varias dimensiones,** y otras funciones poderosas de navegación.

La administración general y la administración de sistemas requieren de la siguiente:

---

<sup>6</sup> Ibid. p. 210-214.

- El modelado inicial de datos en donde son consideraciones clave elegir las dimensiones correctas y su garantía, prever como se accederán los datos y seleccionar los filtros apropiados para cargar los datos desde el repositorio de información.
- Transferencia periódica y actualizaciones en bloque, debido a que las actualizaciones en incrementos son un reto y casi imposibles mientras que la base de datos esta en uno.
- Adición, resúmenes y pre-cálculo durante el proceso de carga.
- Capacitación en aun tecnología diferente y uso de nuevas habilidades.
- Escritura de nuevas aplicaciones en lenguaje de propietario para ampliar y mejorar los procesos frontales comunes de la base de datos.
- Los efectos colaterales de depositar datos a niveles mas gruesos de granularidad ( resumen, adición, y datos precalculados y derivados) significa que la profundizaciones no alcanzan el nivel de detalle.
- La seguridad y acceso están disponibles a niveles altos; no hay privilegios base en el usuario o controles a nivel de subconjuntos.
- Los cambios en las estructuras dimensionales requieren reorganizar la base de datos multidimensional; son limitadas las operaciones de respaldo y restauración de alta disponibilidad.
- La necesidad de proceso frontales especializados limita las opciones. Las ampliaciones al proceso frontal de aun base de datos multidimensional no se trasladan con facilidad a otra base de datos multidimensional.

### ***2.5.3 Depósito de datos relacional***

En este enfoque en servicios OLAP los datos se almacenan en tablas relacionales y pueden tener un tamaño de cientos de gigabytes. Aunque los datos de almacenan en forma relacional (columnas y renglones), se presentan al usuario empresarial en forma de dimensiones empresariales. A fin de "ocultar" la forma de depósito, se crea una capa semántica de datos. Esta capa ubica las dimensiones para las tablas relacionales.



También se crean datos adicionales para cualquier resumen o adición con el fin de mejorar el tiempo de respuesta. Todos estos datos se almacenan en la base relacional creando generalmente, otros depósitos de esos mismos pero mucho más especializados. La solución general del repositorio de información de forma relacional, necesita de estos para su mejor mantenimiento y administración lo que incrementa de forma considerable su tamaño.<sup>7</sup>

Las funciones y elementos principales que se ofrecen para auxiliar la toma de decisiones son las siguientes:

- Una visión multidimensional de los datos relacionales.
- Soporte de la jerarquía dimensional.
- Funciones de cálculo, estadísticas y financieras, con ampliaciones de usuario.
- Profundización al nivel de detalle.
- Elección de herramientas de proceso frontal.
- Aprovechamiento de las inversiones existentes para respaldo y restauración de las administración de base de datos, y subdivisión de la base de datos para análisis individuales.
- Niveles múltiples de seguridad con privilegios de usuarios.

La administración general y la administración de sistemas de este enfoque requieren de lo siguiente:

- Que no haya carga inicial o actualizaciones periódicas del servidor OLAP.
- Una de los procesos estándar existentes para respaldo, restauración y seguridad.
- Diseño del "modelo dimensional" para proporcionar la visión empresarial de los datos.

---

<sup>7</sup> Ibid. p. 215-217

- Administración, sincronización y mantenimiento de todos los datos nuevos dentro de la solución general del repositorio de información.
- Vigilancia de la utilización para afinar el desempeño. Dicha afinación puede afectar el modelo de base de datos, los segmentos, o los niveles de adhesión y resumen. La afinación es complicada. El retiro de normas puede incrementar tanto la falta de datos como el tamaño del repositorio, requiriendo así de búsquedas más largas, más operaciones de entrada y salida de disco y más desperdicio del buffer de datos.

### ***3 Data Warehousing***

#### ***3.1 Definición de Data Warehouse***

Un Data Warehouse (Almacén de Datos o Bodega de Datos) es usualmente definido como una colección de datos integral, variable en el tiempo, no volátil y orientada a temas importantes para el soporte de la toma de decisiones en la administración de la organización.<sup>1</sup> Es necesario resaltar las cuatro características básicas para que un repositorio de información sea reconocido como un Data Warehouse.

- El Data Warehouse debe ser integral. No importando de donde provengan las fuentes de datos que lo integran deberán formalizarse como un solo tipo descrito de datos.
- El Data Warehouse debe estar organizado en relación al objetivo principal de la aplicación y las variables más importantes de la organización. Esta información es el soporte real para la toma de decisiones por lo que su objetivo particular debe ser servir a la misión central de la organización.
- El Data Warehouse debe ser incremental en el tiempo. Debe almacenar información de varios años.
- El Data Warehouse no debe ser volátil. No debe permitir agregaciones, modificaciones, o eliminaciones de los datos que contiene cuando estas provengan del usuario.

---

<sup>1</sup> Definición y características del Dr. William Inmon Creador del concepto y pionero en el desarrollo del tema. Silverston, Inmon y Graziano, The Data Model Resource Book, 235-268 pp.

Como se mencionó en el capítulo anterior la necesidad de utilizar modelados de datos multidimensionales en repositorios de información para resolver favorablemente a las interrogantes organizacionales corresponde en gran medida al concepto que acabamos de describir. Por lo tanto, el Data warehouse es por sí mismo, como su concepto lo describe, una colección de datos como lo podría ser la misma base de datos operacional sumando las características requeridas.

Es decir la base o bases que soportan los sistemas de transacción comunes de la organización, como ventas, inventarios, compras, etc. también podrían proveer los servicios para el soporte a la toma de decisiones; pero también es necesario recordar que los análisis para la toma de decisiones requieren de usos exhaustivos de los sistemas, por la necesidad de consultas muy específicas que seguramente serán cambiadas y requeridas desde otra perspectiva, lo que aniquilaría el rendimiento de la operación normal de las transacciones.

La operación de los sistemas es necesaria y tan importante como los impactos que una mala decisión pudiese ocasionar en la organización, surge entonces la idea de proveer de datos a los DSS sin afectar la operación de los sistemas de misión crítica. (descritos como OLAP en el primer capítulo) La verdadera explotación de esta colección de datos consiste en los usos y provechos que se obtengan de ellos mediante todo un sistema completo que obtenga los datos y los ponga en el escritorio del tomador de decisiones.<sup>2</sup>

---

<sup>2</sup> Barquin, Ranton C., A Data Warehousing Manifesto, Planning and Designing the Data Warehouse, p. 3-8.

### 3.2 Definición de Data Warehousing

Data Warehousing es el proceso de extracción y transformación de datos obtenidos en fuentes operacionales (OLTP) llevándolos a una base centralizada reconocida como un Data Warehouse. Por lo que una vez en el repositorio es explotada usando herramientas para la toma de decisiones. Data Warehousing responde a un esquema Cliente / Servidor donde por un lado es cliente de bases de datos operacionales o transaccionales y por otro es servidor de clientes finales con necesidades específicas de consulta. (Ver Figura 3.1)

#### Modelo General del Data Warehousing

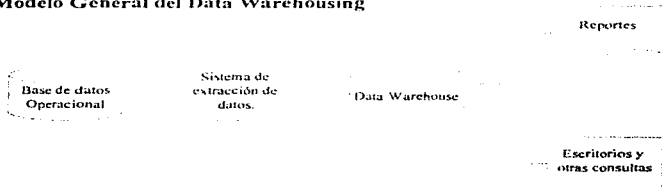


Figura 3.1

El modelo tiene cuatro objetivos principales:

- Proteger los sistemas de producción del acceso indiscriminado de usuarios hambrientos de información, que podrían afectar el rendimiento de estos sistemas de misión crítica
- Proveer un ambiente de información protegido y bien administrado para la toma de decisiones, ya que representa un activo importante para la organización.
- Construir un modelo de datos organizacional que permita una estandarización en el manejo de información tanto de los sistemas de producción como de los sistemas para el soporte a la toma de decisiones.

- **Mantener independencia entre las aplicaciones de acceso del usuario y la administración de la información, manejándolas como problemáticas independientes.**

En general es necesario considerar los siguientes elementos para establecer con claridad que un DSS es Data Warehousing:<sup>1</sup>

- **Fuentes originales de información y datos, tales como bases de datos que soportan el sistema o sistemas de misión crítica de la organización.**
- **Herramientas de migración y concentración de datos que logren el acceso y conectividad a las fuentes de datos y las transformen.**
- **Un repositorio de información (Data Warehouse)**
- **Elementos de recuperación, formateo y análisis del contenido del repositorio.**
- **Herramientas de administración y monitoreo de todo el proceso y del repositorio.**

Los elementos anteriores necesariamente deben ser considerados para la elección de una arquitectura, es necesario estructurar los niveles en que el sistema integrado operará, por lo que una definición de niveles es una solución de arquitectura para el problema del control de costos en la administración del Data Warehousing, en particular en un ambiente de datos inestable o cambiante. Relacionando ambos conceptos, el de OLAP y el de Data Warehousing se requiere de una sólida arquitectura para el desarrollo de sistemas que respondan al proceso de toma de decisiones correctamente en un tiempo muy reducido.

---

<sup>1</sup> Edelstein, Herb, "An Introduction to Data Warehousing", Planning and Designing the Data Warehouse, p. 36-37.

- **Mantener independencia entre las aplicaciones de acceso del usuario y la administración de la información, manejándolas como problemáticas independientes.**

En general es necesario considerar los siguientes elementos para establecer con claridad que un DSS es Data Warehousing:<sup>3</sup>

- Fuentes originales de información y datos, tales como bases de datos que soportan el sistema o sistemas de misión crítica de la organización.
- Herramientas de migración y concentración de datos que logren el acceso y conectividad a las fuentes de datos y las transformen.
- Un repositorio de información (Data Warehouse)
- Elementos de recuperación, formateo y análisis del contenido del repositorio.
- Herramientas de administración y monitoreo de todo el proceso y del repositorio.

Los elementos anteriores necesariamente deben ser considerados para la elección de una arquitectura, es necesario estructurar los niveles en que el sistema integrado operará, por lo que una definición de niveles es una solución de arquitectura para el problema del control de costos en la administración del Data Warehousing, en particular en un ambiente de datos inestable o cambiante. Relacionando ambos conceptos, el de OLAP y el de Data Warehousing se requiere de una sólida arquitectura para el desarrollo de sistemas que respondan al proceso de toma de decisiones correctamente en un tiempo muy reducido.

---

<sup>3</sup> Edelstein, Herb, "An Introduction to Data Warehousing", Planning and Designing the Data Warehouse. p. 36-37.

### 3.3 Arquitectura del Data Warehousing

Al igual que otros métodos para el desarrollo de sistemas, descritos bajo un modelo Cliente / Servidor la segmentación de un Data Warehousing en varios niveles permite diferenciar varios elementos de la arquitectura en particular aquellos encargados de hacer ciertas tareas específicas. (Ver figura 3.2)

Proceso del Data Warehousing en dos Niveles entre el Data Warehouse y el cliente.

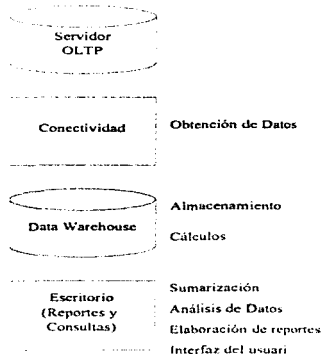


Figura 3.2

Considerando lo anterior la pregunta esencial que debe contestarse con respecto a la arquitectura de un warehouse es: ¿cuántos niveles deben emplearse y en qué momento dentro del ciclo de vida del proyecto?.



Existen factores básicos que deben tomarse en cuenta al considerar la aplicación de niveles dentro del Data Warehousing:

- La volatilidad semántica del dominio de los datos del warehouse. ¿Con qué frecuencia los datos requieren ser adquiridos de las bases de datos en operación?
- La independencia de organización de la comunidad de usuarios. ¿Cuál es la diferencia entre los usuarios más notable en cuanto a la cantidad y concentración de la información así como su posición en la organización?
- El riesgo del ciclo de vida del proyecto. ¿Cuál es el presupuesto asignado y calendarización de los resultados?
- El costo de implantación de una arquitectura más compleja.

Esto ha originado un segundo modelo de niveles basado en tres niveles que establecen una arquitectura más compleja en tamaño pero respondiendo al desempeño crítico con más celeridad. (Ver figura 3.3)

Proceso del Data Warehousing  
en tres Niveles entre el  
Data Warehouse y el cliente.

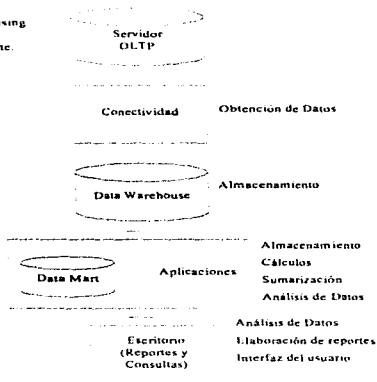


Figura 3.3

En la actualidad podemos identificar la evolución en el mercado de cuatro estrategias generales para la implantación de este tipo de sistemas consolidando los dos conceptos para soportar la toma de decisiones, respondiendo de distinta manera a las tres preguntas planteadas con anterioridad y estableciendo diferentes niveles en la solución.

Las diferentes arquitecturas también están asociadas necesariamente al desarrollo de un tipo de proyecto y una metodología distintas entre sí aunque lo bastante similares para generar un producto final que se conceptualiza de la misma forma.<sup>4</sup>

<sup>4</sup> Jenks, Bruce; Data Warehouse de Niveles, (Tiered Data Warehouse), Soluciones Avanzadas Año 4, No 34, 1996.; p. 58

### ***3.3.1. Data Warehousing ligero.***

Esta aproximación plantea la instalación de software de acceso (gateway) que conecta directamente y en línea la aplicación del cliente con la información de sistemas específicos de operación. Este enfoque, aunque limitado, permite aplicaciones particulares y de bajo volumen tanto de información como de accesos soportar sistemas para la toma de decisiones. La forma de implantar este tipo de arquitectura es incluir el uso de nuevas tecnologías basadas en ambientes cliente-servidor, en lugar de usar las herramientas tradicionales fundamentadas en el mainframe o sistemas propietarios.

Algunos autores han identificado esta estrategia de implantación como "Warehousing Lite", y básicamente consiste en un sistema de reportes para una aplicación en particular; las principales ventajas de esta aproximación residen en incorporar el potencial de las herramientas gráficas fundamentadas en computadoras personales para explotar la información residente en los ambientes propietarios, logrando con esto una rápida implantación sin la necesidad de incorporar demasiados elementos nuevos en la infraestructura, siendo la inversión relativamente moderada. Por otro lado, este tipo de sistemas está limitado en su crecimiento, dado que accede directamente a los sistemas en producción, además de poder llegar a presentar problemas de integración en un ambiente corporativo y respeta las estructuras definidas en los sistemas OLTP.<sup>5</sup>

### ***3.3.2. Data Warehousing virtual.***

Esta estrategia se ha adoptado como la evolución del uso de Gateways (Puentes) a bases de datos, incorporando una visión más corporativa, donde el uso de diversas herramientas de middleware (intermedio en el esquema Cliente / Servidor) permiten el

---

<sup>5</sup> Miramón Commons Bernardo; Data Warehousing Estrategias Generales de implantación; Soluciones Avanzadas Año 4, No 34, 1996; p. 48.

acceso directo y en línea de las aplicaciones de los clientes DSS a diversos sistemas de operación, que se pueden estar ejecutando en múltiples bases de datos o estructuras de archivos. El middleware brinda la capacidad al usuario de ver la información de ambientes de cómputo distribuidos como una unidad estándar, sin preocuparse de su localización física o el formato en el que se encuentra.

Existen herramientas que permiten una vista relacional y uniforme de la información de sistemas de operación, la implantación se fundamenta en la instalación de un repositorio centralizado de metadatos (Datos modelados obtenidos de la información de las bases de datos en operación), donde se manejan las reglas del negocio y los apuntadores a donde se encuentra localizada físicamente la información, ya que no existe un Data Warehouse físico.

Esta aproximación facilita el acceso de datos de manera corporativa, permitiendo la incorporación de diversas aplicaciones sin la necesidad de invertir en infraestructura de hardware para soportar su operación; sin embargo, este enfoque es limitado, ya que cuando se manejan grandes volúmenes de información, con toda certeza se puede llegar a afectar el desempeño de los sistemas OLTP.\*

### ***3.3.3. Data Warehousing corporativo.***

Como visión original, contempla la creación de un almacén de datos independiente a los sistemas de operación, dedicado exclusivamente a satisfacer las necesidades de los sistemas para el soporte a la toma de decisiones. La arquitectura general se fundamenta en la creación de un modelo de datos corporativo, lo que implica un estandarización completa en la administración de la información a través de la implantación de un repositorio físico

---

\* Ibid.; p. 48

de metadatos global, sin embargo, también implica un árduo trabajo de definición antes de implantar el modelo en producción.

Este enfoque es sin lugar a dudas el más completo, ya que cumple con los objetivos planteados en la visión original, pero requiere de fuertes inversiones al inicio del proyecto y largos tiempos de desarrollo, por lo cual ante los ojos de diversas organizaciones, el desarrollo de un Data Warehouse corporativo no cumple con las necesidades dinámicas de las empresas actuales.<sup>7</sup>

### *3.3.4. Data Warehousing con mercados de datos.*

Esta aproximación se creó como un complemento a la implantación de almacenes centrales de datos, y su objetivo original es el crear pequeños almacenes satélites alrededor del Data Warehouse corporativo, para atender necesidades específicas a una área geográfica o departamental sin tener que conectarse al almacén central, esto implica una organización de datos distribuida que debe mantenerse en sincronía absoluta. La estrategia del data marketing (mercado de datos), no solamente ha sido usada como un complemento al repositorio central, sino como una estrategia sustituto; esto quiere decir que en lugar de crear un gran almacén de datos corporativo, se implantan pequeños almacenes que satisfacen necesidades particulares, y que extraen la información directamente de los sistemas de operación.

Este enfoque permite crecer modularmente sin la necesidad de realizar grandes inversiones al inicio del proyecto, obteniendo resultados en tiempos relativamente cortos. Por otro lado, en este tipo de arquitectura es posible enfrentarse al problema de no contar

---

<sup>7</sup> Ibid.; p. 48.

con un modelo de datos corporativo y centralizado, lo que podría acarrear problemas de integración e incrementar las actividades de administración de la información.<sup>3</sup>

### ***3.4 Aplicación de las arquitecturas.***

La evolución tradicional de los proyectos de data Warehousing se dio como el resultado natural del crecimiento tanto de necesidades, como de usuarios. Esto provocó que los proyectos para soporte a la toma de decisiones aplicaran una estrategia diferente, dependiendo de la etapa del ciclo de vida por la que atravesaban.

El decidir usar alguna de estas estrategias dependerá, en primera instancia, directamente de las necesidades de la organización y, posteriormente, de un análisis de nuestra infraestructura de cómputo y comunicaciones en relación con el volumen de información y la cantidad de usuarios DSS en la organización.<sup>4</sup>

El empleo de una forma de trabajo ordenada es un factor de importancia en el desarrollo e implantación de proyectos de Data Warehousing, y la tendencia general busca lograr, a través del uso de una metodología, recortar los tiempos de desarrollo y programar la inversión de recursos de manera eficiente. El desarrollo de este tipo de proyectos se ha asociado constantemente con una inversión excesiva y rendimientos a largo plazo por lo que el soporte de la justificación del proyecto no siempre es bien vista.

---

<sup>3</sup> Ibid. p. 48.

<sup>4</sup> Jenks, Bruce: Op. Cit. p. 59.

## ***4 Proyecto para la Toma de Decisiones utilizando Data Warehousing***

### **4.1 Selección de la estrategia.**

Uno de los pasos más importantes a seguir consiste en decidir la estrategia general de implementación. La decisión tiene mucho que ver con la cultura de la organización y se basa en cómo se llevan a cabo los procesos dentro de esta. Las siguientes son estrategias que han demostrado su eficacia desde las perspectivas que abordan la problemática.

#### ***4.1.1 Enfoque de lo general a lo particular.***

Primero se identifican los requerimientos organizacionales o del tomador de decisiones en referencia a que debe cubrir el Data Warehousing propuesto, este enfoque para planearlo e implementarlo se recomienda bajo los siguientes escenarios:

- Cuando la organización está familiarizada con la tecnología y posee una amplia experiencia en el desarrollo de aplicaciones con base en la identificación de requerimientos organizacionales de lo general a lo particular.
- Cuando se prevén un conjunto claro de objetivos para el data warehouse, ello implica que se entienden perfectamente tanto el proceso como sus beneficios.
- Cuando se tiene una idea clara de dónde encaja el data warehouse en la estructura y en la organización como herramienta de apoyo a las decisiones.
- Cuando se tiene una idea clara de cómo la utilización del data warehouse es un subproceso de un proceso organizacional ya establecido.

Un enfoque de lo general a lo particular es útil cuando la tecnología es ya madura y se comprende bien, y cuando están claros y bien entendidos los problemas a resolverse. Aplicar este enfoque proporciona un excelente ajuste entre la tecnología y los objetivos organizacionales de tal manera que cuando se hace adecuadamente, produce el máximo efecto sobre la inversión.<sup>1</sup>

#### ***4.1.2 Enfoque de lo particular a lo general.***

Generalmente, este enfoque comienza con experimentos y prototipos basados en tecnología. Se selecciona un subconjunto específico bien entendido, de la problemática organizacional y se formula una solución para este subconjunto. Por lo regular, la implementación de esta solución es rápida, ya que comprende un número reducido de personas tomando menos decisiones para resolver un problema minimizado.

El enfoque de lo particular a lo general es útil en las primeras etapas de madurez de la tecnología. Este enfoque permite que una organización avance con un gasto considerablemente menor y que evalúe los beneficios de la tecnología antes de establecer compromisos significativos. En el área del Data Warehousing, por lo regular se toma un enfoque de estas características para implementar un mercado de datos, un sistema de información ejecutiva, o un data warehouse departamental que está claramente orientado a responder consultas bien escogidas en un dominio determinado y sobre todo muy específico.

El enfoque de lo particular a lo general para la planeación e implementación de un Data Warehousing se recomienda en los casos siguientes:

- Existe una oposición grande y conservadora ante la nueva tecnología sin una consideración suficiente o bien existen pocos argumentos para establecer la necesidad.

---

<sup>1</sup> Harjinder S. Gill y Prakash C. Rao, Data Warehousing, p. 79-80



- Cuando la organización todavía no está comprometida con la tecnología del data warehouse, pero busca una evaluación tecnológica para determinar cómo, dónde y cuándo desplegar esta tecnología.
- Cuando la organización trata de obtener una idea de los costos y gastos adicionales de implementar y desplegar la tecnología del Data Warehousing.
- Cuando no están tan claros los objetivos organizacionales que deben satisfacer la implementación propuesta del data warehouse, ni tampoco este claro cuál de los procesos actuales o propuestos, serán afectados, por el uso del Data Warehousing.
- Cuando la organización este realizando una inversión oportunista, en vez de estratégica, en la tecnología Data Warehousing.

Un enfoque de lo particular a lo general es útil para realizar evaluaciones de tecnología y es una buena técnica para organizaciones que no la implementan por no ser del todo satisfactoria. Este enfoque también sirve para que una empresa aproveche las tecnologías que se encuentran en su ciclo temprano de madurez sin comprometerse a riesgos mayores.<sup>2</sup>

#### ***4.1.3 Enfoque combinado.***

En el enfoque combinado una organización puede explotar la naturaleza planeada y estratégica del “enfoque de lo general a lo particular”, al tiempo que conserva la rápida implementación y la aplicación oportunista del enfoque de lo particular a lo general. Este enfoque depende de dos componentes:

- Una arquitectura de lo general a lo particular, basada en estándares y un equipo de diseño que aplique la experiencia adquirida de proyecto en proyecto y que pueda retroceder y convertir las decisiones tácticas en decisiones estratégicas.

---

<sup>2</sup> Ibid. p. 81

- Un equipo de trabajo de lo particular a lo general que se concentre en implementar una solución organizacional muy enfocada, estrecha, pero de largo alcance, en un período de tiempo reducido.

El enfoque combinado tiene las ventajas de los dos anteriores, sin embargo, es más complejo de manejar como proyecto. Este enfoque es recomendable bajo las siguientes circunstancias:

- Cuando la organización que realiza la implementación existen arquitectos experimentados. La organización ha tenido una base para la construcción, documentación, aplicación y mantenimiento de arquitecturas de datos, arquitecturas de tecnología y los modelos de la empresa pueden pasar con facilidad de lo concreto (como son los datos dentro de sistemas operacionales) a lo abstracto (como son modelos lógicos o abstracciones basadas más bien en la naturaleza del negocio que en la tecnología empleada para implementar sistemas).
- Cuando la organización tiene un equipo de trabajo comprometido que tiene un enfoque claro de en dónde desea aplicar la tecnología del data warehouse. Con frecuencia, este equipo es un departamento compuesto por usuarios finales y personal de Sistemas dedicado al manejo de la información que tiene una visión clara para resolver una necesidad organizacional concreta y actual.

El enfoque combinado se adapta mejor al despliegue rápido y oportunista de la tecnología del data warehouse, al tiempo que se reserva el derecho de construir una solución estratégica que tenga valor a largo plazo.<sup>3</sup>

---

<sup>3</sup> Ibid. p 82.

#### 4.2 Desarrollo de los objetivos organizacionales.

El primer paso para efectuar cualquier tarea importante y compleja en desarrollo de sistemas, consiste en desarrollar una lista de objetivos organizacionales que el sistema debe satisfacer. Un segundo elemento a considerar antes de iniciar el desarrollo, será el analizar estos objetivos organizacionales y generar los requerimientos para que estos sean correctamente satisfechos con el sistema. Es imposible comenzar a construir un data warehouse simplemente con base en la comprensión de las diversas fuentes de datos que deben alimentarlo, sin estructurar algún análisis inicial con los usuarios finales del data warehouse.

El establecimiento de objetivos del Data Warehousing es una tarea compleja. Esto se debe a que el data warehouse, de manera muy parecida a las bases de datos que lo alimentan con información, tiene una capacidad a cualquier plazo, debiendo proporcionar un amplio rango de información para un extenso conjunto de analistas y usuarios finales. Incluso en general resulta difícil comenzar a documentar las necesidades de dichos usuarios. Otra fuente de dificultades surge de la diferencia en perspectivas, terminología y definiciones de información desde el punto de vista del personal especializado en tecnología de la información que maneja los recursos de datos y los usuarios finales o bien de analistas que desean derivar un apoyo para las decisiones a partir del data warehouse.<sup>4</sup>

Considerar el Data Warehousing como un "producto" interno para una organización es una simplificación, es un sistema de software cuyos objetivos son apoyar decisiones para una audiencia puramente interna. Con esta visión, es posible que el grupo de desarrolladores se concentren en establecer los objetivos del Data Warehousing, lo que por lo regular se hace atendiendo los siguientes elementos:

---

<sup>4</sup> Ibid. p. 91

- **El mercado de destino ( la audiencia potencial) del data warehouse**

Se basa en el grado de necesidad y valor percibidos. Las personas con una mayor necesidad percibida requerirán de un más alto valor percibido de los servicios del data warehouse. La selección de la audiencia de destino se complica por las necesidades de estaciones de trabajo, conectividad en red y un grado de familiarización de la audiencia con la tecnología de cómputo.

- **Las plataformas en uso actual o planeadas**

Se refiere a un amplio rango de dimensiones de plataformas, tales como servidores, conectividad y redes, estaciones de trabajo, clientes GUI (Graphical User Interfaces, Interfaces Gráficas de Usuario), y dispositivos de base de datos (en particular, sistemas de administración de bases de datos relacionales o multidimensionales y comunicaciones).

- **Las capacidades planeadas en términos de características y funciones**

Se relaciona con el valor que asignan los usuarios finales a cada característica. La inversión requerida para desarrollar capacidades de valor dudoso representa un desperdicio de recursos. Además, para que la solución de data warehouse sea útil se debe considerar si satisface un conjunto mínimo de características. Las características y funciones caen dentro de las dos siguientes categorías:

- Características y funciones visibles que pueden ser utilizadas por usuarios finales externos del data warehouse.
  - Características y funciones que no son visibles de manera externa pero que deben implementarse dentro del data warehouse a fin de ofrecer sus capacidades.
- **Las diversas fuentes de datos que pueden y/o deben integrarse al Data Warehousing**

Se tienen implicaciones tecnológicas basadas en la dificultad de extracción de información de las bases de datos en operación. Además de considerar las fuentes

externas de datos que proporcionan información de mercados, clientes y de la competencia.

- **Inicio de operaciones del Data Warehousing**

Se basa en muchos factores. Los competidores ya están usando el data warehouse para identificar sus clientes y ofrecer servicios mejorados. Un departamento dentro de la organización ya no puede llevar a cabo sus funciones debido a la falta de disponibilidad de información importante sobre clientes y datos históricos, etc.

### ***4.3 Estudio de los requerimientos.***

En la mayoría de las organizaciones y como sucede con muchos otros sistemas, la motivación principal del proyecto del Data Warehousing es conseguir una primera implementación que produzca beneficios inmediatos a un grupo de usuarios. Después de definir un rumbo general de objetivos para el data warehouse, se hace necesario derivar con rapidez un ámbito limitado para la primera implementación o el desarrollo de un prototipo evaluable y replicable.

El volumen de acopio de requerimientos que se debe realizar depende del enfoque de desarrollo que se tome. Si se toma el enfoque de lo general a lo particular, la actividad de acopio de requerimientos es significativamente grande, pero debido a que las actividades están conducidas por el negocio están bien entendidas y pueden, por lo tanto, catalogarse con facilidad. Si se toma el enfoque de lo particular a lo general, la mayoría de los requerimientos se construyen de manera oportunista con base en esperanzas y expectativas. Aquí el conductor principal es la implementación rápida de la cual se pueda aprender. El enfoque combinado intenta hacer una mezcla de ambas que necesariamente implica un grupo de trabajo muy comprometido.

En este sentido y considerando que las estrategias del desarrollo de un proyecto ofrecen tres perspectivas; es necesario abordar el tema del desarrollo del proyecto aquella que resulte ser la más sensata en relación a las posibilidades reales de culminación y éxito del proyecto. Por lo que es necesario determinar una serie de pasos para obtener esta respuesta razonada y así elegir el desarrollo de un prototipo o el enfoque de lo particular a lo general como la estrategia más adecuada para la gran mayoría de los proyectos.<sup>5</sup>

Primero es necesario saber quiénes usarán el data warehouse y para qué fines, la selección de un número limitado de departamentos puede reducir de manera significativa los requerimientos de datos y metadatos del data warehouse. También es necesario tener un primer estudio sobre las consultas que determinan en forma directa las adiciones, resúmenes integración y reingeniería que debe realizarse sobre los datos provenientes de las fuentes de datos. El ámbito de consultas organizacionales debe además estar restringido por el número de dimensiones de datos, la variedad de los reportes y la cantidad de transferencias de escritorio que colocan demandas sobre la implementación del Data Warehousing.

Entre más específicos sean los formatos de consulta será más fácil definir las dimensiones y las especificaciones de adición y resumen, así como el rango de entidades empresariales que deben rastrearse dentro del data warehouse.

Por otro lado, para determinar el ámbito desde la perspectiva de la tecnología, deben analizarse y conciliarse otros elementos entre los que destaca el metamodelo del data warehouse (El metamodelo es el modelo que contiene la definición de los datos del data warehouse). La definición se guarda dentro del catálogo del administrador del data warehouse y lo utilizan todas las herramientas de consulta y reporte. El tamaño del metamodelo es un indicador directo del tamaño de los datos que deben manejarse dentro

---

<sup>5</sup> Carol L. Burleson y David E. Tabler; "Data Warehousing - Putting It All Together", Planning and Designing the Data Warehouse; p 118

del data warehouse. Por lo tanto, la restricción del metamodelo inicial restringe la cantidad de datos que debe manejar la primera implementación de un data warehouse.<sup>6</sup>

Otro elemento técnico importante para el establecimiento de un prototipo es que el data warehouse no solo contiene datos actuales, sino también datos históricos reunidos durante varios años. El grado de resumen determina qué parte de estos datos se encuentra comprimido y resumido. Si el data warehouse va a proporcionar la capacidad para que las consultas organizacionales “profundicen” en el registro histórico real, se requiere que el data warehouse soporte una administración de datos para volúmenes enormes de información. El tamaño de los datos tiene una relación directa con el tiempo de ejecución de las consultas. La realización de elecciones sensatas con respecto a el proporcionar la profundización necesaria y suficiente sólo cuando se requiere y el permitir manejar registros resumidos en vez de detallados, son técnicas para restringir el ámbito de la implementación en prototipo.

El tercer elemento importante para establecer los criterios de utilización del Data Warehousing es como se alimenta con la información proveniente de bases de datos de producción o de misión crítica que manejan las operaciones diarias. Después de que los datos de estas fuentes se cargan en el data warehouse, se deben mantener sincronizados con las fuentes originales. Esta es una carga significativa para el data warehouse, en especial cuando las fuentes no son susceptibles y congruentes con las nuevas características de bases de datos relacionales, como los servicios de duplicación que facilitan las tareas de sincronización. Por lo tanto, la selección inicial de fuentes de datos, su tipo de base de datos y el número de fuentes tiene una fuerte influencia en la capacidad de llevar a cabo una primera implementación.

En este sentido, las bases de datos de herencia presentan retos significativos en la forma en que organizan los datos. se convierte en un problema la calidad de datos que

---

<sup>6</sup> Harjinder S. Gill y Prakash C. Rao, Op Cit p. 87

contienen, a menudo por la ausencia de filtrado y depuración de datos durante la captura. La elección de tales fuentes de datos puede provocar demoras significativas en la implementación. Mientras se atienden y resuelvan estos retos, es de mayor importancia contar con documentación que describa la organización de los datos en las bases de herencia. La calidad de esta documentación determina en gran medida el éxito para extraer datos de estas fuentes. Relacionada con esta cuestión está la disponibilidad de las personas que estén familiarizadas con estas fuentes de datos de herencia y que puedan llenar los huecos en la documentación disponible.<sup>7</sup>

La implementación de un data warehouse es más rápida en un ambiente de diccionarios de datos, depósitos y catálogos de bases de datos que estén actualizados y contengan definiciones precisas de los datos. El volumen de extracción y carga a partir de las fuentes de datos depende de la comprensión y estructura de los datos fuente.

El data warehouse debe ser una extensión de la arquitectura de datos de una organización. La disponibilidad de administradores de bases de datos, administradores de datos y programadores que ya comprendan la arquitectura de datos y hayan trabajado con ella diariamente, acelera la implementación de un data warehouse. Esta implementación sobre las plataformas existentes u otras similares también reduce el reto tecnológico que plantea el desarrollo sobre nuevas plataformas.

En general es necesario considerar una serie de preguntas que el análisis de requerimientos debe contestar claramente, estas son:

- ¿Por qué construir un data warehouse o un mercado de datos?
- En caso de segmentarlo ¿Qué problema organizacional abordará?
- ¿Cuáles son los objetivos?
- ¿Cuánto costará?

---

<sup>7</sup> Ibid. p.89



- ¿Cuándo estará listo?
- ¿Cuál es el impacto sobre la gente? ¿Sobre las habilidades? ¿Sobre la organización?
- ¿Cómo afecta nuestras inversiones actuales en cómputo?
- ¿Tenemos la capacidad para hacerlo?
- ¿Cuáles son los riesgos?

#### ***4.4 Equipo de trabajo***

Es importante establecer claramente el propósito a lograr al formar el equipo de trabajo. Este equipo será el encargado de revisar cuáles son las condiciones actuales de la organización, deberá definir y justificar todo el desarrollo así como hacer una planeación de la inversión que permita un flujo de efectivo desahogado para la operación normal de la organización.

Este equipo de trabajo debe inicialmente formarse por pocos pero muy entusiastas implicados que conozcan profundamente la organización y tengan los elementos dentro de ella para desarrollar los objetivos organizacionales descritos anteriormente. Específicamente es deseable que los miembros del equipo estén caracterizados por los siguientes elementos:

- Disponibilidad al reto
- Deseo de innovación
- Jerarquía alta o media
- Antigüedad en la organización

Es necesario que el equipo inicial incluya a dos elementos básicos, por un lado los encargados del área de informática o responsables del proyecto, encargados de la definición de requerimientos, y por otro lado los usuarios tomadores de decisiones que establecerán

junto con el personal de sistemas del proyecto los requerimientos de sus consultas y sus necesidades de información.

Este mismo grupo de trabajo debe abocarse a plantear el programa de actividades en relación al objetivo organizacional a cubrirse, la definición de las metas específicas deben estas asociadas a una revisión objetiva por parte de los usuarios del Data Warehousing así como el presupuesto de gastos en que el proyecto incurra lo largo de su desarrollo. <sup>8</sup>

La implementación del data warehouse es un esfuerzo de equipo. Después de identificar a los diferentes miembros del equipo, es importante definir sus papeles dentro del proyecto. Esto se hace mediante la asignación de diversas funciones de miembros del equipo dentro de los bloques y capas del proyecto. Establecer los requerimientos como se mencionó antes es responsabilidad tanto de usuarios tomadores de decisiones como personal del proyecto mismo.

#### ***4.5 Justificación de proyecto***

Resulta bastante complicado justificar proyectos de sistemas para múltiples organizaciones, es bien sabido que los desarrollos tienen muy poco impacto visual en los tomadores de decisiones, el caso del Data Warehousing es más extremo aún, si se justifica la inversión en base al costo beneficio de forma general; es muy complicado establecer un beneficio tangible sobre la correcta toma de decisiones, sin embargo es importante justificar la inversión utilizando los siguientes argumentos:<sup>9</sup>

- Mejores elementos para la planeación

---

<sup>8</sup> Carol L. Bureson y David E. Tabler, Op. Cit. : p. 128-129

<sup>9</sup> Neal C. Doug. "How to Justify the Data Warehouse and Gain Top Management Support", *Planning and Designing The Data Warehouse*, p. 91-92

- Costo de oportunidad
- Beneficios sobre competidores
- Ventajas por áreas problemáticas
- Mejora continua
- Mayor penetración en el mercado
- Mejora de procesos

La justificación debe hacerse contener en un documento explicativo primero del concepto mismo, ya que Data Warehousing no es fácil de explicara a tomadores de decisiones si no es en los términos que ellos reconocen para su trabajo, además de los beneficios que podrían obtener de él. Es necesario justificar el Data Warehousing mucho más que cualquier otro proyecto en términos de beneficios poco tangibles tales como documentos estratégicos en general específicamente:<sup>19</sup>

- Planes de mercadotecnia
- Planes estratégicos
- Presupuestos
- Reportes comparativos y tendencias

Muchos de los proyectos de Data Warehousing han fallado por que los beneficios nunca alcanzan la inversión, esto probablemente es debido a que el desarrollo del proyecto se sustenta en soluciones que requieren demasiada inversión tanto en hardware como en software, en desarrollos complejos y con resultados a mediano plazo. La metodología detallada en el siguiente capítulo, se sustenta en la necesidad de hacer sistemas para el soporte de toma de decisiones utilizando Data Warehousing sin tener que justificar inversiones y desarrollos incoesteables.

---

<sup>19</sup> Ibid. p. 92-94.

## *5 Metodología para el desarrollo de Data Warehousing*

### *Nota aclaratoria sobre la metodología*

Las metodologías ayudan a finalizar un proyecto en tiempo y dentro de un presupuesto. Además, ayudan a evitar los olvidos y fallas comunes; proporcionan un lenguaje común, logrando que exista comunicación; permiten la incorporación de nuevos miembros al equipo, siendo productivos inmediatamente.

La presente metodología representa una propuesta para desarrollar un proyecto de Data Warehousing utilizando una estrategia de lo particular a lo general descrita en el capítulo anterior, incorporando el análisis multidimensional en arquitecturas basadas en OLAP multidimensional tal como esta descrito en el capítulo segundo. Sobre la arquitectura general del Data Warehousing se propone la utilización de mercado de datos descrita en el capítulo tercero.

Esta metodología se propone en proyectos para la toma de decisiones en una área o departamento, sin tener que modelar a toda la organización, sin invertir sumas considerables en Hardware y Software así como para esperar obtener resultados en el mediano plazo. Específicamente, esta metodología se adecua a organizaciones que desean probar la utilización de Data Warehousing sin arriesgar sumas considerables o que su dimensión y problemática es tan amplia que la elaboración de un modelo general de la organización resultaría sumamente complejo.

## ***5.1 Análisis de Requerimientos***

La fase de análisis de requerimientos del Data Warehousing es la elaboración de una especificación precisa de las funciones que se obtendrán del proceso completo. Además de las características y funciones necesarias, los requerimientos describirán con claridad el ambiente operativo en el que funcionará el data warehouse.

### ***5.1.1 Requerimientos del usuario final.***

El usuario final ve el proceso de Data Warehousing como una gran caja negra cuyo acceso principal es a través de aplicaciones y herramientas de consulta y reportes, junto con cierto tipo de ubicación de la información contenida dentro del data warehouse. Los requerimientos del usuario final pueden ubicarse en una o más de las siguientes categorías:

- **Flujo de trabajo**

Los flujos de trabajo son una tendencia del desarrollo de sistemas para las organizaciones que afecta directamente el proceso automatizado y que requiere de grandes cantidades de información para poder tomar decisiones sobre el mismo flujo de trabajo, es importante analizar que ofrece el data warehouse con el flujo de trabajo diario del usuario final.<sup>1</sup>

- **Requerimientos de consulta**

En la fase de planeación, se mencionó el empleo de escenarios de uso organizacional como una valiosa herramienta para hacer prototipos de la capacidades del Data Warehousing y para establecer las expectativas del usuario final. Los requerimientos de consulta simplemente captan las consultas expresadas en la terminología del usuario final, por lo que se espera:

---

<sup>1</sup> Silverston, Immon y Graziano, The Data Model Resource Book, 254-268 pp.

- Tener un acceso rápido, buena manipulación y excelente presentación de información.
- Satisfacer las necesidades de una variedad de usuarios.
- Permitir a los usuarios la creación de sus propias consultas utilizando términos que ya conocen y ofrecer estructuras de datos consistentes.
- Profundizar sin repetir el acceso.
- Requerir de una capacitación y soporte mínimos.

Cada uno de los usuarios finales antes descritos tiene diversos requerimientos de reportes. Los usuarios finales pueden además especificar los tipos de análisis de datos que desean realizar sobre los datos recuperados del data warehouse estos también son parte del proceso de Data Warehousing y se reconocen como los siguientes:

- **Tipos de actividades**
  - Rebanar y picar - Separar elementos de datos en varias formas.
  - Penetración - Exponer más detalle de manera progresiva.
  - Navegación - Examinar de una manera indirecta.
  - Copiar y realizar modificaciones locales.
  - Construir modelos empresariales, por ejemplo mediante hojas de cálculo.
- **Visualización de datos:**
  - En dos dimensiones - Hojas de cálculo y relacional.
  - En múltiples dimensiones.
  - Reportes y cuadros.
  - Base de datos de muestra activa.

### **5.1.2 Requerimientos de la arquitectura.**

Se deben compilar una serie de requerimientos que coincidan con la visión del usuario, así como un conjunto de requerimientos que refleje la implementación de la tecnología. La calidad del esfuerzo de arquitectura determinará lo siguiente:

- El rango de funciones y características a ofrecer.
- El rango de plataformas necesarias para la implementación.
- El uso de estándares e interfaces abiertas.
- La flexibilidad para incorporar mejoras.
  
- **Requerimientos de la arquitectura de datos** - La cual describe los elementos de datos y sus relaciones. La caracterización de datos se considera como una actividad fundamental. La razón es que las aplicaciones no pueden desarrollarse sin definir los datos que deben crear o modificar.
- **Requerimientos de la arquitectura de aplicación** - Un Data Warehousing es un sistema que representa una combinación de diversas aplicaciones. Juntas, estas aplicaciones producen la funcionalidad de un sistema completo. La arquitectura de aplicación se define después de caracterizar la arquitectura de datos. La arquitectura de aplicación es un catálogo de aplicaciones junto con las funciones que ofrecen y las interfaces entre ellas.
- **Requerimientos de la arquitectura de tecnología** - Una arquitectura de tecnología es una descripción de todos los componentes de tecnología. una arquitectura de tecnología se construye dividiendo un sistema de elementos de componentes de tecnología como lo sería una computadora servidor, una estación de trabajo de usuario, la interfaz gráfica de usuario, el manejador de base de datos multidimensional y el diccionario/depósito de datos.

La arquitectura de datos se desarrolla mediante la identificación de los diversos elementos de datos y metadatos. La arquitectura de aplicación se desarrolla mediante una lista de diversas aplicaciones y sus características y funciones. La arquitectura de tecnología

es la conjunción de las anteriores en relación a los proveedores y otros elementos de desarrollo de la tecnología.<sup>2</sup>

El desarrollo requiere que las arquitecturas de datos, de aplicación y de tecnología formuladas, se subdividan aún más en aplicaciones, interfaces, computadoras, bases de datos, comunicaciones, y pantallas de interfaz de usuario específicas. Por lo tanto, los requerimientos de desarrollo son un refinamiento de los generales de la arquitectura, con decisiones tomadas respecto a la selección de plataformas y la separación de la arquitectura de datos y la arquitectura de aplicaciones sobre las plataformas seleccionadas. También los requerimientos del desarrollo se relacionan con descripciones detalladas de la arquitectura de tecnología para la especificación de elementos tales como el lenguaje de programación, el acceso al RDBMS y los protocolos de comunicación.

- Para el bloque de Fuentes de Datos , se requieren fuentes, extracción, almacenamiento, administración y manejo de datos, redes y comunicaciones, procesadores OLTP y ambientes operativos, administración del flujo de trabajo y estándares.
- Para el bloque del data warehouse, se necesita incorporar el refinamiento y reingeniería de datos, redes y comunicaciones, procesadores de data warehouse y ambiente operativo, almacenamiento de datos en el data warehouse, catálogo de metadatos, administración del flujo de trabajo y estándares.
- Para el bloque de Acceso del Usuario Final y Herramientas, se requiere de middleware (Middleware es dentro del modelo Cliente/ Servidor, cualquier elemento intermedio de comunicación tanto de Hardware como de Software para establecer la comunicación entre el Cliente y el Servidor) de acceso y recuperación ; almacenamiento local; ambiente de datos multidimensional, navegación de metadatos y herramientas de

---

<sup>2</sup> Carol L. Burleson y David E. Tabler; Op. Cit. , p. 129-131



reporte; estándares, herramientas de usuario empresarial como análisis y reportes, modelado empresarial, minería de datos (búsqueda de patrones ocultos), navegación de datos y procesamiento analítico en línea (OLAP) y nuevas aplicaciones de producción.

• **Requerimientos de despliegue** - Los requerimientos de despliegue se relacionan con la capacidad del data warehouse para proporcionar acceso y para distribuir información de manera oportuna y conveniente. El data warehouse debe proporcionar un rango de métodos de acceso para herramientas de usuario final y aplicaciones especializadas del data warehouse. Debe además proporcionar un rango de trayectorias conectadas a una red de área local. Por lo tanto, deben considerarse los siguientes temas de requerimientos para el despliegue efectivo del data warehouse:

- Métodos de acceso.
- Métodos de recuperación.
- Herramientas de acceso.
- Requerimientos de conectividad.
- Requerimientos de plataforma cliente.

• **Requerimientos de disposición para la operación del Data Warehousing**

Además de los aspectos relacionados con el despliegue, varios requerimientos se relaciona con el aprovechamiento en la producción cuando se vuelve de Misión Crítica y se utiliza con frecuencia para apoyar las decisiones operacionales y estratégicas.

Los requerimientos de disposición para la producción se relaciona principalmente con el manejo de la solidez y la disponibilidad, la conservación de la consistencia y la precisión de la información, el manejo del desempeño al crecer el almacenamiento de datos, la definición de políticas y procedimientos para la actualización, y el mantenimiento tanto del metamodelo del data warehouse como de los datos, proporcionando control de acceso y procedimientos de seguridad. Entre los requerimientos de disposición para la producción están los siguientes:

- Mantener la consistencia, confiabilidad y actualidad de la información.
- Manejar los metadatos y el metamodelo del data warehouse.
- Asegurar que los mecanismos de transporte, base de datos, computadoras y mecanismos de comunicación estén listos y disponibles todo el tiempo.
- Proporcionar soporte técnico inmediato y capacidad de ayuda en línea para asistir a los usuarios cuando falle el sistema o cuando tengan preguntas operacionales.
- Proporcionar políticas y procedimientos de seguridad de acceso y autenticación.
- Administrar el tamaño de las bases de datos que utiliza el data warehouse, incluyendo la selección de tecnología MOLAP.
- Mejorar el tiempo de respuesta de acceso. Se obtienen mejoras en el desempeño mediante una recuperación eficiente; así como por medio del manejo del tamaño de los datos que debe mantener el data warehouse. Se logran mejoras a través del uso de arreglos, índices múltiples, mecanismos de apagado o bloqueo selectivo, purga de datos no utilizados y respaldo de datos de baja utilización.<sup>3</sup>

#### **Requerimientos para el desarrollo y despliegue del personal y sus habilidades**

Las distintas fases del desarrollo del Data Warehousing requieren de diferentes conjuntos de habilidades. Algunas de estas son muy costosas y únicamente son necesarias por períodos cortos, una útil actividad de acopio de requerimientos consiste en identificar cuáles habilidades se requieren para qué fase del ciclo de vida del desarrollo del Data Warehousing. Otro método para identificar habilidades es utilizar la arquitectura del Data Warehousing, para separar las habilidades en categorías, con base en los bloques y capas.

---

<sup>3</sup> Ibid. p. 125-127

### ***5.1.3 Requerimientos importantes del diseño***

- **Areas tema**

Estos son los temas de interés de diversas funciones organizacionales. Una selección cuidadosa de las áreas tema contiene el ámbito de implementación del data warehouse al tiempo que maximiza su utilidad.

- **Granularidad**

Se refiere al nivel de detalle de la información requerida. La granularidad tiene una relación directa con las actividades de resumen y adición que deben realizarse sobre los datos fuente. A menor granularidad, mayor cantidad de detalle. En general, los datos operacionales se consideraran como el nivel más bajo de granularidad. Para incrementar su granularidad y su utilidad para quienes toman las decisiones, los datos operacionales deben resumirse y acumularse todavía más. Por lo regular, entre mayor se la granularidad, mayor será la cantidad de procesamiento requerido para convertir y resumir los datos operacionales. Al mismo tiempo los datos con alta granularidad requieren de menos volumen de almacenamiento y se pueden también consultar con rapidez y conveniencia. El costo de resumen y adición es continuo para cada actualización del data warehouse cuando las fuentes de datos operacionales reúnen nueva información.

- **Dimensiones**

Un data warehouse organiza un gran conjunto de datos operacionales e históricos mediante múltiples dimensiones de categorización. Una importante es el *tiempo*. A los datos operacionales se les asigna un registro de tiempo (fecha y hora) en la fuente, para establecer una referencia de tiempo con los datos. Después, el data warehouse es capaz de agrupar todos los datos ocurridos dentro del mismo rango de tiempo en respuesta a una solicitud de

consulta, por ejemplo, la venta de un producto específico en el último trimestre. <sup>4</sup> Las siguientes dimensiones son de uso común en las consultas organizacionales:

- Tiempo.
- Grupos de clientes.
- Familias de productos o servicios
- Geografía y ubicación
- Estructura de organización.

En resumen, la fase de análisis del ciclo de desarrollo del Data Warehousing significa convertir los requerimientos, en un conjunto de especificaciones que puedan apoyar el diseño. El proceso de análisis consiste en derivar modelos físicos y lógicos de datos para el data warehouse y los mercados de datos y definir los procesos necesarios para conectar las fuentes de datos, el data warehouse, los mercados de datos y las herramientas de acceso del usuario final.

Es importante en el análisis definir el área o aplicación que estará directamente ligada a las utilidades de la organización; el área o aplicación que estará directamente ligada a los gastos o inversiones más altas de la organización; el área o aplicación que consumirá más recursos tanto del personal de sistemas, como recursos de cómputo, y el área o aplicación que maneja un volumen muy alto de datos.

Debido a la complejidad de la tarea de análisis y a la necesidad de habilidades especiales en tecnología y relacionadas con el dominio, ciertos fabricantes están ofreciendo "modelos de industria estándar" para industrias específicas como banca, seguros y aerolíneas. La hipótesis aquí es que las organizaciones de una misma industria comparten retos empresariales, terminología similares, así como ciclos y vínculos de negocios parecidos.

---

<sup>4</sup> Silverston, Inmon, Graziano, Op. Cit. , p. 257-270

## **5.2 La arquitectura Tecnológica**

Como a estas alturas no hay un Data Warehousing que se venda comercialmente, el área de la tecnología de la información debe construirlo a partir de los componentes disponibles. La meta de la presente metodología es ofrecer a las organizaciones más prudentes una arquitectura de solución a sus requerimientos de Data Warehousing consistente en minimizar la construcción y maximizar la compra efectiva y útil al corto plazo.

Después de analizar las necesidades organizacionales y construir una serie de requerimientos de arquitectura como resultado, necesita examinar las estructuras de los fabricantes y los productos que existen para la compra. Como parte de este examen se debe desarrollar una lista de los componentes y las técnicas relevantes del data warehouse que existen actualmente en la organización, y cuáles estándares y políticas de cómputo se aplican, en potencia, a la construcción y el despliegue del data warehouse.<sup>5</sup>

Se debe además seleccionar los componentes y las técnicas que se ajusten y se apliquen mejor, en relación a los tipos de arquitectura descritos en el Capítulo tercero. Analizar las inversiones existentes en infraestructura de comunicaciones; por ejemplo, técnicas para tecnología de la información, redes de comunicaciones, administración de redes y sistemas, así como:

- La capacidad y rendimiento de la organización en tecnología de la información.
- Las inversiones existentes, elección de plataformas de computo y tecnología de base de datos.
- Los nivel de satisfacción con los proveedores actuales.

---

<sup>5</sup> Harjinder S. Gill y Prakash C. Rao: Data Warehousing, p. 266-269.

### ***5.2.1 Análisis general de Fabricantes***

El data warehouse puede ser la aplicación determinante que conducirá las soluciones de la tecnología cliente/servidor. El data warehouse es un área en la que las organizaciones probablemente invertirán recursos significativos. Por ello, el mercado del Data Warehousing presenta con rapidez un gran número de fabricantes con ingresos por varios miles de millones de dólares, con estructuras amplias y múltiples productos. Al mismo tiempo, no hay un sólo fabricante que tenga la amplitud de productos para satisfacer por completo las necesidades.

En este ambiente, la tecnología de la información debe conducir la selección de fabricantes y productos comprendiendo tanto el mercado en tecnología del data warehouse como la dinámica de los fabricantes. Esto ayuda a comprender y administrar los riesgos. Por lo tanto, el análisis y selección de fabricantes debe realizarse desde diferentes perspectivas.

### ***5.2.2 Proveedores de soluciones versus proveedores de componentes***

Desde la perspectiva de la organización, los fabricantes de data warehouse se clasifican en dos categorías: fabricantes de soluciones completas y proveedores de productos específicos.<sup>6</sup>

Los proveedores de soluciones completas, se definen como fabricantes que proporcionan diversos componentes de la arquitectura, integrados de tal forma que el cliente tenga una solución viable que produzca las características que prometen los componentes. La integración de componentes se logra, por lo regular, a través de sociedades o alianzas estratégicas entre fabricantes de componentes, en donde uno de ellos

<sup>6</sup> Ibid. p. 262-263

actúa como líder. La meta de muchas sociedades o alianzas consiste en ampliar su oferta de soluciones para ayudar a la empresa a aprovechar sus inversiones existentes en componentes que no ofrece el fabricante de la solución, sino más bien el fabricante de componentes.

Es posible ampliar la clasificación de los fabricantes de soluciones, por el nivel de integración y ámbito de la solución que ofrecen (figura 5.1). Una solución o conjunto integrado implica que los productos componentes han sido o están siendo diseñados, desarrollados, probados y desplegados como una solución de arquitectura unificada. Invariablemente, las soluciones integradas ofrecen un mejor desempeño ya que están optimizadas para funcionar juntas. Otra ventaja importante de las soluciones integradas es el potencial de que las herramientas de vigilancia y administración se desarrollan y prueban junto con los componentes de la solución, todos dentro de la misma arquitectura.

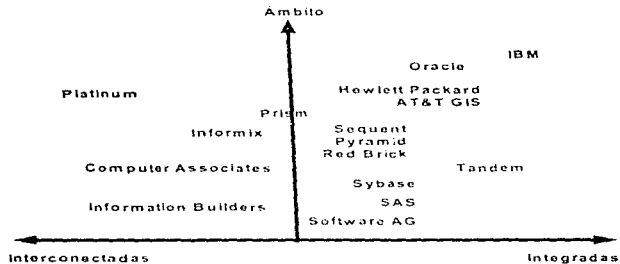


Fig. 5.1 Perspectiva del Data Warehousing desde los proveedores de soluciones integradas e interconectadas

Por supuesto la desventaja es que se reducen las opciones de componentes en la solución. Esto necesariamente se traduce en un incremento sustancial en el costo total de la aplicación además de una dependencia hacia el proveedor para futuras modificaciones y en un riesgo constante si el proveedor desaparece o sufre un rezago tecnológico.

**Los proveedores de componentes**, son fabricantes que proporcionan productos de punta para bloques específicos de la arquitectura. El área de Informática de la organización o los participantes del proyecto, deben integrar estos productos. Casi todos los proveedores de componentes son miembros de varias sociedades o alianzas que forman los proveedores de soluciones. Desde la perspectiva de la organización, el aspecto fundamental es en qué grado los programas asociados ofrecen y cumplen "una solución completa".

El riesgo con los proveedores de componentes podría compararse con el de comprar los elementos para una casa y construirla, sin embargo para proyectos donde el conocimiento del proveedor de sistemas OLTP es abundante, así como el esquema de soporte ha pasado aun etapa de madurez e independencia tal que muchos de los requerimientos de soporte son atendidos por la organización misma la comparación sería similar a adquirir solo accesorios para una casa que ya existe e instalarlos uno mismo.

Un enfoque correcto puede utilizarse en donde la organización construye el data warehouse utilizando una arquitectura racional. Se seleccionan los productos de fabricantes conocidos y el equipo de Informática es el responsable de la actividad de integración de sistemas. Se aplica la arquitectura para definir la arquitectura técnica y la funcionalidad y los productos requeridos. La estrategia consiste en seleccionar los mejores productos de su clase para satisfacer los requerimientos empresariales y las restricciones de presupuesto. Este enfoque es similar a construir una casa a la medida. La elección de los proveedores y la integración de los sistemas son obstáculos formidables.

Para muchos en un esquema así los proyectos de Data Warehousing son complejos y confusos. Haciendo a un lado las sociedades y alianzas, lo importante es qué tan bien



trabajan juntos los fabricantes para satisfacer las necesidades del cliente. La mejor forma de determinar esto es discutir las implementaciones y experiencias reales con los clientes de una sociedad.

### ***5.2.3 Estructura de fabricante***

Los principales fabricantes de datos y plataformas de cómputo ofrecen estructuras de data warehouse para influir y guiar el mercado. La mayoría de las estructuras son similares en ámbitos y sustancia, con diferencias provocadas en gran medida por la tecnología principal del proveedor. Por ejemplo, las estructuras de los de bases de datos relacionales se centran en sus ofertas de sistemas administradores de bases de datos relacionales (RDBMS). La elección de una estructura en particular limita las opciones para ciertas funciones y ofrece, en potencia, opciones de productos de lo mejor en su clase para otras funciones del Data Warehousing. Es importante hacer notar que esta metodología propone el uso de una estructura distinta en relación a la oferta de los proveedores de ROLAP. La siguiente es una lista de proveedores y el nombre de su producto:<sup>7</sup>

- IBM- Data Warehouse *Plus!*
- Oracle- Warehouse Technology Initiative (WTI).
- Hewlett Packard- OpenWarehouse.
- Sybase- Warehouse WORKS.
- Informix- Data Warehouse Framework.
- Pyramid Technology- Smart Warehouse.
- AT&T GIS- Enterprise Information Factory.
- Prism Software- Data Warehouse Framework.

---

<sup>7</sup> Ibid. p. 265

- Platinum technology- Data warehousing as a part of POEMS framework.
- SAS Institute- Data Warehousing Initiative.

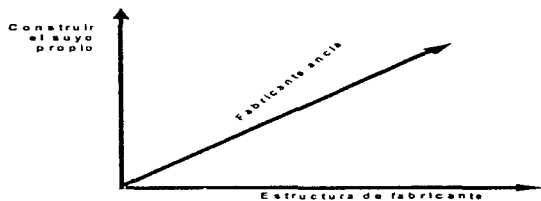


Fig. 5.2 Enfoque de elección de fabricantes

### **5.3 Diseño**

En la fase de diseño, los modelos lógicos desarrollados en la fase de análisis se convierten en modelos físicos. Los procesos identificados en la fase de análisis para conectar las fuentes de datos con el data warehouse, el data warehouse con los mercados de datos y el data warehouse/mercado de datos con las herramientas de estación de trabajo del usuario final, se convierten en diseños para programas que realizarán las tareas requeridas por los procesos.

También se identifican y detallan los procesos que requiere, de manera interna, cada bloque de la arquitectura. De ellos destacan el diseño detallado de la arquitectura de datos y el diseño detallado de la arquitectura de aplicación. Las actividades del diseño de la arquitectura de datos comprenden las siguientes:

- Desarrollo de modelos físicos de datos para las bases de datos de almacenamiento del data warehouse y mercado de datos. En forma adicional, pudiera requerirse también el

desarrollo de modelos físicos para el almacenamiento local necesario para las herramientas de usuario final.

- **Correspondencia de los modelos físicos de datos de las fuentes de datos con los modelos físicos del data warehouse/mercados de datos.** Esta correspondencia ayuda a los procesos de extracción, refinamiento y reingeniería a efectuar sus funciones dentro del data warehouse.

El diseño de la arquitectura de aplicación está comprendido de manera amplia en las siguientes familias de aplicaciones:

- **Procesos que son internos a las fuentes de datos y se relacionan con depuraciones o extracciones parciales de información y procesos que conectan las fuentes de datos con el data warehouse (o los mercados de datos si no necesitan un data warehouse).**
- **Procesos que son internos al data warehouse y se usan para fines de manejo interno.**
- **Procesos que conectan al data warehouse con los mercados de datos (si existieran estos)**
- **Procesos que son internos a los mercados de datos (si se usan) y se emplean para fines de manejo interno.**
- **Procesos que conectan al data warehouse (o a los mercados de datos) con herramientas de usuario final.**
- **Procesos que son internos al data warehouse y a los mercados de datos y para iniciar herramientas de análisis**
- **Procesos que sustentan tareas de manejo, administración y prácticas internas para el data warehouse como sistema.**

El diseño de un DW deberá, necesariamente, estar definido en forma menos precisa que el diseño de sistemas operacionales. Esto se debe a que estos últimos automatizan procesos de negocios perfectamente bien definidos, mientras que un DW está orientado a mejorar el proceso de toma de decisiones, que es un proceso definido en forma mucho menos precisa. Entendiendo la precisión la fase de diseño consta de los siguientes componentes:

- Diseño de la base de datos
- Requerimientos de la extracción de datos
- Diseño del sistema de extracción
- Preparación de los datos
- Diseño de los metadatos
- Diseño de la administración de los datos

En esta etapa también es factible incluir la creación de un prototipo, siempre tomando en consideración el tamaño del proyecto y siempre que sirva para verificar la funcionalidad y viabilidad del software y hardware elegidos.

#### **5.4 Construcción.**

La fase de construcción es aquella donde se implementan físicamente los diseños desarrollados durante la fase anterior. Por medio de decisiones juiciosas “hacer en comparación con comprar”, es posible integrar una solución de data warehouse bastante rápido, además, mediante una sopesada incorporación de las inversiones existentes es posible acelerar la tarea de desarrollo. La construcción del data warehouse es similar a la construcción de un sistema de base de datos relacional grande. La mayoría de las aplicaciones que se necesitan construir son las siguientes:

- Programas que creen y modifiquen las bases de datos para el data warehouse y los mercados de datos. Los fabricantes ofrecen capacidades de generación automática para dichos programas.
- Programas que extraigan datos de fuentes relacionales y no relacionales. Algunos fabricantes ofrecen facilidades para la generación automática de estos programas.

- Programas que realicen transformaciones de datos, tales como integración, resumen y adición. También los fabricantes ofrecen facilidades para la generación automática de estos programas.
- Programas que realicen actualizaciones de bases de datos relacionales.
- Programas que efectúen búsquedas en bases de datos muy grandes. Varias herramientas de usuario final han optimizado utilerías de búsqueda para las consultas que generan. También varios fabricantes de bases relacionales ofrecen utilerías de aceleración de consultas y capacidades para recuperación en paralelo.

Las convenciones de nomenclatura (estándares); calendarización de procesos; carga del sistema; carga de los metadatos; seguridad, y administración, son elementos indispensables durante la construcción del Data Warehousing. En esta etapa también se incluye la puesta en operación a modo de prueba.

La explotación del Data Warehousing utilizando herramientas que cumplan con las reglas de OLAP, es la prueba decisiva de la utilización. El plan de pruebas si bien no requiere ser tan exhaustivo como de un sistema operacional si requiere de las pruebas mínimas necesarias tales como:

- Pruebas de sincronía de datos
- Pruebas de resultados de operaciones estadísticas
- Pruebas de colocación en rangos de datos
- Tiempos de respuesta
- Tiempos de despliegue
- Tiempos de almacenamiento temporal.

- **Programas que realicen transformaciones de datos, tales como integración, resumen y adición. También los fabricantes ofrecen facilidades para la generación automática de estos programas.**
- **Programas que realicen actualizaciones de bases de datos relacionales.**
- **Programas que efectúen búsquedas en bases de datos muy grandes. Varias herramientas de usuario final han optimizado utilerías de búsqueda para las consultas que generan. También varios fabricantes de bases relacionales ofrecen utilerías de aceleración de consultas y capacidades para recuperación en paralelo.**

Las convenciones de nomenclatura (estándares); calendarización de procesos; carga del sistema; carga de los metadatos; seguridad, y administración, son elementos indispensables durante la construcción del Data Warehousing. En esta etapa también se incluye la puesta en operación a modo de prueba.

La explotación del Data Warehousing utilizando herramientas que cumplan con las reglas de OLAP, es la prueba decisiva de la utilización. El plan de pruebas si bien no requiere ser tan exhaustivo como de un sistema operacional si requiere de las pruebas mínimas necesarias tales como:

Pruebas de sincronía de datos

Pruebas de resultados de operaciones estadísticas

Pruebas de colocación en rangos de datos

Tiempos de respuesta

Tiempos de despliegue

Tiempos de almacenamiento temporal.

## **5.5 Implantación**

La implantación del Data Warehousing debe responder los retos de instalación, puesta en servicio y uso de la solución de data warehouse con lo que se implica una capacitación en el uso de la herramienta y la administración de su operación. Las actividades necesarias para la implantación son:

- Proporcionar la instalación inicial, incluyendo facilidades para las conexiones básicas de datos con las fuentes y para la actualización y sincronización de datos.
- Proporcionar capacitación y orientación a todo tipo de usuarios.
- Planeación e implementación de la actualización de plataformas y el mantenimiento necesario por el Data Warehousing cuando se requiere.
- Proporcionar la administración de usuarios y sistemas.
- Proporcionar la capacidad de generar archivos permanentes y respaldos.
- Proporcionar la capacidad de recuperación.
- Asegurar la integración dentro de la infraestructura existente.
- Proporcionar controles de acceso y seguridad.
- Asegurar la completa disponibilidad y los procesos para manejar caídas de los sistemas y sus componentes de infraestructura.

De la implantación se deriva la liberación del sistema con manuales y documentación técnica, con esto debe terminar también el proyecto como desarrollo comenzando una etapa donde el compromiso es la operación y sus mejoras tanto en uso como en administración de la información.

## ***6 Proyecto para la implementación de un Data Warehousing***

### ***6.1 Estrategia General***

Como se mencionó, desarrollar un sistema para el soporte a la toma de decisiones utilizando Data Warehousing debe iniciarse con un estudio serio de los requerimientos y las necesidades de la organización. La Universidad Nacional Autónoma de México como una organización que requiere de diversos elementos informativos para la toma de decisiones estratégicas debe iniciar un proyecto que permita a diferentes instancias utilizar los grandes volúmenes de información acumulados en un acervo, que ofrezcan un servicio valioso para la planeación y en su caso para una mejor presupuestación.

Considerar la problemática completa de la U.N.A.M. para diseñar e implementar un modelo de Data Warehousing y así auxiliar a diferentes dependencias a la toma de decisiones sería definitivamente una tarea demasiado amplia y ambiciosa, para este trabajo de tesis considerando la gran variedad de dependencias, con misiones distintas, problemáticas dispares, plataformas, equipos y fuentes de datos incompatibles, etc. y en este intento por modelar completamente la gran cantidad de necesidades de información estratégica, correría el riesgo de perder su objetivo.

La gran cantidad de servicios de información requeridos por diferentes áreas resalta más aún, la necesidad de llevar principalmente información estratégica a todas las dependencias, pero sólo temas comunes, es decir, una consulta organizacional importante para las autoridades académicas no necesariamente es lo que requieren para su planeación y toma de decisiones las áreas administrativas.



Si bien es cierto que la necesidad de información para la toma de decisiones es muy amplia, también la diversidad de los orígenes por lo que la estrategia correcta es la elección de una arquitectura, en el capítulo tercero se describieron una serie propicia de ellas, la estrategia general propuesta se fundamenta en una solución para esta diversidad.

Así, la estrategia es entonces ir de lo particular a lo general, es decir, elegir un problema específico con sus necesidades limitadas, resolverlo y replicar la solución como se plantea en la arquitectura Data Warehousing de Mercado de Datos (Capítulo Tercero) esta solución elimina el modelo completo y general de la organización que en caso de la UNAM sería muy complicado elaborar.

## ***6.2 Descripción del problema***

La Dirección General de Asuntos del Personal Académico de la Universidad Nacional Autónoma de México, tiene entre sus funciones la recopilación, análisis y presentación de información del Personal Académico, vertiente en la que se inscribe la publicación "*Estadísticas del Personal Académico de la UNAM*"; obtenidas de la Nómina General de la UNAM. La información se presenta en un libro conocido como "Libro azul" donde la información se encuentra desglosada en cuatro series estadísticas: Serie "A", Serie "B", Serie "C" y Serie "D".

La Serie "A" hace alusión a las diferentes poblaciones del Personal Académico que se pueden obtener, dependiendo de la unidad de registro (Personal Académico de la UNAM en su conjunto, Personal Académico en las Dependencias y Nombramientos) y de las diferentes clasificaciones de entidades académicas que se consideran (Grupos de Dependencias, Dependencias y Subdependencias).

La Serie "B" se refiere al Personal Académico en las Dependencias y a las variables sociodemográficas ya mencionadas; la información se clasifica Por Dependencias (69 Dependencias) y Por Grupos de Dependencias (9 Grupos); asimismo , se presentan tablas estadísticas y gráficas para cada una de las variables consideradas.

La Serie "C" muestra, para el total del Personal Académico de la UNAM Tanto por Subdependencias, como para cada uno de los grupos, la descripción de las variables consideradas de la nómina; también, para esta serie se describe al Personal Académico por Figura (Ayudante de Profesor y de Investigador, Técnico Académico, Profesor de Asignatura, Profesor de Carrera e Investigador) y personal que labora por Prestación de Servicios.

La serie "D" tiene como objetivo presentar el comportamiento de la Planta Académica de la UNAM durante el periodo 1987-1996; esta serie contiene tablas estadísticas y gráficas del Personal Académico en las Dependencias de acuerdo con algunas de las variables sociodemográficas que se abordan en las series "A" y "B"; Edad, Antigüedad Laboral Académica, Figura, Categoría y Nivel Ocupacional. Tipo de Nombramiento o Contrato Académico y Número de Personas por Número de Horas Académicas Contratadas, así como información obtenida directamente de la Nómina Generala de la UNAM (1ª quincena de julio 1988-1996), referida a las Horas Académicas Contratadas de las diferentes Figuras Académicas.

Para poder llevar a cabo comparaciones periódicas y lo más objetivas posibles, se estableció desde entonces, que el análisis estadístico de la información contenida en la nómina tendría una periodicidad anual, tomando como fecha de referencia la primera quincena del mes de julio. La Nómina General de la UNAM contiene información administrativa que es factible utilizar para propósitos estadísticos de carácter académico. Por ello, desde 1987 la DGAPA utiliza los datos de esta fuente, seleccionando la población para su estudio y determinado variables de la estructura programática de la nómina lo que permite elaborar estas estadísticas.

La información almacenada en las bases de datos de la DGAPA aborda variables sobre la magnitud y distribución de la Planta Académica. La periodicidad anual de la publicación antes mencionada, hace posible construir series de tiempos y hacer el seguimiento sistemático del número de profesores, investigadores, técnicos académicos, ayudantes de profesor, de investigador y del Número de Horas Académicas Contratadas, información que ha demostrado su utilidad para la elaboración de diagnósticos y estudios orientados a la definición de problemas y toma de decisiones en la Institución.

El tener como fuente de información la Nómina General de la UNAM, presenta la ventaja de que estos datos se encuentran referidos a una misma fecha, asimismo, se cuenta con información de todas las dependencias donde labora Personal Académico y que, en el universo estadístico, sólo faltan los individuos cuyo nombramiento se encuentra en trámite al momento de emitir la nómina de la primera quincena del mes de julio.<sup>1</sup>

### **6.3 Objetivos de la organización (D.G.A.P.A.)**

La Dirección General de Asuntos del Personal Académico (DGAPA) de la U.N.A.M., fue creada con el propósito fundamental de impulsar la superación del Personal Académico de la Universidad. Para cumplir con su cometido, la DGAPA apoya a las dependencias académicas y colabora con autoridades centrales universitarias y órganos coligados competentes en todas aquellas actividades de recopilación, análisis y presentación de información; en el diagnóstico y evaluación sobre el Personal Académico que permitan emprender programas de superación, estabilización y fortalecimiento de la Planta Académica; así como en la administración de estos programas.

---

<sup>1</sup> La descripción del problema así como, los objetivos de la D.G.A.P.A. fueron tomados de: Estadísticas del personal Académicos de la UNAM 1996, que es una publicación anual.

#### ***6.4 Justificación del proyecto***

La necesidad de un proyecto que ponga a la vanguardia tecnológica a la U.N.A.M. siempre es justificable debido al compromiso que tiene la máxima casa de estudios con la investigación y el conocimiento de todo el país, además la importancia de la confianza depositada en sus autoridades para la correcta dirección y toma de decisiones hace de esto un punto clave el uso de información estratégica para la organización.

Dada la infraestructura, equipos, software y conocimiento del personal el proyecto se limita exclusivamente a un buen desarrollo del concepto de Data Warehousing, la DGAPA cuenta con lo descrito además, de con la base de datos Sybase residente en un equipo SUN Sparc10, de donde se obtienen y almacenan los datos históricos, de 10 años de estadísticas obtenidas de la nómina general de la UNAM.

La inversión como se describirá más adelante se limitará a una adquisición de software. Las posibilidades de desarrollar un prototipo que pueda servir de modelo para el resto de la comunidad universitaria que cuenta con equipos similares son muy altas, cada escuela y facultad tendrá equipos similares y manejadores de base de datos idénticos para su administración escolar por lo que las posibilidades de crear sus propios almacenes de datos en el sentido académico y organizacional de cada una de ellas es enorme, sin eliminar la posibilidad de crear un gran modelo organizacional de toda la UNAM.

La existencia de esta publicación como muchas otras que muestran datos estadísticos de la U.N.A.M. los cuales son útiles para la toma de decisiones, reflejan la necesidad de la comunidad universitaria por la valiosa información que contienen, si a la existencia de esta información añadimos las posibilidades que ofrece un Data Warehousing con el manejo inteligente de la información, el resultado será asombroso.

#### **6.4.1 Beneficiados directos del proyecto**

La infraestructura de red instalada de la UNAM dentro y fuera del campus de Ciudad Universitaria, la pone como una de las redes más grandes de América Latina, por el número de nodos de Internet, usuarios atendidos y servicios provistos. Esta misma infraestructura que conecta a cada Facultad, Escuela e Instituto, así como a otras muchas dependencias de la U.N.A.M. sería la mejor vía de utilización del presente proyecto.

Cada Facultad tiene un cierto número de personal académico, básicamente profesores de distintas categorías y niveles, los Institutos cuentan también con investigadores y técnicos académicos, por lo que su necesidad de información se limita a consultas donde una variable (la dependencia) es constante; sin embargo las instancias académicas, asesores de la rectoría y otras instancias requieren no solo conocer los resultados de consultas específicas por facultad, sino resúmenes globales y otras funciones multidimensionales de la totalidad de los datos.

Los beneficiados serían primero observados en los niveles tácticos, estratégicos directivos de la U.N.A.M. incluyendo decididamente instancias académicas pero también administrativas, además de que los beneficios serían llevados a niveles donde las Facultades, Escuelas e Institutos, podrían utilizar la información que les beneficiara, como consultas mucho más específicas que acarrearían mejores planeaciones y toma de decisiones internas sobre el personal académico.

Una lista preliminar de beneficiados podría ser la siguiente:

- El Consejo Universitario.
- El Rector y sus asesores.
- Los Consejos Académicos de Área.

- La Secretaría General con sus dependencias como la Coordinación de Programas Académicos y la Coordinación de Servicios Académicos, entre otros.
- La Secretaría Administrativa con la Dirección General de Personal y la Dirección General de Normatividad y Sistemas Administrativos.
- La Secretaría de Planeación.
- Entre otros.

Dentro de un ámbito un poco más táctico los beneficiados podrían ser:

- La Coordinación de Humanidades, donde se ubican Centros de Investigación en Áreas de Humanidades.
- La Coordinación de la Investigación científica que concentra a los Institutos de Ciencias Exactas y Experimentales.
- Las Facultades y Escuelas
- Entre otros.

## **6.5 Análisis de requerimientos**

### ***6.5.1 Requerimientos de los procesos***

Existen dos tipos fundamentalmente, de usuarios diferentes, los primeros requieren de resultados estratégicos y los segundos tácticos, es decir la visión de las dependencias enumeradas en primera instancia debe considerar todos los valores almacenados en resúmenes y cálculos, las enumeradas posteriormente requerirán una especificidad en sus consultas limitada en su dependencia pero con las mismas operaciones principales.

Se requiere entonces una aplicación abierta, que pueda utilizarse fácilmente, donde los usuarios utilicen elementos gráficos para construir sus propias consultas con los campos que les interesan, así como los cálculos que desean obtener de ellos.

Se requiere seguridad en los accesos a la información, esta información representa un trabajo valioso y una inversión de la UNAM en sistemas que deben ser bien aprovechados, por lo que deberán entregarse una serie planeada de claves de acceso y validación de estas a los niveles más altos de las dependencias participantes.

Se requiere que las aplicaciones tanto de los niveles tácticos como estratégicos emitan al menos los reportes de la misma forma que aparecen en el libro azul, tratándose específicamente de cada uno.

Se requiere que las aplicaciones cliente, también generen los gráficos al menos de la misma forma en que aparecen en el libro azul para las mismas consultas.

Se requiere que las consultas puedan ser depuradas con otros criterio totalmente nuevos, basados en variables distintas. (Pivoteo)

Se requiere un tiempo de respuesta por consulta de no más de 7 segundos, usando la infraestructura de RedUNAM en el campus de C.U. y no más de 30 segundos fuera de él.

Los procesos completos de consulta generarán un archivo dentro del cliente de las últimas 15 consultas realizadas, esto con el fin de revisar consultas antes echas en comparación con actuales.

### ***6.5.2 Requerimientos de los datos***

Desde el primer levantamiento censal que se realizó en la DGAPA (1981) se planteó la dificultad para determinar el número exacto de personas que conforman la población en estudio, dado que el Personal Académico que labora en la Institución puede trabajar en más

de una dependencia de la UNAM y tener más de un nombramiento en cada lugar donde presta su servicios.

Por lo anterior, se definieron tres conceptos que atienden a las distintas formas de obtener la unidad de registro censal (el individuo):

- a) Personal Académico de la UNAM en su conjunto
- b) Personal Académico en las Dependencias
- c) Nombramientos

Estas formas de referirse a la unidad de registro censal, surgen de la diferencia que existe entre un total de individuos que trabajan como académicos en la Universidad y el conjunto que resulta de la suma del Personal Académico adscrito a las distintas dependencias, denominado "Personal Académico en las Dependencias ". Estos dos conceptos del Personal Académico difieren del denominado número de nombramientos en la UNAM, ya que una misma personal puede tener uno o más nombramientos en una o varias dependencias.

De esta manera, se determinó el universo estadístico con la totalidad de personas que trabajan en la Universidad en calidad de Personal Académico; es decir, todas las personas que tienen alguna o algunas de las figuras, categorías y niveles académicos estipulados en el Estatuto del Personal Académico de la UNAM.

La estructura de los datos se establece con base en tres conceptos que atienden a las distintas formas de obtener la unidad de registro censal:

- Personal Académico de la UNAM (personas)
- Personal Académico en las Dependencias
- Nombramientos

ESTA TESIS NO DEBE  
SALIR DE LA BIBLIOTECA



Esta estructura permite determinar el número exacto de personas que conforman la población bajo estudio. La población se clasifica por:

- Grupos de Dependencias
- Dependencias
- Subdependencias;

A partir de estas clasificaciones, se analizan algunas características demográficas y laborales del Personal Académico tales como:

Edad

Figura

Categoría

Nivel Ocupacional

Tipo de Nombramiento o Contrato Académico

Antigüedad Laboral Académica

Número de Horas Académicas Contratadas por Persona

Número de Horas Académicas Contratadas por Figura

## **6.6 Elección de plataforma y Diseño**

La plataforma sugerida para una metodología como la descrita en el capítulo anterior, dadas las necesidades específicas de la U.N.A.M. es la ofrecida por Sybase, que ha sido proveedor de tiempo atrás. Su solución en Conjunto se presenta como Sybase Warehouse WORKS y pretende con las mejores herramientas del mercado en materia de middleware promover su producto multidimensional para soluciones de Data Warehouse en mercados de Datos.

Conocido como Sybase IQ, el producto permite diseño multidimensional y no exige altos requerimientos en Hardware ni por el lado de procesamiento ni por almacenamiento, su tecnología "bit-wise" (bit listo, se refiere a la inteligencia de un solo bit en columnas y renglones que modelan una o varias bases de datos en operación), ofrece diversos beneficios sobre otros productos en el mercado (Ver ANEXO).



Modelo propuesto de explotación de datos.

FIGURA 6.1

La tecnología bit-wise ofrece una gran cantidad de beneficios destacando el almacenamiento, Sybase ofrece una compresión de datos, es decir un modelado de los datos existentes en las bases de datos en operación de aproximadamente un quinto del total del tamaño de la base, por ejemplo una base que en operación mide 1 GB, modelada utilizando bit-wise solo utilizaría alrededor de 200 MB.

Otra ventaja de la utilización del Sybase IQ es el número de operaciones de lectura escritura que tiene que establecer por consulta prototipo, es decir, mientras un manejador de bases de datos relacional hace alrededor de 125,000 operaciones de entrada/salida, Sybase IQ hace menos de 235, con el mismo tamaño de página de memoria.

La reducción en los índices es dramática, mientras un manejador relacional crece hasta en un 80 % dependiendo de la complejidad de la normalización de la base, Sybase se mantiene en la proporción de un quinto de la base en operación. La diferencia radical en el desarrollo es la calidad del modelado multidimensional que permita representar solo con falso y verdadero los valores de información requeridos.

Como se describe en el análisis de los datos, las diversas variables de las estadísticas del personal académico se relacionan como se muestra en el "libro azul" se pueden formar diversas combinaciones para obtener información relacionada, pero se depende totalmente de los datos mostrados ahí, algún otro dato debe ser recalculado por el interesado.

Mientras que en el manejador relacional se almacenarían los datos de la siguiente forma:

RFC	Categoría	Nivel	Dependencia	...etc.
XXXXX999999XXX	Tec. Acad. Aux	A	FCA	
XXXXX999999XXX	Tec. Acad. Aso	B	DGSCA	

En un modelado multidimensional en relación a la información que realmente se requiere debe elaborarse una tabla probablemente con un gran número de columnas que se construirían de la siguiente forma:

Spongamos que la categoría es un elemento importante de decisión, además del nivel, y la dependencia para construir la tabla se puede modelar con varias dimensiones dentro de las cuales se encuentran fundamentalmente la dimensión de la categoría, relacionada con el nivel y la dependencia. Con la utilización de Sybase IQ un modelo de datos un pre-diseño podría mostrarse de la siguiente manera:

Tec.	Tec	Otras	Nivel A	Nivel B	Nivel C	Dependen	Dependen	Otras
Acad	Acad	Académicas				U.S.	DGSCA	
Aso	SI	Aux	SI	Académicas				
0	1	-----	1	0	0	1	0	-----
1	0	-----	0	1	0	0	1	-----

En consecuencia las consultas organizacionales no pre-establecidas pueden realizarse con gran facilidad utilizando los conteos de las columnas y renglones, por

ejemplo: ¿Cuántos Académicos Auxiliares A tiene la Facultad de Contaduría y Administración?, la respuesta será el recorrido rápido donde la Dependencia sea igual a FCA, el Nivel igual a A y la categoría igual a Auxiliar.

Otras consultas podría completarse exitosamente con alta velocidad sin haberse establecido previamente, tales como:

¿Cuántos Académicos tiene la DGSCA?

¿Cuántos Académicos Nivel A tiene la FCA sin importar su categoría?

¿Cuántos Académicos no tienen categoría de investigador ?

De la misma forma en que se establecieron las variables anteriores como variables de decisión se pueden establecer otras en relación a las variables expuestas en el análisis de los datos tales como, la edad que podría modelarse en rangos de edad de la misma forma en que aparecen en la publicación de las estadísticas.

En general todas las consultas pueden ser obtenidas de esta forma y el desarrollo de la aplicación o en su defecto la adquisición de una herramienta existente en el mercado para la explotación de estos datos puede distribuirse entre los tomadores de decisiones de las organizaciones interesadas en aprovechar estos datos para sus planeaciones internas y globales.

Es factible construir aplicaciones específicas para la problemática de la UNAM utilizando los modelados de datos como base para hacer clientes ligeros y que incluso podrían ser explotados mediante páginas de HTML o lenguaje JAVA, con esto se abarataría el desarrollo y los usuarios no podrían tener mas libertad y apertura en las plataformas utilizadas dentro de cada dependencia.

## Conclusiones

La toma de decisiones rápidas y correctas en las organizaciones hoy en día puede ser la diferencia radical entre una organización y otra, esto puede afectar significativamente su permanencia en el mercado, utilidades, presupuestos u otros factores tan significativos como los económicos.

El proceso de toma de decisiones, como muchos otros procesos racionales, requiere de abundantes elementos de información para el análisis y la evaluación de las posibles alternativas, dentro de las organizaciones una ventaja competitiva puede ser proveer de mayor y mejor información a los tomadores de decisiones, por lo que la utilización de sistemas de soporte a esta toma de decisiones es una necesidad plenamente justificada.

Data Warehousing es una tecnología relativamente reciente que permite modificar los procesos de toma de decisiones radicalmente, su propuesta innovadora, frecuentemente ha caído en mitificaciones sobre sus costos y factibilidad. Por el contrario, Data Warehousing es un concepto que impacta los subprocesos de análisis y evaluación de alternativas de decisión de manera ágil y con información abundante, sin afectar la operación de los sistemas transaccionales de la organización.

El concepto de Data Warehousing como herramienta útil de explotación de información, rebasa los intereses de los proveedores tanto de Hardware como de Software que han sido los principales promotores así como las grandes consultorías de soluciones integradas, del Data Warehouse como monstruo devorador de presupuestos.

Es posible hacer Data Warehousing con relativamente pocos recursos, atendiendo problemáticas específicas y generalizando posteriormente. Es conveniente elaborar

desarrollos prototipo y replicarlos posteriormente, donde la complejidad de las organizaciones es demasiado amplia.

Como cualquier desarrollo, Data Warehousing requiere de establecer un proyecto claro, con objetivos bien delimitados y con una metodología accesible. Las metodologías pueden variar dependiendo del tipo de arquitectura que se intenta desarrollar con el Data Warehousing.

La U.N.A.M. como organización sin fines de lucro debe incorporar la utilización de Data Warehousing a sus procesos de planeación y presupuestación por el compromiso de correcta utilización de recursos públicos y vanguardia tecnológica que tiene para con todo el país.

## **ANEXO Proveedores**

### **Hewlett Packard**

Hewlett Packard conduce su oferta de data warehouse mediante su programa Open Warehouse, que se caracteriza como una estructura para construir data warehouses basada en componentes de hardware y software de primera calidad de HP y de otros fabricantes. Las anclas de lo que ofrece HP son sus plataformas UNIX de alto rendimiento y su producto Intelligent Warehouse, un software de administración del data warehouse. La estructura Open Warehouse permite elegir un RDBMS, las herramientas de refinamiento y reingeniería, y las herramientas de acceso de datos. El programa de sociedades de Open Warehouse se orienta específicamente a reclutar socios que ofrezcan las opciones mencionadas.

El programa Open Warehouse también ofrece metodología de consultoría y servicios para un rápido despliegue del data warehouse. Intelligent Warehouse incluye una de las pocas herramientas para vigilar el uso del data warehouse. Open Warehouse es inoperante fuera de equipos HP, por lo que la arquitectura también basada en OLAP relacional se hace más compleja para operar en esquemas donde se requiere más apertura.

### **IBM**

La solución Data Warehousing de IBM se denomina: A Data Warehouse *Plus!* El enfoque de IBM consiste en entregar un conjunto completo de productos y servicios; su meta está en ofrecer una solución integrada con base en una sola arquitectura. Es probable como lo indican muchos expertos que IBM tiene la mejor solución integrada de data warehouse, basada en una sólida arquitectura". La familia DB2 es el ancla de la estrategia del data warehouse de IBM. Sin embargo los costos son impresionantes.

IBM tiene la ventaja de que la mayoría de los datos operacionales que se van a extraer y almacenar en el data warehouse residen en sistemas IBM. De ahí que la integración apegada sea un resultado natural. El reto en este momento es que casi todos los productos IBM son para plataformas

**IBM. La oferta de IBM en esta área es bastante débil. Actualmente, IBM tiene un programa de sociedades para reclutar más socios de productos y servicios.**

**IBM ofrece tres soluciones:**

- **Mercado de datos independiente** - Se concentra en un departamento o función empresarial de la organización, se maneja con una ayuda mínima de la organización en tecnología de la información.
- **Mercado de datos dependiente** - Es similar al mercado de datos independiente, pero la organización en tecnología de la información controla y administra la conectividad con las fuentes de datos.
- **Data warehouse global** - Lo implementa y administra la organización en tecnología de la información, y se apoya en una arquitectura de empresa. Esto puede implicar un data warehouse centralizado o uno distribuido con los mercados de datos.

Una solución global de data warehouse puede basarse en DB2 para MVS, o DB2 para AIX Parallel Edition. La solución visual para el data warehouse se asienta en DB2 para OS/2 o DB2 para AIX, y se ofrece como el punto de entrada de bajo costo. IBM aborda la administración de metadatos con la familia DataGuide. En la minería de datos, IBM tiene una familia reciente de herramientas de descubrimiento de conocimientos. Las técnicas de descubrimiento de conocimientos que aplican estas herramientas son las asociaciones, los patrones secuenciales, los clasificadores y el agrupamiento. Además, IBM Research ofrece a Parallel Visual Explorer, una poderosa técnica de análisis para visualizar espacio dimensional con las coordenadas paralelas, una alternativa de representación geométrica para datos multidimensionales.

## **Oracle**

Oracle ofrece una solución amplia, con un enfoque en las atribuciones principales en el almacenamiento y la administración de datos, las aplicaciones de mercado vertical y las herramientas de acceso y desarrollo de datos. La solución del data warehouse se caracteriza por dos atributos: la extensión de la línea de productos de ORACLE, y la cantidad de socios en su Warehouse Technology Initiative



(WTI). Oracle también aprovecha su Systems Management Tools Initiative (SMTI) para cubrir las necesidades de administración de sistemas y vigilancia del desempeño.

La fuerza de la oferta de Oracle proviene de su motor RDBMS, Oracle7, el cual mejora constantemente para satisfacer los requerimientos de funcionalidad del data warehouse, de sus aplicaciones de mercado vertical que ofrecen el potencial de data warehouses prefabricados, la amplitud de su tecnología para desarrollo y análisis de datos y la disponibilidad de productos de software de otros fabricantes. Sin embargo la solución de Oracle para Data Warehousing se sustenta en la arquitectura de análisis de datos OLAP relacional, montado en equipos de gran capacidad que soportan pesadimas consultas optimizadas en su manejador relacional.

## SYBASE

Las estrategias corporativas de Sybase se concentran en tres mercados: el procesamiento en línea de las transacciones, el data warehousing y el soporte de decisiones, y el "despliegue masivo" de información en toda la empresa. Su estrategia de data warehouse se incluye en su estructura "Warehouse WORKS". La fortaleza de Sybase en el soporte a ala toma de decisiones reside en su motor de middleware conocido como: OmniCONNECT, y en su herramienta de desarrollo: Powerbuilder.

Sybase ofrece una de las mejores heramientas del mercado como manejador de bases de datos multidimensionales con modelado de datos, el Sybase IQ, su oferta para consultas inteligentes y modelado multidimensional en mercados de datos lo ha hecho bastante exitoso en varias aplicaciones de misión critica a nivel mundial. Sybase IQ se presenta como una opción muy razonable para el modelado multidimensional por su rápida explotación de información sustentada en la tecnología de bit-wise que modela los datos de bases en operación y permite hacer consultas organizacionales no predeterminadas con gran facilidad y rapidez.

## BIBLIOGRAFÍA

BARQUIN Ramon y EDELSTEIN Herb editores, Planning and Designing the Data Warehouse, USA, 1997, 311 pp.

De esta compilación se utilizaron las participaciones de:

BARQUIN Ramon, A Data Warehousing Manifiesto

LIVINGSTON Glen y RUMSBY Bob, Database Design for Data Warehouses: The Basic Requirements.

BURLESON Carol L. y TABLER David E., Data Warehousing: Putting it All Together.

RADEN Neil , Coping the righth OLAP technology

HERB Edelstein, An Introduction to Data Warehousing

NEAL, C. Doug, How to Justify the Data Warehouse and Gain Top Management Support

Estadísticas del Personal Académico de la U.N.A.M. 1996, México, C. U., 1996, 416 pp.

HARJINDER S. Gill y PRAKASH C. Rao, Data Warehousing, México, 1996, 382 pp.

KONTZ, Harold y WEHRICH, Heinz, Administración, 10ª ed., México, 1993, 746 pp.

ORFALI Robert, HARKEY Dan y EDWARDS Jeri, The Essential Client/Server Survival Guide, 2ª ed., U.S.A., 1996, 676 pp.

SILVERSTON Len, INMON H. William, CRAZIANO Kent, The Data Model Resource Book, U.S.A., 1997, 355 pp.

## HEMEROGRAFÍA

Boletín, Sybase Warehouse WORKS, U.S.A., 1996, no numerado.

JENKS Bruce, "Data Warehouse de niveles (Tiered Data Warehouse)", Soluciones Avanzadas, Año 4, Número 34, junio 1996, 56 p.

McDERMONT Karl, "Muchos datos poca información útil para el análisis y toma de decisiones", Soluciones Avanzadas, Año 4, Número 34, junio 1996, 61 p.

MIRAMÓN Comons Bernardo, "Data Warehousing, estrategias generales de implantación", Soluciones Avanzadas, Año 4, Número 34, junio 1996, 47 p.