



Universidad Autónoma de Querétaro
Facultad de Informática
Maestría en Ingeniería de Software Distribuido

**Técnicas para Sistemas de Inteligencia de Negocios en
las organizaciones para la toma de decisiones**

Que como parte de los requisitos para obtener el grado de :

Maestro en Ingeniería de Software Distribuido

Presenta:

Rodolfo Luna Pérez

Dirigido por:

DR. Ubaldo Chávez Morales

SINODALES

DR. Ubaldo Chávez Morales

Presidente

Firma

M.C. Alberto Lamadrid Alvarez

Secretario

Firma

M.I.S.D. Carlos Alberto Olmos Trejo

Vocal

Firma

M.S.I. Gerardo Rodriguez Rojano

Suplente

Firma

M.I.S.D. Juan Salvador Hernández Valerio

Suplente

Firma

M. en C. Ruth Angélica Rico Hernández
Director de la Facultad

DR. Luis Gerardo Hernández Sandoval
Director de Investigación y Posgrado

Centro Universitario
Querétaro, Qro.
Febrero 2009
México

La presente obra está bajo la licencia:
<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.es>



CC BY-NC-ND 4.0 DEED

Atribución-NoComercial-SinDerivadas 4.0 Internacional

Usted es libre de:

Compartir — copiar y redistribuir el material en cualquier medio o formato

La licenciante no puede revocar estas libertades en tanto usted siga los términos de la licencia

Bajo los siguientes términos:



Atribución — Usted debe dar [crédito de manera adecuada](#), brindar un enlace a la licencia, e [indicar si se han realizado cambios](#). Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante.



NoComercial — Usted no puede hacer uso del material con [propósitos comerciales](#).



SinDerivadas — Si [remezcla, transforma o crea a partir](#) del material, no podrá distribuir el material modificado.

No hay restricciones adicionales — No puede aplicar términos legales ni [medidas tecnológicas](#) que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia.

Avisos:

No tiene que cumplir con la licencia para elementos del material en el dominio público o cuando su uso esté permitido por una [excepción o limitación](#) aplicable.

No se dan garantías. La licencia podría no darle todos los permisos que necesita para el uso que tenga previsto. Por ejemplo, otros derechos como [publicidad, privacidad, o derechos morales](#) pueden limitar la forma en que utilice el material.

Resumen

Algo que día con día se palpa dentro de las organizaciones actuales es la gran competencia debido a la globalización y apertura de nuevos mercados, por lo que la necesidad de información estratégica para la toma de decisiones es determinante para crear diferencia entre una organización y otra. De ahí es que Business Intelligence (BI) viene a revolucionar la forma en que las organizaciones accesan y utilizan la información. Muchas de dichas organizaciones cuentan con gran cantidad de datos refundidos en sus servidores, pero mientras estos datos no se conviertan en información valiosa para la compañía no dejan de ser un activo muerto. El secreto es: ¿Cómo transformar esos activos de información en conocimiento para la toma acertada de decisiones?

Este proceso de transformar los datos en conocimiento se lleva a través de actividades como la recolección, integración y organización de la información para los cuales tenemos que hacer uso de técnicas de datawarehouse y herramientas de tipo OLAP (Online Analytical Processing). En donde un datawarehouse, básicamente es un repositorio de información integrada de varias fuentes que sea fácilmente accesada por los usuarios. El otro punto que menciono en la construcción de un BI es la parte de OLAP que nos ayudará con la parte de la extracción de la información.

Este documento presenta una metodología que ayude en la creación de una arquitectura de BI dentro de una organización, inyectando de información valiosa que contenga el conocimiento y ayude en la toma oportuna de decisiones.

(Palabras Clave: BI, Datawarehouse, OLAP)

Summary

Something that can daily be detected among existing organizations is the great amount of competition due to globalization and the opening of new markets. As a result, the need for strategic information for decision making is a determining factor in creating a difference between one organization and another. Business Intelligence (BI) has revolutionized the way in which organizations access and utilize information. Many of these organizations have a large quantity of data collected in their servers, but until these data can be converted into valuable information for the company, they are dead assets. The secret is: how to transform these knowledge information assets so they help in decision making. This process of transforming data into knowledge is carried out through activities such as the collection, integration and organization of information for which we must use datawarehouse techniques is basically a repository for information from different sources that can be easily accessed by users. The other point I mention in the construction of a BI is the OLAP part which helps in extracting information. This paper presents a methodology which helps in the creation of a BI architecture within an organization, injecting valuable information that contains knowledge and aids in decision making.

(Key words: BI, Datawarehouse, OLAP.)

Dedicatorias

A Dios:

Por darme la fuerza de vivir día con día.

A mi Esposa e Hijos:

Maria Soledad, Marysol y Rudy por todo el cariño, apoyo, comprensión además de estar siempre conmigo y apoyarme incondicionalmente.

A mis familiares y amigos:

Por expresarme su cariño en cada momento.

Agradecimientos

Agradezco a mi asesor, el Dr. Ubaldo Chávez Morales, por su guía, críticas y consejos durante la asesoría y revisión de este trabajo de tesis.

A los catedráticos de la Maestría por compartir sus conocimientos y experiencia, enriqueciendo mi formación académica.

A mis compañeros y amigos de la Maestría por compartir conmigo los momentos de trabajo y convivencia.

A la Universidad Autónoma de Querétaro, por las atenciones brindadas a través de su personal: docentes, administrativos y de servicio, facilitando mis labores y haciendo más agradable mi estancia durante la maestría.

Índice

Página

<i>Resumen</i>	<i>i</i>
<i>Summary</i>	<i>ii</i>
<i>Dedicatorias</i>	<i>iii</i>
<i>Agradecimientos</i>	<i>iv</i>
<i>Índice</i>	<i>v</i>
Capítulo 1	1
Introducción	1
1.1. Antecedentes y Justificación	2
1.2 Definición del Problema	3
1.3 Objetivos Generales y particulares.	4
1.4 Alcances y Limitaciones	5
1.5 Organización del Proyecto	6
Capítulo 2	8
Sistemas de Inteligencia de Negocios (BI)	8
2.1 Introducción	8
2.1.1. La Era de la Información.....	9
2.1.2. El Valor de la Información.....	10
2.1.3 ¿Por qué las organizaciones requieren distintos sistemas de información?.....	13
2.1.4. La Información que las empresas necesitan.....	16
2.2. Origen de BI	19
2.3. Definición de BI.	21
2.4. Técnicas de Análisis BI.	25
2.4.1 Sistema de Información Ejecutiva (Executive Information System –EIS-).....	25
2.4.1.1 ¿Qué es un EIS?	26
2.4.1.2 ¿Información para todos?	28
2.4.1.3. Lo que debe contener un EIS	30
2.4.2 Sistema de Soporte de Decisiones (Decision Support System -DSS-).....	33
2.4.2.1 ¿Qué es un DSS?	33
2.4.2.2 ¿Qué debe contener un DSS?	36
2.4.3 Reporteo.....	37
2.4.3.1. Herramientas de consulta y reporte.	40
2.4.3.2. Características de una herramienta de reporte.	42
2.4.4. Minería de Datos.....	43
2.4.4.1. Qué es y qué no es la minería de datos.	44
Capítulo 3	45
Fase de Preparación	45

3.1 Definición del Equipo de Trabajo.....	45
3.1.1 Preferencias de trabajo	46
3.1.2. Tipos y preferencias de trabajo	47
3.2. Identifique a los Beneficiarios.....	52
3.3. ¿Quién necesita soluciones de Business Intelligence?.....	53
3.4. Defina Metas.....	54
3.4.1. ¿Qué puede hacer Business Intelligence?	55
3.5. ¿Construir o Comprar?.....	58
3.6. Software de Business Intelligence	58
3.6.1 Cuadrante mágico de posicionamiento.	60
3.7. Examine la situación Actual.....	61
3.8. Preguntas Estratégicas de Negocio que el BI va a servir para aclarar.	62
3.9. Identifique Posibles Fuentes de Información.....	64
3.9.1. Internet y Servicios de Bases de Datos Comerciales	65
Capítulo 4.....	67
Análisis y Diseño	67
4.1 Fase de Preparación.....	67
4.2 Análisis de Requerimientos.	67
4.2.1 Propiedades de los Requerimientos.	71
4.2.2 Personal involucrado en la Ingeniería de Requerimientos	73
4.2.3 Técnicas y herramientas utilizadas en el análisis de requerimientos.....	74
4.2.4 Definir los Límites del Sistema.....	75
4.2.5. Uso de listas de control para el análisis de requerimientos.	77
4.2.6 Proveer Software de Apoyo en las Negociaciones.....	78
4.2.7 Planear los conflictos y resolución de conflictos.	80
4.2.8 Priorizar los Requerimientos.....	82
4.2.9 Clasificación usando acercamiento dimensional.....	83
4.2.10 Uso de Matrices de interacción para encontrar conflictos y traslapes.....	84
4.2.11 Evaluar el riesgo del requerimiento	86
4.3 Evaluar la calidad de los Datos.....	88
4.3.1. Introducción.....	88
4.3.2 Proceso de Evaluación de la Calidad.	90
4.3.3 Ejemplo ilustrativo.....	90
4.3.4 Modelando el sistema de integración de datos.....	91
4.3.5. Modelo de DAG.....	92
4.3.6 Propiedades asociadas al cálculo de la calidad.	94
4.3.7 Algoritmos de Evaluación.....	97
4.4 Modelo de Datos de Negocio.....	97
4.4.1 Introducción.....	97
4.4.2 Definición de Base de Datos	98
4.4.3 Los modelos de datos.....	98
4.4.4 Objetivos del modelo de datos.....	99
4.4.5 Modelo de entidades y relaciones (MER)	100
4.4.6 Ejemplo.....	103
Capítulo 5.....	105

Implementación	105
5.1 Modelo Físico de Datos (Creación del Datawarehouse).....	105
5.1.1. ¿Por qué se justifica construir un Datawarehouse?	106
5.1.2. Comenzando a construir un DW	107
5.1.3. Análisis del Problema	108
5.1.4. Construcción de un DW	109
5.1.4.1 Definición de una arquitectura	109
5.1.4.2. Construcción en Incrementos, Datamarts.....	110
5.1.5. Consideraciones de Implementación mediante DataMarts.	111
5.1.6. Metodología de desarrollo de DW	112
5.1.7. Evolución del DW	113
5.1.7.1 Evolución de la Implementación	113
5.1.7.2 Evolución de la Arquitectura	113
5.1.8. Conclusiones y comentarios del DW	114
5.2 Extracción de la Información	116
5.2.1 Antecedentes.	116
5.2.3. ¿Qué es un Datamart?	117
5.2.4. Metadatos.	118
5.2.5 Definición de ETL.	118
5.2.6 Una solución universal para la integración de información.	119
5.3 Transformación de la Información en BI.	121
5.3.1 Introducción.	121
5.3.2 Minería de Datos.....	123
5.3.2.1. Qué es y qué no es la minería de datos.	123
5.3.2.2. Fases de un Proyecto de Minería de Datos	124
5.3.3 Análisis Multidimensional.	127
5.3.3.1 Hechos y dimensiones.....	128
5.3.3.2 Múltiples vistas de negocio.	129
5.3.3.3. ¿Cuándo es necesario OLAP?	130
5.4. Generación de informes.....	131
5.4.1 Informes completos y decisionales	132
5.4.2. Etapas en la creación de informes	133
Capítulo 6.....	134
Conclusiones y trabajo futuro	134
6.1. Conclusiones.	134
6.1.1. La receta.....	134
6.2. Trabajos futuros.....	136
Referencias Bibliográficas	137
Glosario	141

Capítulo 1

Introducción

La situación actual de las compañías preocupadas por la reducción del costo y aumento de rentabilidad a través de métodos superiores de administración de la cadena de valor, requieren reemplazar el método tradicional de análisis a través de habilitación de tecnología y al mismo tiempo emerger un nuevo tipo de administradores. *Administradores Multidimensionales* que operan en un mundo complejo y multidimensional de toma de decisiones.

Business Intelligence permite a estas compañías acelerar la tasa en la cual los administradores pueden procesar información permitiendo a las organizaciones acceder el gran volumen de información con el que cuentan y rápidamente traducirlo en conocimiento transformando los activos de información con los que cuenta la empresa en información valiosa para la toma acertada de decisiones.

En esta investigación se analiza el papel que juega la tecnología de Business Intelligence como mecanismo para generar una ventaja competitiva en las organizaciones. Para ello se realiza un análisis detallado de los componentes más importantes de Business Intelligence para poder llegar así a presentar un modelo de esta naturaleza.

1.1. Antecedentes y Justificación

Richard Conelly, Robin Mc Neill and Rolland Mosimann, (1996) mencionan que algo que día con día se palpa dentro de las organizaciones actuales es la gran competencia debido a la globalización y apertura de nuevos mercados, por lo que la necesidad de información estratégica para la toma de decisiones es determinante para crear diferencia entre una organización y otra. De ahí es que Business Intelligence viene a revolucionar la forma que las organizaciones accedan y utilicen la información. Muchas de dichas organizaciones cuentan con gran cantidad de datos refundidos en sus servidores, pero mientras estos datos no se conviertan en información valiosa para la compañía no dejan de ser un activo muerto. El secreto es: ¿Cómo transformar esos activos de información en conocimiento para la toma acertada de decisiones?

Los *sistemas transaccionales* conocidos como ERP (Planner Resource Enterprise) son de importancia crítica para *operar el negocio* pero...

Inteligencia de Negocios es crítica para que el **NEGOCIO** sea **EXITOSO**

LEJEUNE Miguel (2001) dijo *INTELIGENCIA DE NEGOCIOS* es el proceso de obtener y analizar la información necesaria para la toma de decisiones. La mayoría de las veces, esa información está disponible dentro de la empresa, pero se halla oculta entre los datos del negocio que se guardan en archivos y bases de datos.

“SI TU LO PUEDES **MEDIR**, TU LO PUEDES **ADMINISTRAR**”. Kaplan & Norton (Jan-Feb 1996)

1.2 Definición del Problema

Actualmente las organizaciones que carecen de sistemas para la toma de decisiones se encuentran ante 2 grandes problemas **GENERAR** y **OBTENER** información:

GENERAR información	OBTENER información
<ul style="list-style-type: none">• Existen muchos requerimientos continuos de información al área de sistemas	<ul style="list-style-type: none">• Al tener recolección manual de los datos, los tiempos de respuesta son lentos
<ul style="list-style-type: none">• Islas de Información entre las áreas de la organización	<ul style="list-style-type: none">• Falta de información oportuna
<ul style="list-style-type: none">• Recolección Manual de Datos	<ul style="list-style-type: none">• Poco tiempo para analizar la información
	<ul style="list-style-type: none">• Información no integrada ni consolidada
	<ul style="list-style-type: none">• Mecanismos inflexibles para analizar la información

Al analizar e implementar una metodología que ayude en la creación de una arquitectura de BI dentro de la organización, inyectamos de información valiosa que contenga el conocimiento y ayude en la toma oportuna de decisiones.

1.3 Objetivos Generales y particulares.

El Objetivo general de esta tesis es realizar el análisis de las técnicas necesarias para el establecimiento de una arquitectura de información que transforme datos en información de negocio para la toma de decisiones dentro de una organización a través de la recolección, extracción y organización de la información, utilizando técnicas de Datawarehouse y herramientas tipo OLAP(Online Analytical Processing).

Los objetivos particulares son:

- Establecimiento de una metodología para la construcción de una arquitectura de Bussiness Inteligent a través del análisis dentro de una organización.
- Permitir un pensamiento multidimensional en la petición de la información para la toma de decisiones cubriendo las necesidades de clientes, procesos internos, innovación/aprendizaje y finanzas.

1.4 Alcances y Limitaciones

La tesis consiste como se mencionó en el objetivo general en realizar el análisis de las técnicas necesarias para el establecimiento de una arquitectura de información que transforme datos en información de negocio para la toma de decisiones dentro de una organización a través de la recolección, extracción y organización de la información, utilizando técnicas de Datawarehouse y herramientas tipo OLAP(Online Analytical Processing).

Entre los alcances logrados están los siguientes:

- Determinar la arquitectura necesaria para un sistema de inteligencia de negocios.
- Determinación de fuentes de información.
- Establecimiento de Indicadores de Negocio.
- Descripción de herramientas para extracción de la información.
- Técnicas de Datawarehouse.
- Selección de herramientas tipo OLAP.
- Permitir el análisis multidimensional de la información.
- Visualización de Información.

1.5 Organización del Proyecto

El presente proyecto estará dividido en 6 capítulos en los cuales mediante **Método Analítico** .el cual consiste fundamentalmente en el estudio de todas y cada una de las partes que componen el objeto en estudio, es decir; en este método separaré la investigación en las partes que la forman y estudiaré a cada una de éstas de manera independiente pero guardando su relación y función dentro de el todo en el objeto de la investigación. Dentro de estas fases se estará haciendo uso de encuestas y entrevistas las cuales deberán ser medidas cualitativa y cuantitativamente.

Los Capítulos y secciones de esta tesis son los siguientes:

- **Capítulo 1:** Este capítulo contiene la introducción, los antecedentes, la justificación, la definición del problema, los objetivos de la tesis y su organización.
- **Capítulo 2:** Capítulo en donde se describe el estado actual de la Inteligencia de Negocios.
- **Capítulo 3:** Este capítulo consiste en la primer parte dentro de la arquitectura que corresponde a la fase de preparación donde se integrará el equipo de trabajo, definirán metas, se determinará el tipo de preguntas que el BI deberá contestar y se identificarán las fuentes de información.
- **Capítulo 4:** Capítulo de análisis y diseño donde se realizará el análisis de requerimientos, se evaluará la madurez de las fuentes de información y se establecerá el modelo de datos de negocio.
- **Capítulo 5:** En el desarrollo del capítulo de implementación se establecerá el modelo lógico de datos transformado en un Datawarehouse, establecimiento de reglas y herramientas para la extracción de la información, realización de minería de datos y visualización de la información.

- **Capítulo 6:** Conclusiones obtenidas de la tesis y trabajo futuro.
- **Referencias Bibliográficas:** Lista de referencias bibliográficas y ligas a documentos WEB utilizados en la realización de la tesis.

Capítulo 2

Sistemas de Inteligencia de Negocios (BI)

2.1 Introducción

Desde que las organizaciones comenzaron a guardar los datos de sus operaciones en medios de almacenamiento físico, con el fin de permitirles una mayor administración y control de la información, ha existido de la mano una necesidad de utilizarla para atender las necesidades propias del negocio

Anteriormente, las empresas se enfocaban simplemente a producir, producir y producir. Los clientes no tenían opción de seleccionar lo que querían consumir, sino que adquirían lo que existía en el mercado. La competencia era mínima, con lo cual los consumidores eran cautivos y, al mismo tiempo, dependientes de los caprichos de la producción. La era de la industrialización se encuentra totalmente ligada a la producción en serie, acompañada de un consumo obligado, en ese sentido, el poder que podría tener la información que, normalmente va de la mano con los servicios, era mínima. A las empresas no les interesaba brindar servicios acompañados de sus productos para generar valor adicional sobre los competidores, tampoco había un interés para lograr la lealtad en sus clientes, quienes eran vistos más como un ente que cerraba el ciclo económico y no como la fuente principal de su existencia. La información y su importancia estratégica comenzaron a surgir cuando la competencia se hizo patente, y cada vez más y más productos similares, de diferentes compañías, se ponían a la venta, en ese momento el consumidor tuvo la opción de seleccionar aquello que más le conviniera o lo que más se adecuara a sus gustos y preferencias. Surge entonces la necesidad de brindar servicios adicionales para obtener la lealtad de los clientes, quienes poco a poco comenzaron a ver, no solo el producto que compraban, sino como eran atendidos, que garantías se ofrecían sobre su compra, que ventajas habría entre diferentes productos y, en general, evaluar todo lo que genera la diferenciación entre las compras que realizan.

Cuando las empresas no tienen garantizada la venta de lo que producen, realizan un cambio paulatino hacia obtener de los datos toda información útil y estratégica para mantenerse en el mercado, dándole un lugar preponderante al cliente.

2.1.1. La Era de la Información.

La Era de la Información se hace patente, detonada por la globalización de los mercados, así como su dinámica y agresividad actual.

Richard Conelly, Robin Mc Neill and Rolland Mosimann, (1996) mencionan que algo que día con día se palpa dentro de las organizaciones actuales es la gran competencia debido a la globalización y apertura de nuevos mercados, por lo que la necesidad de información estratégica para la toma de decisiones es determinante para crear diferencia entre una organización y otra.

En la Era de la Información se le da un peso específico muy importante a la información como el principal conocimiento que sostiene el negocio. Existen empresas que, de modo predominante, ofrecen servicios y giran su negocio principal sobre el manejo de la información (bancos, aseguradoras, casas de bolsa, internet, etc.), en ellas es fácil identificar la importancia de la información, si no existiera ésta dejarían de existir; sin embargo, hay otras en las que su giro principal es alrededor de la producción, en ellas la información debe identificarse para analizar y perfeccionar su producción (porcentajes de desecho, líneas de producción, distribución de materias, suministro, inventarios y almacenes, procesos internos, publicidad y mercadotecnia, preferencias del cliente, etc.). De hecho, en cualquier empresa se está tratando de convertir, por todos los medios posibles, esa información en conocimiento que mejore los procesos y, a su vez, se traduzca en ventajas competitivas en los mercados. La idea de las empresas sedientas de información no surge de súbito, en realidad desde que se almacenan los datos debe entenderse que tendrían un fin utilitario en algún momento, caso contrario, cualquier dato de control sería desechado instantáneamente. Lo que si

surge de súbito es la imprescindible necesidad de dar respuesta rápida y asertiva a los requerimientos de información e incluso anticiparse y prospectar oportunidades de negocio, basadas en los datos históricos y el motivo es muy sencillo: en un mercado altamente globalizado, agresivo y competitivo, aquellas empresas que no respondan a las tendencias actuales de comercialización y competitividad se vuelven obsoletas y sin remedio serán absorbidas o desaparecerán de la escena comercial.

2.1.2. El Valor de la Información.

En la época actual, que se caracteriza por un crecimiento exponencial de las nuevas tecnologías de la información y las telecomunicaciones, los activos más valiosos de una empresa ya no son activos tangibles o los depósitos en los bancos, sino los conocimientos, habilidades, valores y actitudes de las personas que forman parte de una empresa. De hecho, para generar riqueza es suficiente tener conocimiento sobre un tema determinado y explotarlo de la mejor manera posible. Los factores de la producción como capital, tierra y trabajo, han sido sustituidos por el capital intelectual, que comprende todos aquellos conocimientos tácitos o explícitos que generan valor económico para la empresa.

Prácticamente nadie cuestiona el hecho de que vivimos en la era de la información y que la información tiene un valor concreto en pesos, esto se evidencia por el hecho de que existen empresas cuyo único negocio es alrededor de la venta de información -Gartner Group, Empresas de Internet, Amazon, etc.-. En mercadotecnia, el conocimiento es el único camino posible para sostener ventajas competitivas. Es más, en la actualidad, la información y el conocimiento son considerados como el capital intelectual que soporta la riqueza de una organización.

Para identificar el valor concreto de la información pueden hacerse dos ejercicios:

1. Todas las corporaciones tienen un modelo del mundo de negocios basado en la información que poseen, por ejemplo: ¿Qué influencia la compra y la demanda?, ¿en dónde hay oportunidades de negocio?, ¿qué es lo que mueve la calidad del producto y la demanda de los clientes?, etc. A medida que esta información se vuelve más exacta, la capacidad de la corporación para competir se incrementa. Visto así, la información corporativa es claramente un activo de la empresa que genera valor y su inexistencia genera "desvalor", o sea, pérdidas en caso de que existiera la información y ésta desapareciera, o bien, "no ganancias" en caso de que no exista.

2. Otra forma de entender la información como dinero es mediante la transformación de ésta en conocimiento tácito o explícito. Las "organizaciones que aprenden" tienen procesos que detonan cuando hay eventos nuevos o diferentes, y trasladan las experiencias a sistemas o a individuos para enriquecer su negocio. El conocimiento tácito es el que tienen las personas producto de la experiencia, los estudios y la educación; los conocimientos explícitos son los que se almacenan en medios magnéticos como cintas, disquetes, etc. En el momento que una persona decide cambiar de empleo se está llevando consigo información, conocimientos y está vendiendo su fuerza intelectual por un mayor precio; el campo laboral nos indica que la fuerza de trabajo intelectual aumenta su costo con dos factores básicos que generan conocimiento: la experiencia y la educación. Por su parte, si un sistema que posee información eventualmente desaparece o falla, generará pérdidas a la empresa, incluso por cada minuto que esté detenido. En la actualidad las empresas están apostando mucho por la tecnología y los individuos para que juntos tengan un conocimiento suficiente que acerque la visión interna de ambos a la realidad exterior, en la medida que esa brecha disminuye, las decisiones

tomadas se acercan más a la realidad exterior, generando decisiones más asertivas y en menos tiempo; si la brecha o "Gap de información" aumenta, puede ocasionar grandes pérdidas para la organización. Es fácil entenderlo, supongamos una situación hipotética en la cual un nuevo auto es diseñado con lujo, pero con algunos toques de un auto deportivo y, sin realizar ningún tipo de estudio previo más que la intuición y el sentido común, se pretende lanzarlo para que sea adquirido por adultos mayores de 30 años. Para ello, una vez que se encuentra listo para la distribución, comienzan las campañas de publicidad y presentaciones orientadas precisamente a ese mercado potencial, al cabo de cierto tiempo se dan cuenta que las campañas que lanzaron no han tenido mucho impacto en ese segmento, pero curiosamente un porcentaje similar de las ventas a la fecha se han dado en personas entre 25 y 30 años. La realidad indica que ese auto tiene un impacto mayor en un segmento distinto al que suponía, en caso de haber tenido información suficiente sobre las preferencias de los distintos segmentos, la historia de las ventas y, sobre todo, un estudio previo de mercado se habría sabido con anticipación hacia donde dirigir los esfuerzos de las ventas con dos resultados benéficos: en primer lugar, la publicidad no habría sido inefectiva y el dinero utilizado en las campañas no se habría desperdiciado; y en segundo lugar, se habría atendido a los verdaderos clientes potenciales, con lo cual las ventas habrían sido mayores. El ejemplo es hipotético, pero la situación es muy similar a la cotidianeidad, muchas empresas utilizan el sentido común y la intuición para tomar decisiones, la información que se traduce en conocimientos acerca la visión interna a la realidad y esa diferencia existente es la que puede representar miles o millones de pesos. Lo que se pretende es acercar el mundo real a la visión interna para generar ganancias, para convertir la información en utilidades, para darle un valor a la información.

2.1.3 ¿Por qué las organizaciones requieren distintos sistemas de información?

Una gran infraestructura se dedica a la tecnología que soporta Sistemas de Información para tener completamente automatizada a la empresa. Este crecimiento tecnológico tiene distintos orígenes, que van desde la implementación, crecimiento, ampliación, integración, etc. Las condiciones actuales de los mercados han provocado, precisamente, el que sea necesaria tecnología cada vez más avanzada para responder a las peticiones muy particulares de información. Sistemas de Procesamiento de Datos (SPD), Sistemas de Manufactura, Administración de Recursos Empresariales (ERP), Sistemas de Información Ejecutiva (EIS), Sistemas Gerenciales, Manejo de Relación con Clientes (CRM), Suministro de la Cadena de Distribución (SCM), etc., son algunos de los sistemas que surgen, se ponen de moda y luego desaparecen acorde a la evolución de las empresas. Lo que es un hecho es que, independientemente del enfoque que esté de moda o sea más útil en el momento, los datos siempre serán almacenados en depósitos para el efecto y esos datos serán el soporte total a las decisiones de la empresa. En función de lo que nunca desaparecerá, podemos entender que distintos datos pueden proveernos de distinta información y al haber distinta información existen distintas necesidades que pueden ser cubiertas, dependiendo de la función que cada empleado desempeñe en la empresa.

Muchos negocios requieren específicamente información de su giro, por ejemplo, los ERP (Administración de Recursos Empresariales) son sistemas muy grandes y complejos en donde gran parte de su contenido se dedica a la producción, sería ilógico adquirir un sistema tan complejo y costoso si nuestra empresa se dedica a los bienes raíces. En ese mismo sentido existen desarrollos comercializados como productos que solo son configurados en una organización en particular, pero tienen el funcionamiento mínimo necesario para cierta industria. Es común encontrar software para la industria automotriz, software para hoteles, retail, transporte, software educativo, etc. El motivo por el cual son distintas las

herramientas utilizadas obedece a que las actividades de misión crítica, que soportan cada una de las industrias, es distinto y como tal, también es distinto el tipo de información que puede solicitar un directivo en cada una de las industrias, de ahí que mucho software dedicado a explotar la información de almacenes de datos no tenga características estándares, sino más bien es adaptable a cada situación. Considerando las distintas necesidades en cada giro, es fácil extrapolar la misma situación a cada empresa, incluso con giros similares, pero lo importante es entender el último nivel en cuanto a la diferenciación de la información solicitada.

La información que fluye en una empresa está destinada a responder a diversos tipos de preguntas de sus usuarios, de ahí la necesidad de que existan sistemas de información para requerimientos muy específicos que permitan la colección y el manejo de datos.

Al interior de una empresa, los puestos son factor importante para determinar la información que comúnmente es requerida por la gente.

Es importante resaltar que todos los sistemas de información tienen un fin muy particular, y se complementan para sostener, de la manera más eficiente, un negocio; sin embargo, no todos pueden solucionar las distintas demandas de los usuarios, precisamente porque son diseñados para alguna área de aplicación muy específica.

El motivo por el cual existen varios sistemas de información es porque los usuarios tienen preguntas muy específicas que no cualquier sistema puede resolver. De hecho, las bases de datos operacionales, que son las indispensables en cualquier organización, no están organizadas para responder a preguntas globales sino a pequeños grupos de datos. Preguntas que involucren consultas complejas podrían resolverse en un lapso extenso, en el cual cabe la posibilidad de que la vigencia desaparezca. Lo importante es destacar que una base de datos

o sistema de información no tienen la capacidad de resolver las necesidades informativas de toda la organización. Aquí algunos de los motivos:

- **Alcance de las decisiones que soporta la información.** El alcance de una decisión estratégica implica información de toda la organización. El alcance de una decisión operacional solo involucra información del departamento de interés. Si el alcance es distinto y existen diversos sistemas por departamento (RH, Nómina, Ventas, Inventarios, etc.), entonces un sistema proveerá información local y otra información global.

- **Objetivos de la información.** El objetivo del usuario operativo es que se le facilite y automatice la operación de una función específica de la empresa; el de un estratega es maximizar la función de la empresa.

- **Usuarios.** El usuario es distinto incluso en la misma pirámide organizacional. Mientras los sistemas operativos tienen interfaces muy especializadas para un usuario que realiza una operación rutinaria, los usuarios estratégicos realizan consultas variadas y no previstas de la información, por lo que los sistemas deben ser sencillos y con toda la información disponible que cubra cualquier consulta requerida, de este caso el software final debe ser orientado a un usuario en particular y, por ende, deberá adecuarse al conocimiento que tenga sobre el tema.

- **Tipos de pregunta.** Las preguntas que responde un sistema operacional son referentes a las transacciones que se realizan diariamente y a nivel registro o suma de registros de un solo tipo. Un usuario operativo realiza frecuentemente preguntas sobre registros como pueden ser el estado actual de una factura, movimientos de un cliente, cantidad surtida por un proveedor, fecha del último movimiento de un distribuidor, etc.; las preguntas de un ejecutivo pueden también ser específicas, pero se orientan más a agrupamientos de datos como pueden ser totales por zona, promedios de clientes, tendencias de ventas e incluso pronósticos. Toda esta información se encuentra de alguna forma en los

almacenes operativos, pero lanzar una consulta como las ventas totales del año anterior puede implicar hasta días en resolverse y otro tiempo para publicar los datos. Un sistema organizado para resolver preguntas de ambos tipos en el menor tiempo posible es lo ideal.

- **Cantidad de datos.** Si un usuario procesa la información de las transacciones se mueve en el nivel registro. Si un usuario procesa información de entidades, se mueve en el nivel agrupamientos de registros, obviamente la cantidad de datos que se necesitan es distinta y debe ser un sistema diferente el que provea de esa información. Para que un director o gerente, quien necesita conocer las transacciones de toda una zona para tomar una decisión, pudiera analizar cierto comportamiento, serían necesarias muchas hojas de reportes con cientos de datos. El usuario operativo que necesita pocos registros no tiene mayor problema por recibir una hoja de reportes, pero el directivo si tendría problemas con una cantidad exagerada de papeles. Se necesitan sistemas que brinden no solo la cantidad ideal, sino también que la entregue en el tiempo óptimo.

Resumiendo, existe una gran necesidad de información en muchos niveles de las organizaciones, pero hasta el momento no existe un sistema de información que esté diseñado para dar respuesta cabal a todos ellos. Cada sistema da respuesta a una parte de los requerimientos de toda la empresa para que, en conjunto, no quede un espacio vacío de información ni en tiempo, ni en forma.

2.1.4. La Información que las empresas necesitan.

La tendencia de las organizaciones actuales es demandar información en los niveles donde antes la administración se basaba en la intuición y el sentido común para tomar decisiones. A pesar de que en los niveles operativos siempre se ha demandado información, históricamente no ha existido restricción alguna para brindarla al usuario. Más bien los mercados dinámicos han obligado a las empresas para que la información estratégica sea puesta en las computadoras de los directivos, este comportamiento se ha generalizado principalmente motivado

por la facilidad y utilidad de la información compartida. En estos momentos la información fluye en todos los niveles de la organización con diferentes fines (comunicación, control, administración, evaluación, etc.) independientemente de los puestos, las empresas están entendiendo que los niveles directivos tienen una gran responsabilidad al tomar decisiones, pues el impacto que generan recae sobre toda la organización, pero también existen más personas que toman decisiones y, a pesar de que éstas no tienen un impacto global, deben ser también correctas y oportunas, pues ciertos grupos dependen de las mismas. Directores, gerentes, supervisores, jefes, todos aquellos que toman decisiones deben tener suficiente información para apoyarse en su trabajo cotidiano, el lugar que ocupen en la pirámide organizacional se vuelve secundario cuando el enfoque es hacia el manejo de procesos y todos los puestos tienen cierta relación y dependencia entre sí.

De modo general en una pirámide organizacional, los requerimientos informativos se dividen en 3 partes:

- **Información Estratégica.** Está orientada principalmente a soportar la toma de decisiones de las áreas directivas para alcanzar la misión empresarial. Se caracteriza porque son sistemas sin carga periódica de trabajo y sin gran cantidad de datos, sin embargo, la información que almacenan está relacionada a un aspecto cualitativo más



que cuantitativo, que puede indicar como operará la empresa ahora y en el futuro, el enfoque es distinto, pero sobre todo es distinto su alcance. Se asocia este tipo de información a los ejecutivos de primer nivel de las empresas. Un punto importante es que la información estratégica toma grandes cantidades de datos de áreas relacionadas y no se enfoca específicamente hacia una sola, de ahí que las

decisiones que puedan ser tomadas impactan directamente sobre toda la organización.

- **Información Táctica.** Información que soporta la coordinación de actividades y el plano operativo de la estrategia, es decir, se plantean opciones y caminos posibles para alcanzar la estrategia dictada por los directivos. Se facilita la gestión independiente de la información por parte de los niveles intermedios de la organización. Este tipo de información es extraído específicamente de un área o departamento de la organización, por lo que su alcance es local y se asocia a gerencias o subdirecciones.

- **Información Técnico Operacional.** Se refiere a las operaciones tradicionales que son efectuadas de modo rutinario en las empresas mediante la captura masiva de datos y sistemas de procesamiento transaccional. Las tareas son cotidianas y soportan la actividad diaria de la empresa (contabilidad, facturación, almacén, presupuesto y otros sistemas administrativos). Tradicionalmente se asocian a las Jefaturas o Coordinaciones operativas o de tercer nivel.

Si consideramos factores internos y externos de una organización podríamos concluir que los requerimientos actuales se orientan a conocer y eficientar los costos de toda la cadena económica. Este se refleja en el interés por tener a la mano los diagnósticos que nos arrojen información específica y clave para determinada área de conocimiento, en el menor tiempo posible. A pesar de que la tendencia actual es que las áreas directivas necesiten en su escritorio la información clave de su empresa; en todos los niveles el requerimiento es similar aunque, evidentemente, tiene objetivos diferentes. El paradigma de la información exclusiva en los niveles directivos para apoyar la toma de decisiones no es obsoleto, simplemente se complementa agregando la información también en otros niveles medios y jefaturas, o sea, en cualquier persona que tenga el poder de tomar decisiones.

2.2. Origen de BI

Los Sistemas de Información son bastante recientes si los comparamos con otras áreas de conocimiento. La historia del procesamiento de información ha existido prácticamente desde el inicio de las computadoras. En esa época las aplicaciones se corrían de forma individual en archivos maestros que almacenaban los datos en medios magnéticos, con la limitante del acceso secuencial. El problema del acceso secuencial es la necesidad de recorrer todos los registros antes de encontrar el que se está buscando. Además, al almacenar operaciones individuales, pronto existió una cantidad enorme de cintas y medios de almacenamiento con redundancia, que hacía difícil su administración.

El concepto de Business Intelligence surgió a partir de la formulación de estrategias efectivas de negocios que respondieran a los nuevos tiempos y sus demandas. En los 70s, cuando aparecen los dispositivos de almacenamiento de acceso directo, se da un gran paso en cuanto a la velocidad para acceder a los datos, pues con ellos las búsquedas ya no eran lineales, sino directas. Junto con ello también aparecen sistemas de administración de bases de datos (DBMS), cuyo propósito era facilitar al programador el acceso a la información al encargarse del almacenamiento e índices. En ese entonces se definió a la base de datos como "una fuente única de información para todo el procesamiento". Tanto la implementación de las bases de datos para el procesamiento en línea como las nuevas tecnologías y lenguajes de cuarta generación (4GL), proveyeron al usuario la facilidad de tener el control directo de los sistemas y la información, dando origen a los primeros sistemas de información formales. Pero también fueron el inicio del paradigma de una sola base de datos que pudiera servir tanto al procesamiento operativo como al procesamiento de alto nivel. Hay que verlo de la siguiente manera, la información almacenada en las bases de datos mantiene el registro total de lo que sucede en la organización. Cuando un usuario operativo desea consultar transacciones unitarias o pequeños grupos de transacciones, se puede acceder directamente y extraer la información en un lapso muy breve (milisegundos); si la base de datos no es muy extensa, incluso un gerente puede

también realizar una consulta (vía asistentes) que traiga información resumida sobre muchos registros e, incluso, sobre toda la base de datos, los tiempos para consultas de ese tipo son muy razonables (segundos). ¿Pero que pasa cuando los datos sobrepasan los límites permisibles para tener información a la mano? Algunos de los factores causales de que las bases de datos se volvieran poco operativas para consultas extensas son las fusiones, la globalización, las alianzas, la diversificación de productos, el crecimiento exponencial de las empresas y, en general, todas las condiciones derivadas de la evolución natural de las empresas que trajo consigo un aumento cuantitativo de los datos que se necesitaba almacenar. A esto hay que agregar que las herramientas o software necesario para obtener la información era muy especializado y rara vez una persona que toma decisiones tenía el background necesario para manipular información. La información primitiva se volvió muy extensa y poco práctica para cierto tipo de consultas, había que desarrollar nueva tecnología que permitiera derivar información calculada o sumariada para satisfacer las necesidades de la administración, además, la información primitiva representa el valor actual, es utilizada y operada en procesos repetitivos, por lo tanto, es posible su modificación. La información derivada no puede ser actualizada porque, normalmente, contiene valores históricos, es operada y utilizada por procesos que se ejecutan aleatoriamente. La información primitiva es operacional apoyando las funciones de empleados, la información derivada es para el soporte de decisiones que normalmente apoya a administradores y ejecutivos. Teniendo tantas diferencias es complicado entender que la misma información pueda residir en una misma base de datos. De hecho así fue, la gente que toma decisiones demandó respuesta a sus necesidades en los almacenes operativos encontrando muchas limitantes en tiempo y forma, a fin de cuentas es posible satisfacer los requerimientos, pero no como los necesita quien toma decisiones, la tecnología intervino para desarrollar arquitecturas especializadas para que resida la información según su tipo y que sea operada por el software adecuado para desempeñar sus funciones, ya sea recorrer la compañía en el caso de la información operativa o dirigir el rumbo de la compañía en el caso de la

información que soporta decisiones. **Business Intelligence** es un concepto que se asocia 100% con los niveles directivos y, por razones obvias, realmente ahí es donde surge, de la necesidad de contar con información para dirigir el rumbo de la empresa por los altos mandos, sin embargo, con el tiempo se ha ido ampliando el alcance de este término hasta llegar prácticamente a toda la empresa, pero no hay que dejar de lado que incluso en la actualidad un gran porcentaje de niveles directivos siguen dictando las reglas del rumbo de este término. A pesar de relacionar completamente el término con conceptos 100% computacionales, sobre todo las herramientas utilizadas para lograr implementar un desarrollo de este tipo, la verdad es que el concepto no se construye basándose en herramientas computacionales, sino de la formulación de estrategias efectivas de negocios que respondieran a los nuevos tiempos y sus demandas. El énfasis es en los requerimientos y de ahí se desprenden las aplicaciones, es decir, los hombres de negocio dictan las necesidades y la gente técnica investiga y adapta la tecnología para resolver favorablemente esos requerimientos con todos los medios a su alcance. *BI se plantea como una sinergia entre los tomadores de decisiones y las herramientas que emplean, la tecnología está claramente vinculada a la administración de las empresas, teniendo como resultado obtener ventajas competitivas, producto de decisiones mejor informadas. En función de esta idea podemos entender a BI como una combinación de tecnología y desarrollo de negocios.*

2.3. Definición de BI.

Como muchos otros conceptos o términos, el de Business Intelligence no escapa a la diversidad de interpretaciones. Se justifica su uso y se entiende el que sea considerado como una tecnología de información, pero no existe un acuerdo en cuanto a su definición.

LEJEUNE Miguel (2001) dijo *INTELIGENCIA DE NEGOCIOS* es el proceso de obtener y analizar la información necesaria para la toma de decisiones. La mayoría de las veces, esa información está disponible dentro de la empresa, pero se halla oculta entre los datos del negocio que se guardan en archivos y bases de datos.

"BI is an umbrella to describe a set of concepts and methodologies designed to improve decision-making in business through the use of facts-based systems. Fact-based systems consist of executive information systems, decision-support systems, online analytic processing, and newer technologies such as data mining, data visualization, and geographic information systems."**Howard Dresner**

“Es el conjunto de tecnologías que permiten a las empresas utilizar la información disponible en cualquier parte de la organización para hacer mejores análisis, descubrir nuevas oportunidades y tomar decisiones más informadas.”

Analizando las definiciones, queda primeramente claro que BI no es un software, sistema o herramienta específica, es más bien un conjunto de tecnologías que van desde arquitecturas para almacenar datos, metodologías, técnicas para analizar información y software entre otros, con un fin común para el apoyo a la toma de decisiones. A partir de elementos comunes es que podemos obtener una definición que abarca BI en cuando a su utilidad y funcionalidad en las empresas.

1.- Información. Es la esencia de BI. El fin último es proveer de información al usuario final para apoyarlo en la toma de decisiones, y esta información puede provenir tanto de los almacenes operacionales como de arquitecturas diseñadas específicamente para el análisis -Data Mart, Data Warehouse. Sea cual fuere el caso, el usuario puede necesitar información de cualquier fuente primitiva o derivada para apoyarse en su labor, para lo cual BI utiliza o construye fuentes de datos o de información interna o externa, que son la principal materia prima de esta Tecnología.

2.- Apoyo a la Toma de Decisiones. Un sistema que exclusivamente brinde información no representa lo que se busca con BI, una segunda característica consiste en organizar y presentar los datos relevantes para que puedan verdaderamente apoyar una toma de decisiones. Esto implica tecnologías, técnicas de análisis y todo aquello que sea necesario para obtener de los datos, solo aquella información relevante y útil a la labor del usuario. Recordando el origen de BI, surge en la toma de decisiones para obtener ventajas competitivas, producto de decisiones mejor informadas. Si su origen cae en el desarrollo de negocios, es lógico entender que BI sea un apoyo para tal efecto. BI abarca cualquier forma de organizar información, siempre y cuando sostenga la toma de decisiones.

3.- Orientación a usuario final. Un factor que dio pauta a pensar en nuevas tecnología para explotar información fue que el usuario final no poseía conocimientos técnicos que le permitieran tener un acceso sencillo y directo a los datos operacionales, pues esa área está reservada para informáticos. Por tal motivo, el usuario final no tenía de primera mano la información que necesitaba y las consultas no definidas, que son las que tradicionalmente realiza un ejecutivo, eran realizadas por terceras personas (léase secretarios, asistentes técnicos o gente de sistemas) con la dependencia consecuente. BI incluye herramientas de explotación de información orientadas a usuario final, para eliminar la dependencia de terceras personas. Se pretende brindar las facilidades necesarias para que, con la tecnología, el usuario actúe solo. Las herramientas de BI son sencillas, intuitivas y fáciles de entender y usar; pueden tener diversos fines, como son: Informar, reportar, permitir análisis, identificar tendencias, proyectar, etc. Cualquiera que sea su función final, el común es el mismo: orientación a usuario final.

Considerando los elementos comunes en cualquier definición, puede implementarse una definición más acabada.

"Business Intelligence es una combinación de tecnologías de colección de datos y manejo de información, que implementa soluciones orientadas al usuario final para apoyar la toma de decisiones, aprovechando la información estratégica disponible en cualquier parte de la organización." **Arturo Mendoza**

Para la colección de datos usa o construye almacenes de datos y los maneja con técnicas de análisis y herramientas orientadas al usuario final. Los almacenes de datos son las fuentes operacionales (bases de datos, archivos de texto, hojas de cálculo, administradores de archivos, etc.), almacenes de datos operacionales (ODS), bases de datos externas, data warehouse y data marts. Las técnicas de análisis principales son los sistemas de información ejecutiva (EIS), sistemas de soporte de decisiones (DSS), data mining (DM) y herramientas de reporte (RT). BI es un término "agrupador". El que sea considerado como un conjunto de conceptos le da un poder enorme, pues pueden integrarse funciones que tradicionalmente estaban separadas, tales como el acceso de datos, reporte, explotación, pronóstico y análisis. De ese modo, al menos en la actualidad en empresas grandes, BI se ha convertido en un apoyo indispensable para la toma de decisiones, en cualquier nivel de la organización y mucha gente está explotando el potencial estratégico de los datos operativos. Bien utilizada, BI puede ser un arma estratégica de la gente de negocios, sustentada en tecnología de sistemas.

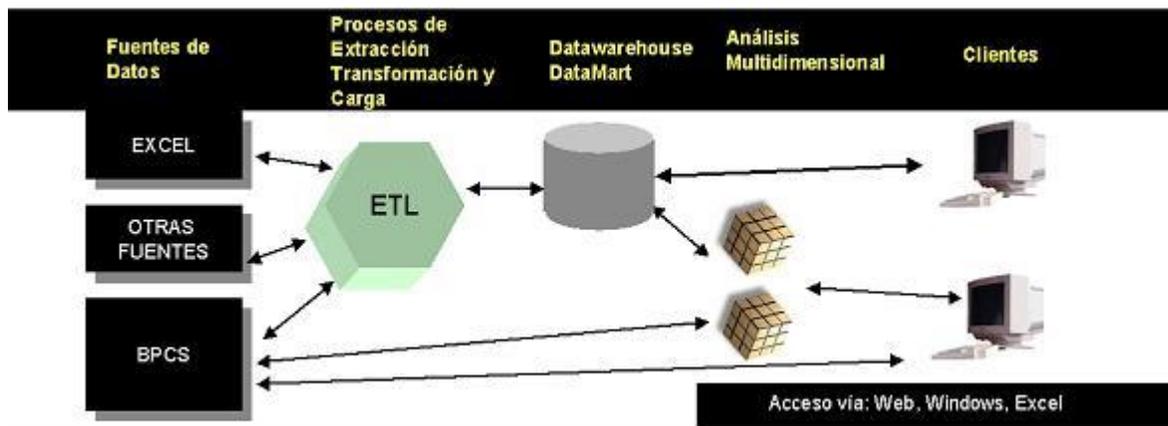
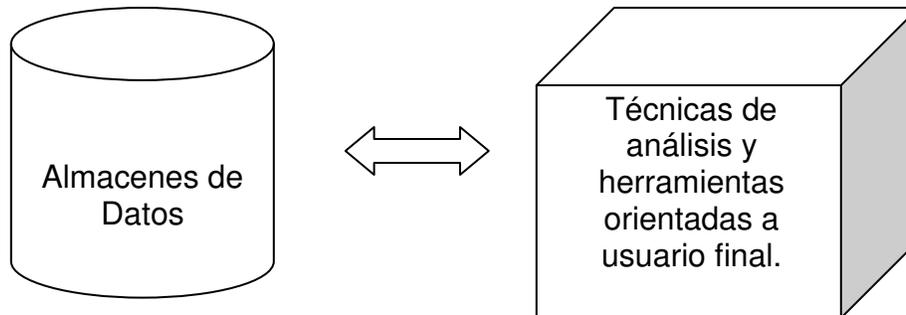


Figura que representa el Set de conceptos que integra BI.

2.4. Técnicas de Análisis BI.

Como se mencionó anteriormente BI se integra por dos grupos:



Las técnicas mayormente utilizadas en BI son:

- Sistemas de Información Ejecutiva (Executive Information System-EIS-)
- Sistema de Soporte de Decisiones (Decisión Support System-DSS-)
- Herramientas de Reporteo.
- Minería de Datos

2.4.1 Sistema de Información Ejecutiva (Executive Information System –EIS-)

Desde que las grandes empresas iniciaron su camino para que sus ejecutivos manipularan la información como deseaban para lograr una visión completa del negocio, hubo muchos intentos y malas acogidas motivadas por razones de tecnología, costos o simplemente cultura. El Data Warehouse fue un esquema de información que albergaba datos para efectos exclusivamente de análisis y estadísticos, con este avance en el almacén, los esfuerzos se centraron en utilizarlo y proveer de la información que necesitaba el ejecutivo. Los primeros intentos se centraron en la incursión al mercado de soluciones con alarmas, instrumentos de consulta, sistemas expertos y mucho más, pero la evolución, incluso a la fecha, es hacia alcanzar a colocar en una pantalla la mayor cantidad posible de datos para realizar análisis gráficos, visuales y rápidos basados en técnicas básicamente de consolidación, agrupamiento y tendencia. Esto dio origen a los Sistemas de Información Ejecutiva.

2.4.1.1 ¿Qué es un EIS?

“EIS se refiere a cualquier sistema de software que muestre información ejecutiva de las diferentes áreas del negocio en un solo sistema, facilitando el monitoreo de la empresa.” **David Abdo**

El EIS es una técnica de más alto nivel dentro de herramientas de BI. Tiene como objetivo primordial proveer de toda la información necesaria a la gente que toma decisiones, de modo fácil y prácticamente sin necesidad de interactuar con el mismo sistema. En términos formales, "un EIS es un sistema de información que permite a los ejecutivos acceso rápido y efectivo a información compartida, crítica para el negocio, utilizando interfaces gráficas." [González, Marisol]. Las interfaces que son utilizadas en estos sistemas deben ser más sofisticadas que los sistemas transaccionales y deben incluir, en el menor número de páginas posible, la mayor cantidad de información que el usuario necesita para monitorear su empresa. La función principal de un EIS es el monitoreo de la empresa o de una área de negocio específica. Se debe entender que el monitoreo implica una visualización de lo que se quiere controlar y, por tal motivo, no debe existir mayor esfuerzo por parte del usuario que simplemente ver las consultas que cotidianamente realiza. La información que un EIS presenta es producto de indicadores empresariales que cotidianamente son evaluados para analizar su comportamiento y lo que permite determinar su cualidad son las propiedades que le son asignadas a cada indicador. Por ejemplo, la cuota de ventas para cada vendedor en una empresa es un indicador de su productividad dentro de la organización. Un vendedor que haya logrado una cuota de 9,000 USD el mes anterior, tiene una cantidad que, por si sola, no dice mucho, no se sabe si es buena o mala la cantidad que logró vender. La compañía previamente diseña las "propiedades" de la métrica, que son indicadores para evaluar el comportamiento. Por información histórica se determina que menos de 8,000 USD está considerado como una cuota baja, entre 8,000 y 10,000 USD se encuentra en un rango normal, y mayor de 10,000 USD es una cuota alta que merece un bono adicional equivalente a un porcentaje sobre la

diferencia. A cada uno de los calificadores se le asignan ciertas características con la intención de identificar rápidamente el rango en que se encuentran. Con esta explicación se podría entender que en este caso, las consultas para determinar las ventas son predefinidas, un EIS siempre debe estar a la mano. Para este ejemplo, un director de ventas puede simplemente abrir una consulta predefinida que le presente aquellos vendedores con una cuota superior a 10,000 USD, quienes se harán acreedores a un bono adicional; de la misma forma puede visualizar fácilmente quienes no hayan logrado su cuota para tomar decisiones correctivas. En caso de que el director deseara ver más información relacionada con un vendedor y entrar a detalle a ver sus ventas, cantidades o clientes, estaría entrando a un Sistema de Soporte de Decisiones (DSS), el cual, a pesar de que en conjunto con un EIS, forman una poderosa solución de BI, pero merece una explicación independiente por la diferencia de su uso. Debido a la función principal de monitoreo, un EIS es considerado como un sistema muy fácil de usar y de desarrollar, pero con funcionalidades analíticas muy limitadas. La diferencia de los EIS con otros sistemas no solo es la vistosidad y facilidad de uso. Aparte del front-end, los EIS interpretan y manipulan de forma diferente la información, pues trabajan con formatos de datos no típicos, tales como Data Warehouse o Data Mart. Prácticamente todos los EIS obtienen sus datos de matrices multidimensionales denominadas "cubos" y las herramientas en las que se desarrollan estos sistemas tienen tecnología que permite realizar consultas amplias y complejas de diversas fuentes de datos en tiempos mínimos. Las partes importantes de un EIS son: la interfaz de usuario y la base de datos multidimensional, esto montado en una arquitectura Cliente/Servidor. Normalmente las pantallas o escenarios que se le presentan al usuario final poseen información que se obtiene vía remota del servidor en donde existe, de forma consolidada, proveniente de múltiples fuentes de datos. El Cliente es precisamente el usuario final, a quien se le debe presentar, de una forma amigable y sencilla, la mayor cantidad de información posible para que le apoye a visualizar sus variables de control del área específica que gestiona. Cada escenario en un EIS se adecua a los casos particulares de cada usuario, proveyendo la

información necesaria para tomar decisiones. En la mayoría de los casos, los EIS tienen en la gráfica su principal medio de representar información al usuario, pues es la forma más sencilla de tener comparativos entre variables. Esto y las tablas son una forma bastante sencilla, pero útil para otorgar información al usuario, pero esos componentes deben complementarse con otras utilerías y otros componentes que permitan no solo ver, sino también analizar lo que está pasando en la empresa con solo mover el ratón. En términos prácticos, un tomador de decisiones puede entender más el comportamiento de su empresa si ve escenarios en donde no sea necesario interactuar, pero que tenga lo que necesita para controlar su negocio y, solo en caso de ser necesario, interactuar de modo intuitivo (casi siempre con doble clic o arrastre) para obtener más detalle.

2.4.1.2 ¿Información para todos?

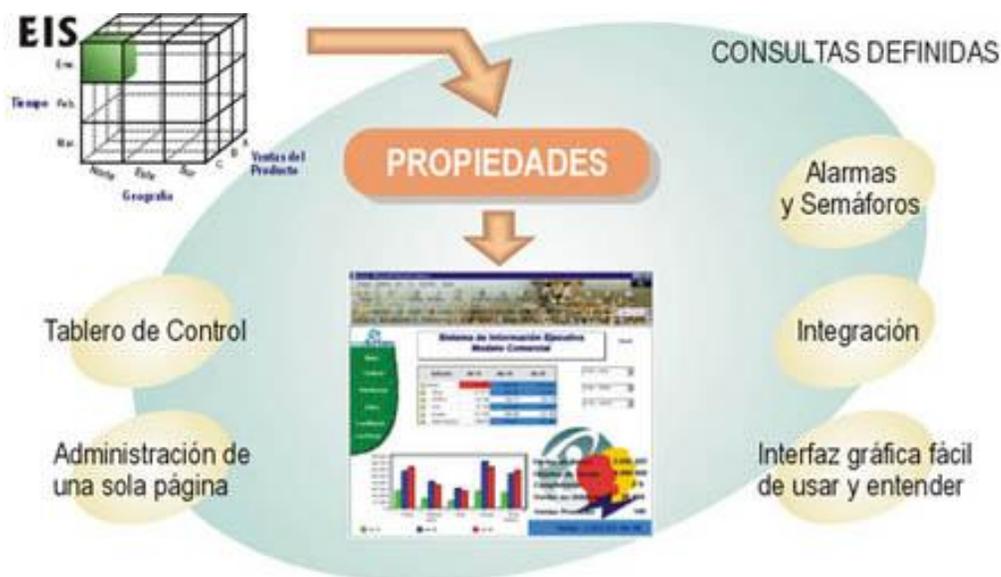
Las empresas se están dando cuenta que la solución de EIS se está convirtiendo en una solución de alcance empresarial. A pesar de que su origen es en los mandos directivos, su funcionalidad, sencillez y utilidad han llevado a que cualquier persona que necesite estar enterada de algún tipo de información para tomar una decisión, sea susceptible a ser usuario de un EIS. El enfoque original hacia el control o gestión de las empresas sigue siendo útil para algunos, pero no para todos, de ahí que la comunicación interna comience a sustituir al control.

Surge, sin embargo, un problema semántico, pues la definición original está orientada al uso ejecutivo, la verdad es que no existe tal problema, dependiendo de la interpretación que se le quiera dar al término 'Ejecutivo'. Si ejecutivo implica la gente que lleva el rumbo estratégico de la empresa, como son directores y gerentes, entonces no aplica una universalidad del término; si ejecutivo implica aquellas personas que toman una decisión, sin importar lo operativo o estratégico de su decisión, entonces si aplica. Lo importante no es definir ejecutivo, sino entender que poco a poco comienza a utilizarse en toda la empresa este tipo de soluciones. En función de eso, si el término ejecutivo genera diferencias, puede

modificarse y de hecho se está modificando hacia un nuevo significado derivado del original. EIS se puede leer como Sistema de Información Empresarial (Enterprise Information System) o Sistema de Información para todos (Everyone's Information Systems). Mucha gente aun no comprende como puede, por ejemplo, utilizar un supervisor o un jefe de algún departamento un EIS. Supongamos una empresa de mensajería y paquetería, en donde existe un operador que necesita constantemente monitorear el estatus de los paquetes y las unidades que los transportan, pues los clientes solicitan información para ver el estado de su envío. Puede hacerlo en una interfaz que le presente en un mapa las unidades de transporte, en la misma pantalla puede ver las rutas y horarios de las unidades, así como la cantidad de paquetes que transporta y sus embarques, y desembarques previos. Si el embarque no presenta ninguna clase de contratiempo y ningún cliente solicita el estatus de su envío, la pantalla lo único que hace es mantener informado al operario con información importante referente al proceso de embarque y que, en un momento determinado, puede apoyar una decisión. De la misma forma puede tener pantallas históricas sobre la eficiencia y eficacia de su flotilla, y su sistema de entregas, así como otras más relacionadas con los servicios brindados por su empresa. Todas estas pantallas, con diferentes formas de reflejar el estado actual y/o pasado de la empresa son EIS. Un jefe de producción, en otro ejemplo, puede tener una pantalla que le esté informando constantemente sobre sus líneas de producción, material, tiempos, trabajadores, etc., para que, si llega a detectar problemas en ese momento, pueda ejercer medidas correctivas; si no hay problemas, continua monitoreando. Sin importar tipos de información o puestos, las empresas se están haciendo mucho más planas y comienzan a sustituir el control por la autonomía, de ahí que los EIS estén siendo más universales y útiles en cualquier parte, pues brindan información útil de un modo sencillo para tener comunicación total.

2.4.1.3. Lo que debe contener un EIS

Si consideramos que una pantalla de información valiosa es un EIS, prácticamente cualquier hoja de cálculo e incluso un reporte de producción podría serlo. Para ser considerado un EIS, un sistema debe reunir una serie de características adicionales a las que por defecto debe contener al formar parte de la tecnología de Business Intelligence, es decir, brindar información y que sirva de apoyo a la toma de decisiones.



- **Interfaz gráfica fácil de usar y ver**

La heterogeneidad en cuanto a la formación académica que un usuario final posee, es enorme. Puestos claves pueden ser asumidos por economistas, administradores, contadores, informáticos, actuarios, ingenieros industriales, etc., el único elemento común es un mínimo de conocimientos necesarios para manejar una herramienta como Excel. Sin embargo, lo que no está bien definido es el límite superior de sus conocimientos. Un EIS está diseñado para un usuario final que no forzosamente domina herramientas computacionales complejas, e independientemente de que las domine, no debe invertir una parte importante de su tiempo en conocer y, posteriormente, utilizar herramientas complicadas. Bajo

esta filosofía es que se menciona como una característica de los EIS el que posean interfaces gráficas sencillas, que tengan una curva de aprendizaje corta y, además, deberán ser vistosas e intuitivas para facilitar la labor de monitoreo del tomador de decisiones. La idea es que se interactúe con el sistema el menor tiempo posible, de hecho, solo cuando alguna propiedad de una métrica indique que es necesario conocer el por qué de un valor.

- **Alarmas o semáforos**

Un típico EIS cuenta con funciones que le permiten al usuario notar rápidamente los errores y los valores destacables de la información. Es una de las características principales, ya que las propiedades que le son asignadas a los valores son las que permiten verdaderamente "monitorear" la información, de otro modo el ver datos planos provocaría que el usuario investigue cada uno de los valores y, posteriormente, los compare contra identificadores antes de poder determinar si son "buenos" o "malos". La tendencia es hacia interactuar cada vez menos con los sistemas y para ellos, las alarmas se vuelven indispensables, ya que disparan indicadores para que el ejecutivo solo ponga atención donde se han sobrepasado ciertos rangos de tolerancia. Estos indicadores pueden ir desde simples colores de letra y fondos, hasta el envío de correos electrónicos o mensajes por radiolocalizador.

- **Tableros de Control**

Surgiendo en Francia con el nombre de Tableau de Bord, el Tablero de Control es una herramienta que en un principio utilizaba indicadores financieros para permitir desarrollar diferentes procesos de negocio. Su especialización ha tomado el camino hacia el CMI (Cuadro de Mando Integral), que ya es una poderosa herramienta para dirección, que no solo utiliza indicadores financieros, sino también los no financieros para dirigir de forma proactiva a la empresa en la consecución de objetivos a mediano y largo plazo. El Tablero de Control no forzosamente implica una metodología de trabajo, en ese sentido es algo más

sencillo que puede llegar a un nivel de especialización tan amplio como uno lo quiera, por ejemplo, empleando la metodología del Balanced Scorecard para implementar un sistema de monitoreo del estado de salud corporativa. Un tablero de control es una utilidad en la cual el usuario puede monitorear con indicadores de cualquier tipo a su empresa o el área que tenga a su cargo. Diversos componentes le apoyan en esa actividad, pues cada indicador puede asociarse a un tipo de control específico. Con el tablero de control se pretende destacar los principales aspectos a monitorear, ya sea para el control o para la comunicación. El fin último es tener la información necesaria para apoyar una decisión en la menor cantidad de pantallas. Hay que considerar también que por la misma evolución de los EIS, la función de controlar con un tablero sigue siendo la más común, pero no es la única, en algunos casos se comienza a sustituir la palabra control por comunicación; no debe sorprendernos si entendemos el motivo.

- **Administración de una sola Página**

De la mano del tablero de control, la administración de una sola página significa colocar el mayor número posible de indicadores destacables de la empresa en el menor número posible de páginas, además, permitir la navegación hacia otras páginas con más información, ya que, definitivamente es prácticamente imposible colocar toda la información valiosa en un solo lugar.

- **Integración de información proveniente de los cubos**

El usuario tiene muchas veces la necesidad de tener información proveniente de múltiples sistemas o bases de datos. De hecho, para tomar una decisión estratégica es necesario poseer información de toda la empresa, no solo de una parte de ella. Un EIS debe permitir integrar información de cualquier aplicación y presentársela al usuario final de una forma transparente para él. Esto es la base para no moverse entre aplicaciones, al integrar la información de múltiples fuentes de información es posible la administración de una sola página en un tablero de control, el no hacerlo representa una limitante al usuario con la

consecuencia de moverse hacia otras pantallas e, incluso, hacia otras aplicaciones con información dispersa.

2.4.2 Sistema de Soporte de Decisiones (Decision Support System -DSS-)

2.4.2.1 ¿Qué es un DSS?

Como mencionamos anteriormente, los límites y el concepto de un DSS no han sido completamente precisados, a pesar de que la utilidad ha sido justificada en las organizaciones. Su uso indiscriminado con frecuencia lo lleva a rebasar límites de su aplicación y confundirse con términos como OLAP, Data Warehouse o EIS, lo cierto es que, independientemente, del término que llegue a usurpar, siempre se asocia al soporte a la toma de decisiones y, de alguna forma, todos los conceptos señalados tienen en la toma de decisiones el punto de encuentro.

"DSS se refiere a cualquier sistema de software que permite análisis de las diferentes variables del negocio para apoyar una decisión." **David Abdo**

Puede considerarse como un sistema que se basa en un almacén de datos y crea una base de datos multidimensional, permitiéndole al usuario procesar analíticamente la información en línea (OLAP), con la habilidad de poder girar y taladrar dentro de los datos. Como utilidad al usuario final, un DSS se valora cuando se profundiza en la información para conocer los "porqués" a los indicadores presentados, pero la infraestructura y metodologías que soporten el "taladro" de información son las que completan el esquema de un DSS y le permiten hacer uso de OLAP, Data Warehouse y otros conceptos asociados. En términos prácticos, el DSS lo vemos cuando analizamos la información, pero realmente involucra todo un proceso previo antes de poner la información en el escritorio del usuario. Los DSS están asociados a las jerarquías dentro de los conceptos de los cubos del Data Warehouse, por lo siguiente: En el momento que

desea conocerse el "por qué" de un dato visto en un EIS, el DSS debe permitir "profundizar" o "taladrar" la información, o sea, conocer el detalle de la misma y de las partes que la componen, aquí se echa mano de las jerarquías que fueron definidas en los cubos de información, para ver por cual de las rutas se profundiza hacia el detalle. Un DSS inicia cuando se analiza el detalle de una propiedad de la métrica. Los DSS se presentan como la opción viable para poder dar soporte y consistencia a la información que se diversificó enormemente. Con una metodología más estricta y más herramientas a su alcance, los DSS comenzaron a ser la contraparte de los sistemas OLTP. Más que una definición, los DSS pueden entenderse a partir de sus características de funcionalidad en una empresa al ser comparadas con los sistemas OLTP que apoyan la transacción diaria. En un DSS la consistencia se mide globalmente. A pesar de que no importan las transacciones individuales, son precisamente estas las que, en suma, conforman una gran transacción completa y consistente. De hecho, antes de poder formar una gran transacción, a partir de los sistemas operacionales, debe verificarse la calidad de los datos para garantizar la información correcta y coherente. Un sistema DSS procesa una transacción por día (puede ser por cualquier periodo de tiempo, pero el común es por día), pero esa transacción es producto de miles o millones de registros que han sido procesados en el día. En vez de llamarle transacción, se le llama carga de información de producción. En este caso lo importante es el estado de consistencia del sistema antes de la carga y el estado de consistencia una vez que se ha terminado de efectuar la carga. En cuanto a los usuarios y administradores de un DSS, no son los que ingresan cada una de las operaciones en sus PCs, sino los que solicitan una o dos hojas con resúmenes totalizados de miles de operaciones. El usuario de OLTP cambia con frecuencia el tipo de información que solicita, de ahí que sus requerimientos no sean planeados, sino heurísticos. Una pantalla de un sistema DSS deberá contener la información sobre el por qué de una tabla o gráfica, con algunos números y pantallas, al momento de pedir información sobre el por qué de determinado dato, podrá accederse a otra pantalla con la explicación y así, sucesivamente, hasta llegar al último detalle. Una característica importante de los

DSS es que la demanda puede ser muy variable, por lo que es normal implementarlo en una máquina distinta de OLTP. Al estar en una máquina distinta se convierte en un servidor que puede ser accedido por diversos usuarios, pero también puede tener información integrada de múltiples sistemas remotos de OLTP. El tiempo es un factor importante de los DSS visto como dimensión. La inconsistencia temporal se debe evitar en un DSS. Mientras que un OLTP es complicado para explicar la historia, un DSS lo que hace es tomar fotografías instantáneas de la empresa en un momento determinado de la historia, al sumar esa serie de fotografías se conforman capas que pueden explicar como era la empresa en determinado periodo de la historia. Moviendo fotografías estáticas hacia el DSS cada cierto tiempo programado, se resuelven dos problemas de representación de tiempo que se tiene en los sistemas OLTP. Primero, el DSS permanece quieto durante el día cuando los usuarios están lanzando queries. Después, almacenando cuidadosamente la información de cada instantánea del DSS, se puede representar cualquier momento en el tiempo correctamente. Las fotografías instantáneas del sistema OLTP se llaman 'extracción de la información de producción', mientras que el envío hacia el DSS se llama 'migración de la información de producción'. Esta migración se realiza en el mismo momento cada día. El ciclo de vida de un DSS es una situación muy particular. Mientras que un tradicional ciclo de vida de un sistema se deben entender los requerimientos, analizar las etapas del diseño y después desarrollar; en un DSS el ciclo de vida es inverso, es decir, inicia con la información de los sistemas OLTP, una vez que se tiene, es integrada y se prueba para saber cual es real y útil, y cual no. A continuación se escriben los programas para manejar esta información y, por último, los programas se analizan, y finalmente se entienden los requerimientos del sistema. El ciclo de vida de un DSS se orienta a la información, mientras que el de OLTP se orienta a los requerimientos. La interpretación y manipulación de la información es muy distinta a los formatos típicos. A pesar de que la capacidad de extracción de datos puede ser de archivos de texto, tablas, etc., la forma de organizar esa información es lo que difiere, pues la mayoría de los DSS organizan la información vía matrices multidimensionales denominadas "cubos". Los cubos

organizan la información de tal modo que puedan, posteriormente, 'montarse' herramientas para desarrollar sistemas complejos, que permitan realizar una gran cantidad de cálculos, consolidaciones, consultas y despliegues de información, localizadas en múltiples repositorios en un tiempo mínimo. La explotación de la información, a diferencia de OLTP en donde los reportes o las listas sencillas son el medio más común para ver los datos, se realiza al penetrar hacia el detalle mismo de los datos más genéricos. El poder verdaderamente tener información relevante y resumida para tomar una decisión es el equivalente a analizar todos los reportes que la actividad diaria genera, es decir, una gran cantidad de papel.

2.4.2.2 ¿Qué debe contener un DSS?

Análisis Multidimensional (OLAP) El análisis multidimensional no es privativo de arquitecturas multidimensionales, puede también llevarse a cabo en arquitectura relacional, diseñada para tal caso, lo importante para poder hacer análisis multidimensional no son las bases de datos, sino la estructura de la base de datos y las técnicas que se utilicen para su explotación. Las aplicaciones OLAP soportan ese tipo de análisis ya que dos de sus principales características son que permite el análisis y la multidimensionalidad. Dentro de los niveles estratégicos organizacionales, la información se concibe como una serie de hechos multidimensionales, jerárquicos y relacionados; como ejemplo, los datos de inventarios, ventas y compras están interrelacionados y dependen entre sí. La idea del análisis multidimensional es facilitar la consulta y análisis al usuario al presentar una visión muy sencilla de los datos, muy similar a la forma como él ve la organización. La información puede ser accesada desde diferentes variables organizacionales y pivotar la perspectiva. Los almacenes multidimensionales guardan de modo lógico sus datos en arreglos utilizando el concepto conocido como "cubo". En éste, cada una de las caras almacena una dimensión, pudiendo cruzar diferente información en una sola arista de hasta 3 variables. Físicamente, en la base de datos, una celda de información puede almacenar hasta 3

dimensiones y, con mucha facilidad, pivotar el cubo, es decir, cambiar la consulta a otras celdas para analizar más a detalle.

Cuando un esquema similar a la realidad se guarda en medios de almacenamiento y aparte se complementa con herramientas diseñadas para obtener específicamente ese tipo de información, el resultado es poder entender a la empresa a través de la tecnología de la misma forma que lo haría en la actualidad. El análisis multidimensional brinda esa posibilidad al usuario y es la principal característica que debe poseer un DSS.

El análisis multidimensional debe permitir:

- Soportar requerimientos complejos de análisis
- Analizar datos desde diferentes perspectivas
- Soportar análisis complejos contra volúmenes de datos ingentes

Proyecciones de Información. Las proyecciones de negocio ofrecen al usuario un pronóstico de lo que puede ocurrir en el futuro, basándose en análisis estadístico y de regresión.

Tendencias. Utiliza la información presente y pasada para evaluar el comportamiento de determinada variable en el tiempo.

What... If. Análisis prospectivo de un indicador al modificarse una o más variables que inciden en su comportamiento.

2.4.3 Reporteo

Tradicionalmente los reportes han sido el medio principal para tener información en las manos. Tanto reportes en papel como en el escritorio, el usuario depende de ellos para comunicar a la gente lo que está ocurriendo en la

empresa vía los almacenes de información. Los reportes tienen en las bases de datos su principal fuente de alimentación y han brindado al usuario final -entendido éste como cualquier persona que requiera un reporte-, la posibilidad de consultar y publicar lo que las bases de datos poseen, la limitante que siempre ha existido en este sentido es que el generar un reporte implica manejar algunas habilidades técnicas relacionadas con las bases de datos y las herramientas de software.

Los reportes o las consultas son requeridos por cualquier persona dentro de la organización, de hecho, no es una labor que esté limitada a un nivel jerárquico o puesto. Cuando un usuario con poca preparación técnica (secretarias, ejecutivos, etc.) requiere de un reporte o consulta, lo hace vía terceros, el tradicional entorno es hacer la solicitud de información a sistemas y obtenerla después de cierto tiempo, el tiempo puede ser irrelevante o puede significar la pérdida de vigencia de la información reportada.

El usuario técnico realiza los reportes y para ello necesita conectarse a la base de datos, posteriormente, diseñar el formato requerido y, al final, obtener los datos. Tanto la conexión como el formato son tarea de sistemas, la obtención de datos se comparte con el usuario final, con esta lógica se entiende que el usuario final no puede manipular sus reportes en caso de necesitar modificaciones o adecuaciones al mismo, por lo tanto continua la dependencia, incluso, para modificaciones mínimas. Los usuarios con algunas habilidades computacionales pueden hacer uso de los reportadores orientados a los usuarios de sistemas, pero solo para obtener reportes sencillos, no el tradicional reporte complejo que es el más útil, además, la gente que toma decisiones no debe perder su tiempo en procesar la información para después analizarla.

Algunas de las limitantes propias de la generación de reportes en la actualidad son:

- **Integración.** Los reportes complejos que necesitan de más de una fuente de información se realizan extrayendo los datos en un primer momento y,

posteriormente, procesándolos en una herramienta que los integre (hojas de cálculo, archivos de texto, access, etc.). Esto implica preparar la información antes de que se encuentre lista para ser reportada y la preparación normalmente llevan horas o hasta días con la consecuente pérdida de tiempo en funciones secundarias improductivas.

- **Distribución.** Los reportes se imprimen y posteriormente se distribuyen de forma manual o mediante fax. También se generan en pantalla y, luego, se envían mediante fax o correo electrónico, con lo que hay dos actividades involucradas que generan pérdida de tiempo en su ejecución: la publicación y la distribución.

- **Dependencias de sistemas.** Las bases de datos se eficientizan para almacenar datos no para consultarlos. Los lenguajes de consulta se adaptan a la estructura de las bases de datos. En función del conocimiento necesario, tanto de las bases de datos a consultar como de los lenguajes técnicos necesarios, el usuario final no está en posibilidad real de obtener su propia información, por tal motivo la dependencia de gente de sistemas es muy grande y provoca el que se tengan los reportes cuando la gente de sistemas dispone de tiempo y recursos para hacerlo, visto de otro modo, si la gente de sistemas no lo hace, el usuario final no los tendría a la mano.

- **Seguridad.** Al dejar en manos de terceros la generación de reportes se da pauta a que la información valiosa para la empresa pueda ser filtrada y utilizada por personas que no son los directamente interesados en utilizarla. Un reporte de sueldos y prestaciones de los empleados, solicitado por Recursos Humanos, puede ser un detonante de conflictos en manos de inconformes.

Dependiendo del nivel del usuario puede disminuirse o eliminarse la dependencia de sistemas para construir los reportes.

2.4.3.1. Herramientas de consulta y reporte.

Las herramientas de consulta y reporte (Query & Reporting Tools) son una categoría de herramientas de BI. Con las herramientas de reporte orientadas al usuario final se pretende eliminar la dependencia del área de sistemas al manejar una capa intermedia entre complejidad técnica y usuario final.

Las herramientas de reporte orientadas al usuario final son software que aísla la compleja capa técnica propia del lenguaje de sistemas, tal como lenguaje SQL, uniones de tablas y nombres críticos, al organizar los datos de la terminología de negocios. El resultado es que el usuario final o intermedio tiene una vista mucho más parecida a su concepción del negocio, o al menos lo suficiente como para poder generar sus propios reportes y publicación de los mismos, sin depender de los usuarios técnicos.

El usuario de sistemas continúa participando, pero básicamente en el control sobre el acceso a bases de datos, administración, mantenimiento, seguridad, impacto en la red e incluso, creación única de reportes complejos; el usuario final tiene solo un ambiente amigable en el que es posible crear sus propias consultas y reportes.

Una herramienta de reporte puede también publicar los datos que se encuentran almacenados en un Data Warehouse, pues físicamente el Data Warehouse es también un almacén de datos. Desde ese punto de vista, se brinda con la tecnología, la posibilidad de que el usuario final no solo consulte y publique la información a detalle, sino información concentrada y agrupada; de ahí que también sea considerada como el soporte último para el máximo nivel de detalle de un DSS.

Cierta duda ha surgido en cuanto a la verdadera posibilidad para que el usuario final realice sus reportes y posteriormente los explote. Definitivamente depende del nivel en el cual ubiquemos al usuario final y, evidentemente, del nivel de conocimientos en materia de computación que pueda tener. Un usuario final puede ser cualquier persona que requiera hacer una consulta o reporte y éste puede ir desde una secretaria, un asistente, un gerente o un director, cualquiera que necesite información procesada de fuentes de datos es un usuario potencial. Sin embargo, a pesar de que el manejo de una herramienta como excel se ha considerado un estándar que pueda utilizar cualquier administrativo de cualquier nivel, no todas las personas pueden decir que verdaderamente saben utilizar, para fines prácticos, esa herramienta. Precisamente, para ese tipo de usuarios, las herramientas de reporte no serán útiles más que para explotar los reportes que previamente se hayan realizado, es decir, ejecutar reportes. Otro tipo de usuario, con algunos conocimientos, si podrá construir sus propios reportes, que siguen siendo tarea de gente de sistemas, pues requieren de un procesamiento de la información antes de que sea "maquillada" y, para ello, son necesarios procedimientos externos a los datos con lenguajes de consulta altamente técnicos. Dependiendo del nivel del usuario puede disminuirse o eliminarse la dependencia de sistemas y pueden las herramientas de reporte mantener 3 niveles de dificultad:

- Usuarios poco expertos (gerentes tácticos, directores ejecutivos) quienes solicitan la ejecución de reportes o consultas predefinidas, según parámetros predeterminados.
- Usuarios con cierta experiencia (asistentes, gerentes operativos, secretarias especializadas) pueden generar consultas o reportes flexibles, apoyándose en una interfaz gráfica intuitiva.
- Usuarios muy experimentados (sistemas) pueden crear e incluso escribir, total o parcialmente, la consulta en un lenguaje de consulta.

2.4.3.2. Características de una herramienta de reporte.

Una herramienta de reporte orientada al usuario final debe también poseer algunas utilidades adicionales que faciliten la generación y publicación de reportes.

- **Intuitivo.** Como cualquier herramienta de BI, la característica común es su facilidad de uso e intuición. Con apoyo en interfaces gráficas y visuales, un usuario con una formación estándar podrá hacer uso de una herramienta de este tipo.
- **Seguridad.** Deben brindar seguridad para el acceso a los reportes, tanto a nivel usuario como por grupos e, incluso, en el grado de profundidad de cada usuario a la información. Esto con la idea de que la información privada no sea accesible por cualquier persona.
- **Publicación y distribución.** Una función importante de las herramientas es eliminar la doble actividad de ejecutar y publicar primero un reporte y, posteriormente, distribuirlo a quien lo necesita. La publicación de una consulta normalmente se realiza mediante fax, correo electrónico, archivo e, incluso, de mano en mano. El objetivo con las nuevas herramientas es facilitar estas actividades. En el momento que un usuario pueda ejecutar, desde cualquier lugar, una consulta y mandarla imprimir in-site, se está eliminando la doble función de publicación y distribución. Las herramientas de reporte deben permitir realizar consultas desde su PC en modo Cliente / Servidor, Intranet o Internet e imprimir cualquier consulta.
- **Navegación.** La interrelación de reportes es también frecuente para generar sistemas basados en reportes y reportes auxiliares. La navegación entre ellos mediante ligas e hipervínculos es una funcionalidad más.

- **Programación automática.** Generación de instrucciones para que los reportes se ejecuten automáticamente e incluso se distribuyan mediante correo electrónico.
- **Reportes dinámicos.** Permitir el ingreso de parámetros de valor que hagan un reporte flexible y dinámico en el momento de su ejecución. Con esto se economizan esfuerzos al ejecutar muchas consultas a partir de un solo reporte.
- **Reportes por excepción.** Para no consultar una cantidad muy grande de información debe existir la posibilidad de brindar reportes por excepción, es decir, lo que no cumplan con las condiciones de generalidad.

2.4.4. Minería de Datos.

Es la explotación de grandes volúmenes de datos con el fin de extraer información valiosa para el negocio y previamente conocida. El Data Mining es un término usado para describir el conocimiento descubierto en bases de datos, la extracción del conocimiento, la arqueología de los datos, la exploración de los datos, el proceso de modelo de los datos, la información generada y el software.



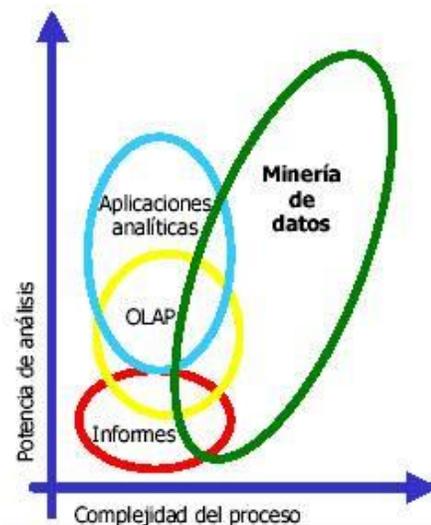
- | | | |
|-----------------------------|-----------------------------------|---------------------------------------|
| ❖ Transacciones | ❖ Segmentación de Clientes | ❖ Tratamiento diferenciado a clientes |
| ❖ Sociodemográficos | ❖ Afinidad entre productos | ❖ Venta cruzada de productos |
| ❖ Respuesta a campañas | ❖ Patrones secuenciales de compra | ❖ Efectividad de promociones |
| ❖ Compra de productos | ❖ Modelos de predicción | ❖ Incremento de cuota de mercado |
| ❖ Hábitos de comportamiento | ❖ Rendimiento de campañas | ❖ Retención de Clientes |
| ❖ Clasificaciones externas | | ❖ Recuperación de Clientes |

2.4.4.1. Qué es y qué no es la minería de datos.

La minería de datos puede definirse como la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de los datos. Para conseguirlo hace uso de diferentes tecnologías que resuelven problemas típicos de agrupamiento automático, clasificación, asociación de atributos y detección de patrones secuenciales. La minería de datos es, en principio, una fase dentro de un proceso global denominado descubrimiento de conocimiento en bases de datos (Knowledge Discovery in Databases o KDD), aunque finalmente haya adquirido el significado de todo el proceso en lugar de la fase de extracción de conocimiento.

Es habitual que los expertos en estadística confundan la minería de datos con un análisis estadístico de éstos (afirmaciones de este tipo pueden encontrarse en documentación de empresas dedicadas al procesamiento estadístico que venden sus productos como herramientas de minería de datos).

En el caso de la minería de datos el proceso es muy distinto: la consulta que se realiza a la base de datos (al Data Warehouse) busca relaciones entre parejas de productos que son adquiridos por una misma persona en una misma compra. De esa información, el sistema deduce, junto a otras muchas, la afirmación anterior. Como podemos ver, en este proceso se realiza un acto de descubrimiento de conocimiento real, puesto que no es necesario ni siquiera sospechar la existencia de una relación entre estos dos productos para encontrarla.



Capítulo 3

Fase de Preparación

Durante el Capítulo 2 se ha visto la importancia de la implementación de un sistema BI en una organización entendiéndose el valor tan grande que actualmente se le da a la información. También se mencionó que las condiciones actuales de los mercados han provocado, precisamente, el que sea necesaria tecnología cada vez más avanzada para responder a las peticiones muy particulares de información y que difícilmente un solo sistema de información cubrirá todas estas necesidades de la pirámide organizacional. Esto dio pauta para entrar en la definición de un sistema BI y que técnicas de análisis se pueden utilizar bajo esta metodología.

Entendiendo esto es momento de entrar en la fase de Preparación.

3.1 Definición del Equipo de Trabajo.

Del equipo de Trabajo depende el éxito del proyecto. De ahí la importancia en su definición. Para que un grupo trabaje verdaderamente en equipo es muy importante que sus miembros desempeñen ciertos roles conforme lo requieren las circunstancias. No se pretende que cualquier miembro sea capaz de ejercer todos los roles positivos. Esto es muy difícil, si no imposible. Se trata de que entre todos los miembros se logre un adecuado ejercicio de los roles. Unos miembros serán más aptos para ciertos roles, y otros lo serán para otros roles. Esto tiene que ver con los estilos personales.

En esta sección analizaremos los roles necesarios para el logro del trabajo en equipo en línea con los estilos personales de quienes desempeñan esos roles.

3.1.1 Preferencias de trabajo

Fundándose en Carl Jung y Myers-Briggs, Charles Margerison y Dick McCann han elaborado un modelo que considera la actuación de la persona en el trabajo. Dichos autores opinan que el estilo de la persona en el trabajo no es el mismo que su estilo fuera del trabajo. La estructura de este modelo tiene mucha analogía con su antecedente. Establece cuatro dimensiones de preferencias -y para cada una indica una alternativa en cuanto a lo que se llama “habilidad funcional”- que responden a cuatro cuestiones clave, a saber:

- ¿Cómo prefieren las personas relacionarse con otras?
- ¿Cómo prefieren las personas recolectar y utilizar la información?
- ¿Cómo prefieren las personas tomar decisiones?
- ¿Cómo prefieren las personas organizarse y organizar a otras?

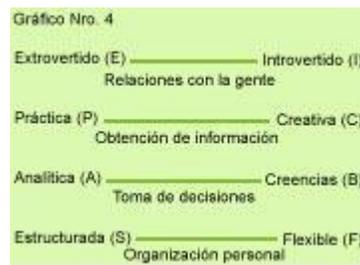
La primera pregunta es enteramente de relaciones entre las personas. Algunas son comunicativas y sociables en sus relaciones con otros en una manera extrovertida (E) mientras que otras son más calladas y no necesitan estar con otras personas, es decir, introvertidas (I).

La segunda pregunta se relaciona con la administración de información. Cuando las personas se reúnen y utilizan la información frecuentemente tienen una tendencia a ser prácticos (P) o creativos (C). Las personas prácticas prefieren trabajar con ideas testeadas y prestar atención a los hechos y el detalle, mientras que las personas creativas frecuentemente están desafiando el statu quo y se presentan con nuevas ideas.

La tercera área se relaciona con la toma de decisiones. Cuando se toman decisiones normalmente se hace en una manera analítica (A) o de acuerdo con las creencias de un individuo (B -por belief). Las personas analíticas erigirán criterios de decisión objetivos y elegirán aquella decisión que maximice el rédito, mientras

que las personas orientadas hacia las creencias tenderán a tomar decisiones que concuerden con sus propios principios y valores personales.

Finalmente, la cuarta pregunta se relaciona con las cuestiones organizacionales. Los administradores parecen organizarse y organizar a otros en una manera estructurada (S -por structured) y flexible (F). Las personas estructuradas están orientadas hacia la operación y quieren concluir o resolver cuestiones, mientras que a las personas flexibles les gusta pasar el tiempo diagnosticando la situación y tenderán a posponer la “conclusión” y “resolución” hasta que hayan recolectado toda la información que puedan.



3.1.2. Tipos y preferencias de trabajo

La Rueda de Administración en Equipo, desarrollada por Charles Margerison y Dick McCann, proporciona un gráfico integrado de las preferencias de trabajo de las personas y las relaciona con las funciones clave que son necesarias en un equipo eficaz.

La Rueda posee ocho sectores y un centro que indican las funciones y preferencias de trabajo, y se incluye una descripción de conducta (ej.: “implementador”) así como la función de trabajo denominada “organizador”. Cada área de función, por lo tanto, posee una descripción de dos palabras como creador-innovador, cumplidor-productor, etc.

En torno de la Rueda se encuentran principales actividades como:

- Exploración
- Organización
- Control
- Asesoramiento

En resumen, las personas pueden preferir un trabajo que posee un fuerte elemento de control o uno donde existe un alto grado de exploración y búsqueda de nuevas oportunidades.



a. Creadores-innovadores: son personas que poseen un número de ideas que pueden desafiar y perturbar la manera existente de hacer las cosas. Dichas personas pueden ser muy independientes y desean experimentar y practicar sus ideas sin considerar los sistemas y métodos actuales.

b. Exploradores-motivadores: usualmente son excelentes tanto en la generación de ideas como en lograr que las personas se entusiasmen con ellas. Saldrán a investigar qué está ocurriendo fuera de la organización y compararán las nuevas ideas con lo que otras personas están realizando. Son también buenos recordando contactos, información y recursos que pueden ayudar a la innovación

en el equipo. Son muy capaces de impulsar una idea aun si no siempre son las mejores personas para organizarla y controlarla.

Usualmente son influyentes, hablan fácilmente, aun sobre temas donde no son expertos, y disfrutan buscar nuevas oportunidades y desafíos. Pueden aburrirse fácilmente y por lo tanto les gusta el desafío de trasladarse de un proyecto a otro. Son particularmente buenos en tomar ideas de otros (creadores-innovadores, por ejemplo) y promocionarlas entusiastamente a otros.

c. Evaluadores-desarrolladores: proporcionan un equilibrio entre las partes de “exploración” y “organización” de la rueda. Frecuentemente buscan maneras y medios de hacer que una idea funcione en la práctica. Su preocupación es ver si el mercado quiere la innovación y por lo tanto la testearán contra algún criterio práctico.

d. Implementadores-organizadores: son las personas que lograrán que se hagan las cosas. Una vez que han sido convencidos de que la “idea” es de interés establecerán procedimientos y sistemas y convertirán la idea en una realidad en funcionamiento. Organizarán personas y sistemas para asegurar que pueden alcanzarse límites.

Ponen énfasis en lograr que las cosas se hagan aun si esto implica que en el camino ciertas “plumas se ericen”. Fijan objetivos, establecen planes, calculan quién debe hacer qué y luego ejercen mucha presión para obtener resultados. Pueden por momentos ser impacientes cuando las personas o cosas se interponen en su camino, pero siendo implementadores-organizadores hallarán obstáculos. Están orientados hacia la tarea, les gusta trabajar con plazos determinados y buscan controlar su mundo más que dejar que éste los controle.

e. Cumplidores-productores: se enorgullecen de fabricar un producto o servicio a un nivel. Harán esto en una base regular y se sentirán satisfechos si pueden entregar “lo que se espera cuando se espera”. Verdaderamente, les gusta trabajar para fijar procedimientos y hacer cosas en una manera regular.

f. Controladores-inspectores: son personas que disfrutan de hacer el trabajo detallado y de asegurarse que los hechos y cifras sean correctos. Serán meticulosos y cuidadosos y frecuentemente críticos de errores o de trabajo poco sistemático.

g. Seguidores-mantenedores: usualmente son personas de fuerte convicción sobre la manera en que se deberían hacer las cosas. Con frecuencia son las personas más sustentadoras de otras en el equipo y pueden proporcionar mucha estabilidad.

h. Informantes-asesores: son buenos generando información y recolectándola de tal manera que pueda comprenderse. Muchas personas usualmente son pacientes, y están preparadas para contenerse de tomar decisiones hasta que sepan cuanto puedan sobre el trabajo a efectuar.

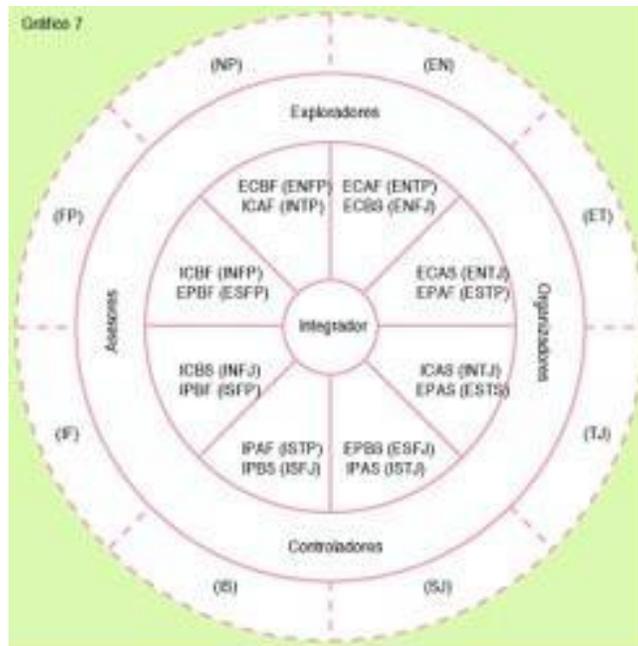
El análisis de la Rueda contribuye a clarificar que, cuando los miembros de un equipo comprenden las diferentes preferencias de trabajo o “marcos de referencia” que las personas poseen, entonces puede resultar un excelente trabajo en equipo.

Por último, vemos que la “coordinación” está en el centro de la Rueda y es una aptitud que todos los administradores pueden aprender. Es importante mencionar que la “coordinación” no es una preferencia sino una aptitud. Por lo tanto, un administrador tendrá preferencias por uno o más de los sectores externos de la Rueda, pero si también puede sostener el espacio del centro como un “integrador” estará en el camino hacia un equipo de alto rendimiento.

Esto sugiere que el administrador actual necesita tener no sólo habilidades técnicas, sino también habilidades interpersonales, particularmente las de

comunicación. Los ejecutivos exitosos del mañana serán aquellos que hayan dominado las habilidades de la coordinación administrativa.

Parece de suma utilidad la relación que Margerison-McCann establecen entre los ocho sectores de tipos de trabajo/conducta con las preferencias enunciadas.



En el gráfico N° 7 se aprecia esta relación; adicionalmente se agregaron entre paréntesis los estilos personales que corresponderían al modelo de Myers-Briggs.

Así, por ejemplo, podemos tomar el sector “cumplidor-productor”. Recordemos que las personas que tienen esta preferencia se enorgullecen de fabricar un producto o servicio a un nivel. Harán esto en una base regular y se sentirán satisfechos si pueden entregar “lo que se espera cuando se espera”. Verdaderamente, les gusta trabajar para fijar procedimientos y hacer cosas en una manera regular. Por ello, según el modelo de McCann-Margerison estas personas responderán a la habilidad funcional: ICAS (introvertida-creativa-analítica-

estructurada) o EPAS (extrovertida-práctica-analítica-estructurada). Para el modelo de Myers-Briggs el estilo personal será: INTJ (introverso-intuitivo-razonador-juzgador) o ESTJ (extroverso-sensorial-razonador-juzgador).

En base a esto deberás crear un grupo de trabajo multidisciplinario que considere los siguientes roles:

- Administrador del Proyecto.
- Especialistas del Negocio.
- Arquitecto del Data Warehouse.
- Modelador de Datos.
- Diseñador de Base de Datos.
- Programador de la Aplicación.
- Entrenadores de Herramienta.
- Clientes.

3.2. Identifique a los Beneficiarios.

¿Quién puede ganar más con la implementación del BI? Quién está ya trabajando con base de datos o emplea metodologías estadísticas.

Dentro de la organización tenemos personal clave que se vería beneficiado directamente con la implementación de un BI. Estas personas son aquellas que actualmente ya están trabajando con base de datos para generar sus propios indicadores tales como % de desperdicio, cumplimiento a clientes, estadísticas de ventas, niveles de inventario, etc.

Imagina un caso práctico en donde un gerente de ventas desea analizar la estadística de ventas de 24 meses, 10 vendedores, 10 diferentes productos, de sus 100 clientes mostrando 5 indicadores. Tan solo esta sencilla consulta que para

muchas compañías son del día con día nos genera 1,200,000 combinaciones diferentes de poderla ejecutar. Para poder generar esto en reportes tradicionales se necesitarían varios cientos de páginas de información que serían imposible de analizar. Este es un caso donde necesitas BI.



Así como el caso del gerente de ventas se deberán de identificar aquellas personas que se verían beneficiados con la implementación de un BI en la organización.

3.3. ¿Quién necesita soluciones de Business Intelligence?

Si usted puede contestar afirmativamente por lo menos a una de las siguientes preguntas, entonces usted es candidato a beneficiarse de las soluciones de BI.

- ¿Pasa más tiempo recolectando y preparando información que analizándola?
- ¿En ocasiones le frustra el no poder encontrar información que usted está seguro que existe dentro de la empresa?

- c) ¿Pasa mucho tiempo tratando de hacer que los reportes en excel luzcan bien?
- d) ¿Quisiera tener una guía sobre las cosas que han sucedido cuando los administradores anteriores implementaban determinada estrategia?
- e) ¿No sabe qué hacer con tanta información que tiene disponible en la empresa?
- f) ¿Quiere saber qué productos fueron los más rentables durante un periodo determinado?
- g) ¿No sabe cuáles son los patrones de compra de sus clientes dependiendo de las zonas?
- h) ¿Ha perdido oportunidades de negocio por recibir información retrasada?
- i) ¿Trabaja horas extras el fin de mes para procesar documentos o reportes?
- j) ¿No sabe con certeza si su gente está alcanzando los objetivos planeados?
- k) ¿No sabe si mantiene una comunicación estrecha entre las diversas áreas de su empresa hacia una estrategia común?
- l) ¿No tiene idea de por qué sus clientes le regresan mercancía?

3.4. Defina Metas.

Las decisiones en BI deben estar basadas en medidas cuantificables.

“Si tú lo puedes medir, tú lo puedes administrar”, Kaplan & Norton.

3.4.1. ¿Qué puede hacer Business Intelligence?

Con BI se puede:

- Generar reportes globales o por secciones
- Crear una base de datos de clientes
- Crear escenarios con respecto a una decisión
- Hacer pronósticos de ventas y devoluciones
- Compartir información entre departamentos
- Análisis multidimensionales
- Generar y procesar datos
- Cambiar la estructura de toma de decisiones
- Mejorar el servicio al cliente

Según Kobana Abukari y Vigía Job [2], “BI es una de las iniciativas administrativas más robustas que los administradores inteligentes pueden emplear para ayudar a sus organizaciones a crear más valor para los accionistas”.

BI ha tenido mucho éxito ya que le da una ventaja a las empresas sobre sus competidores al juntar a las personas y a la tecnología para resolver problemas. La siguiente es una lista de las áreas más comunes en las que las soluciones de inteligencia de negocios son utilizadas:

- **Ventas:** Análisis de ventas; detección de clientes importantes; análisis de productos, líneas, mercados; pronósticos y proyecciones.

- **Marketing:** Segmentación y análisis de clientes; Seguimiento a nuevos productos.
- **Finanzas:** Análisis de gastos; rotación de cartera; razones financieras.
- **Manufactura:** Productividad en líneas; análisis de desperdicios; análisis de calidad; rotación de inventarios y partes críticas.
- **Embarques:** Seguimiento de embarques; motivos por los cuales se pierden pedidos.



Hoy en día conocer a su organización de arriba abajo es crítico para la supervisión y el crecimiento. Para asegurar que su compañía está alcanzando sus metas corporativas, usted tiene que mantener al día en el rendimiento de áreas claves a lo largo de toda la empresa, desde ventas hasta finanzas y a lo largo de la cadena de valor del negocio.

Es importante durante este proceso identificar los indicadores de proceso clave (**KPIs**) a lo largo de la empresa para identificar las causas y los efectos de los éxitos y de los problemas. Para que estas mediciones estratégicas sean

eficaces, la información de funciones individuales de la empresa debe unificarse en una visión completa de la compañía y esta visión debe ser compartida con la mayor audiencia posible.

Se debe atacar primero el área de mayor urgencia en su empresa desde el punto de vista de negocio y luego extender una solución integral a lo largo de empresa, permitiendo administrar de una forma inmediata y confiable el rendimiento a ambos niveles al de la unidad de negocios y al nivel empresarial, transformando sus datos operacionales en información consistente y confiable optimizada a través de reportes con una visión multidimensional en la empresa.

Señal #1: La información “El punto Dulce”
Alguna información es mas importante que otra, esta es el “Punto dulce”. La mas valiosa información para la organización esta en un número pequeño de información “Punto Dulce” y están ligadas con *actividades de administración clave*.

3.5. ¿Construir o Comprar?

¿Construir o Comprar?. Es una gran decisión la cual dependerá de los requerimientos y recursos de la organización. Sin duda comprar reduce grandemente el tiempo de implementación de un sistema de BI y se aprovechan los años de experiencia con los líderes en el mercado de proveer soluciones de inteligencia de negocios empresarial completamente integradas pero implica en mayores costos.



3.6. Software de Business Intelligence

Estas son las soluciones de BI más reconocidas actualmente en el mercado en el caso de que usted haya decidido comprar una aplicación BI.

SAGENT SOLUTION PLATTFORM:



Este sistema integrado extrae, transforma, mueve, distribuye y presenta la información clave para la toma de decisiones en la empresa en un entorno homogéneo.

MICROSTRATEGY:



Provee soluciones a clientes de cualquier industria y/o área funcional con el fin de ayudarlos en la obtención de un mayor conocimiento sobre la información manejada en su empresa.

BUSINESS OBJECTS:



Suministra a los usuarios el poder acceder de forma sencilla a los datos, analizar la información almacenada y creación de informes.

COGNOS:



Es un software que ofrece la funcionalidad de análisis y toma de decisiones. Cuenta con una herramienta especial para modelación, pronóstico – forecasting -, y simulación - what-if - del negocio.

BITAM/ARTUS BUSINESS INTELLIGENCE SUITE:



Herramienta capaz de agrupar la información y utilizarla como un activo que ayudará a la empresa a identificar las oportunidades de negocio, optimizar las áreas de finanzas, clientes, procesos internos, aprendizaje e innovación.

ORACLE9I APPLICATION SERVER:



Permite acceder, analizar y compartir la información y tomar decisiones precisas, basadas en datos en forma rápida.

3.6.1 Cuadrante mágico de posicionamiento.

Un cuadrante mágico es una herramienta analítica creada y promovida por la empresa Gartner y la cual muestra una representación gráfica del mercado compartido en un determinado periodo de tiempo. Los cuadrantes mágicos de Gartner proporcionan a las empresas un medio para identificar y diferenciar a los proveedores de servicios del sector de las tecnologías de la información.

Según define Gartner, los líderes en los cuadrantes mágicos son aquellos fabricantes de software que operan bien hoy día, tienen una visión clara de la dirección del mercado y desarrollan activamente las competencias necesarias para mantener su posición de líderes en el mercado.

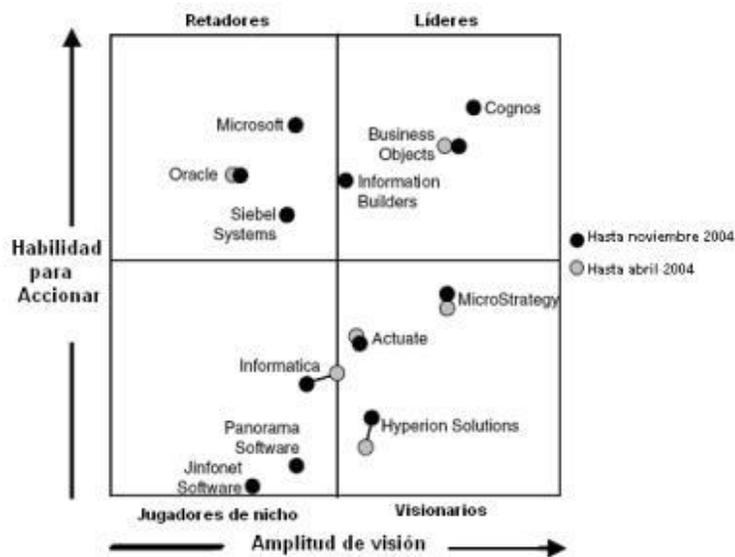


Fig. Cuadrante mágico de posicionamiento de plataformas BI.

El cuadrante mágico debe tomarse como una herramienta y no como una guía específica de acción. En el caso de BI Empresarial, el gran visionario es COGNOS. (Gartner, 2005)

3.7. Examine la situación Actual.

¿Cuál es el nivel de satisfacción de los usuarios con la información disponible?

Entender la situación actual del negocio, las necesidades más apremiantes y cuál es su visión hacia el futuro, con el fin de generar propuestas que respondan a estas necesidades es el objetivo de esta fase.

Metodología: A través de entrevistas con personal clave del negocio se identifican oportunidades de mejora en los actuales procesos de información del negocio.

Entregables: Documento de la situación actual que nos indique cuales son los indicadores actuales, frecuencias de generación, confiabilidad, fuentes de información y grado de satisfacción del usuario.

Es muy importante entender cual es la situación actual del negocio y direccionar todos los esfuerzos hacia la visión y misión del negocio, es decir si el objetivo de la organización es hacia la mayor obtención del margen en los productos, pues hacia ese sentido debemos enfocar las preguntas estratégicas que el BI va a servir para aclarar.

Al entender la situación actual y el grado de satisfacción que los usuarios tienen con la información disponible podemos tener un punto de comparación contra lo que se desea tener una vez implementado BI en la organización.

3.8. Preguntas Estratégicas de Negocio que el BI va a servir para aclarar.

Alguna información es más importante que otra, esta es el “Punto dulce”. La mas valiosa información para la organización esta en un número pequeño de información “Punto Dulce” y están ligadas con actividades de administración clave.

Al entrevistar al personal clave de la organización de cual es la información que les gustaría tener en un sistema BI podemos llenarnos de un sin número de requerimientos, aquí es donde no podemos perder de vista cual es la finalidad de la implementación de inteligencia de negocios y no contaminar el sistema con información no valiosa para la organización. Dentro del capítulo 4 se hablará sobre técnicas de la ingeniería de requerimientos que nos ayudarán en este proceso.

Una técnica que nos podría ayudar es alinear estas preguntas a través de las técnicas de Balanced Score Card donde nuestros indicadores los debemos agrupar en 4 secciones; clientes, procesos de negocio internos, innovación y aprendizaje, y finanzas alineados estos hacia la visión y misión del negocio para poder cubrir con todas las expectativas de los directivos.

Perspectivas de Negocio Clave

Clientes

¿Como deberíamos aparecer ante nuestros clientes consiguiendo nuestra visión?

Procesos de negocio internos

¿Cuales de nuestros proceso de negocios deberíamos sobresalir para satisfacer a nuestros proveedores y clientes?

Innovación y aprendizaje

¿Cómo debemos sustanciar nuestras habilidades para cambiar y tratar de alcanzar nuestra visión?

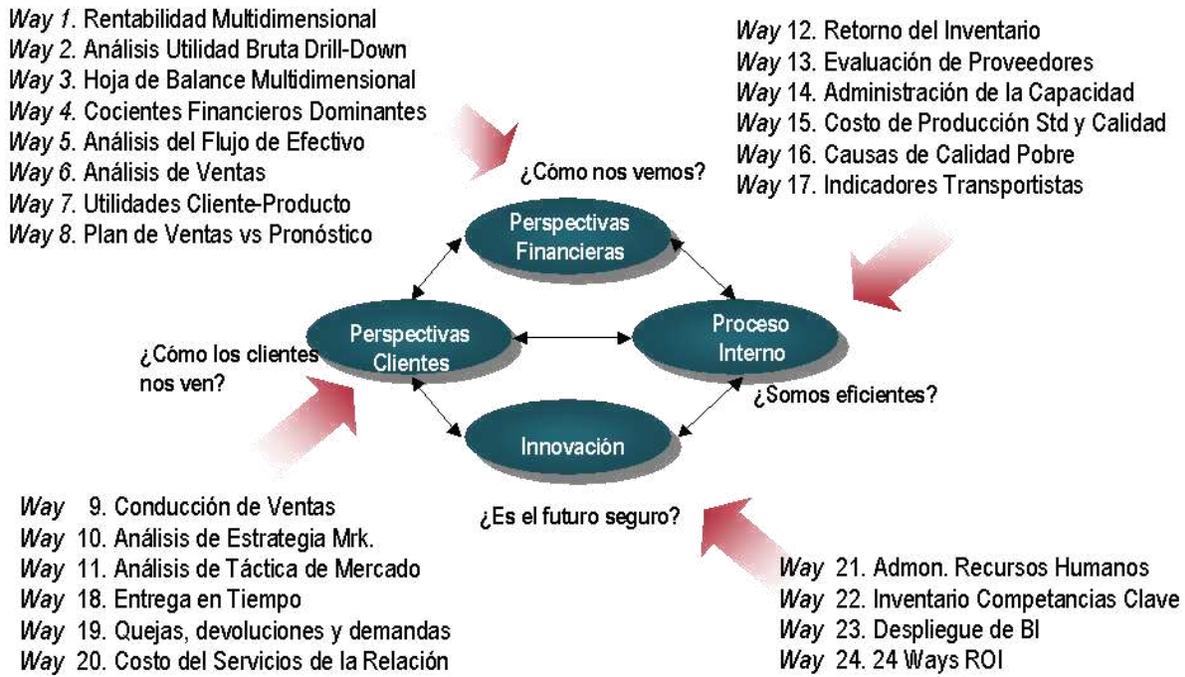
Finanzas

¿Cómo debemos aparecer ante nuestros socios para ser exitosos financieramente?

Estudiosos en el tema nos sugieren 24 indicadores, usted puede tomar estos como referencia para construir los que apliquen a su organización.



Alineando los indicadores con las perspectivas de negocio clave quedarían de la siguiente manera:



3.9. Identifique Posibles Fuentes de Información.

¿Son suficientes las fuentes que tengo para responder las preguntas de negocio planteadas?

Al tener identificadas las preguntas de negocio que el BI servirá para aclarar podrás darte cuenta tal vez que lo que revisaste en la situación actual del negocio y sus sistemas de información no basta para responder las preguntas planteadas.

Algunas aplicaciones de BI utilizan los datos extraídos de diversas fuentes las cuales están clasificadas en: **internas** (donde los datos provienen de diversos lugares. Ejemplos de estos datos son datos sobre personas, productos, excel,

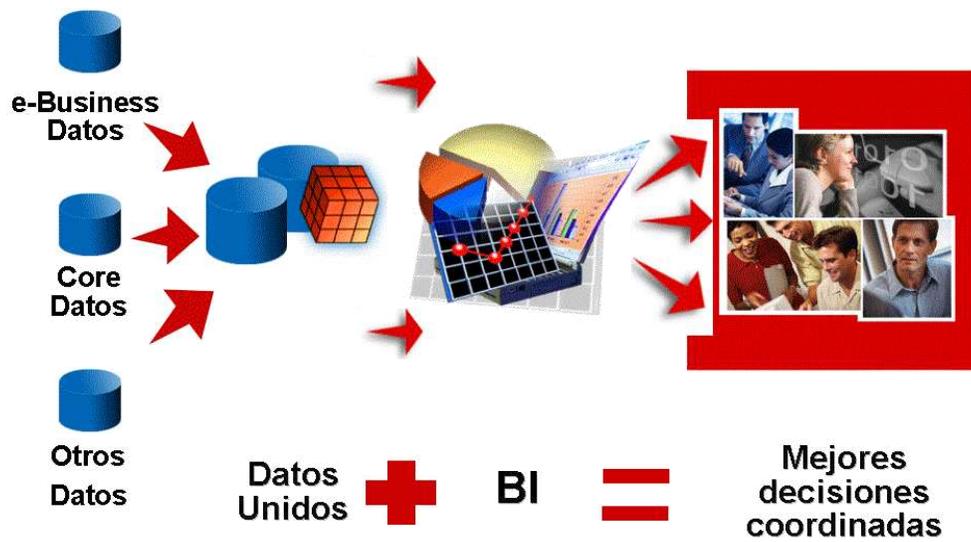
access, servicios y procesos), **externos** (Los datos provienen de fuentes externas. Estos datos pueden estar disponibles en CD-ROM, internet, películas y música o voces. Pueden provenir de todo el mundo y algunos pueden ser irrelevantes para las aplicaciones) y **personales** (donde los datos provienen de los usuarios o de otras corporaciones empleadas para contribuir a su propio expertís por la creación de datos personales).

La calidad y la integridad de los datos son puntos críticos para un BI, ya que forman el núcleo del sistema de toma de decisiones. El principal problema es la sintetización de datos, y el que estos no sean correctos o que no sean almacenados correctamente.

3.9.1. Internet y Servicios de Bases de Datos Comerciales

La recolección de datos de fuentes externas puede ser complicada, una opción es el utilizar un administrador de flujo de datos (DFM) el cual toma la información y la pone donde se necesita, cuando se necesita y en una forma utilizable.

Un DFM es una colección de sistemas y aplicaciones que administran el flujo de datos, son soporte clave en los sistemas de toma de decisiones. El DFM es un procesador central de consulta de datos, un componente de integridad de datos con ligas a proveedores de fuentes de información externos.



Como se podrá ver las fuentes de información pueden ser muchas. A través de técnicas de recolección y extracción de datos lo más recomendable es unir estas en un almacén de datos Datawarehouse lo cual se profundizará mas en el capítulo 5 "Creación del Datawarehouse".

Capítulo 4 Análisis y Diseño

4.1 Fase de Preparación

Dentro de la fase de Preparación Capítulo 3 hemos creado al equipo de trabajo que implementará la solución de BI dentro de la organización, se han identificado a los beneficiarios de dicha solución, se han definido las metas a través de indicadores cuantificables que el BI deberá contestar traducidos a preguntas estratégicas del negocio, se ha realizado un análisis de la situación actual obteniendo el grado de satisfacción de los usuarios con los sistemas actuales de información, se debieron haber identificado las fuentes de información de donde se obtendrán los datos y la maduración de los mismos, además se debió tomar la decisión de comprar o construir. Es momento ya de pasar a la fase de Análisis y Diseño.

4.2 Análisis de Requerimientos.

En la actualidad, son muchos los procesos de desarrollo de software que existen. Con el pasar de los años, la Ingeniería de Software ha introducido y popularizado una serie de estándares para medir y certificar la calidad, tanto del sistema a desarrollar, como del proceso de desarrollo en sí.

Una actividad muy importante de TI dentro de las organizaciones debería estar ligada con el modelado de procesos del negocio y la reingeniería. Un número creciente de herramientas automatizadas han surgido para ayudar a definir y aplicar un proceso de desarrollo de software efectivo. Hoy en día la economía global depende más de sistemas automatizados que en épocas pasadas; esto ha llevado a los equipos de desarrollo a enfrentarse con una nueva década de procesos y estándares de calidad.

Sin embargo, ¿cómo explicamos la alta incidencia de fallos en los proyectos de software? ¿Por qué existen tantos proyectos de software víctimas de retrasos, presupuestos sobregirados y con problemas de calidad? ¿Cómo podemos tener una producción o una economía de calidad, cuando nuestras actividades diarias dependen de la calidad del sistema?

Tal vez suene ilógico pero, a pesar de los avances que ha dado la tecnología, aún existen procesos de producción informales, parciales y en algunos casos no confiables.

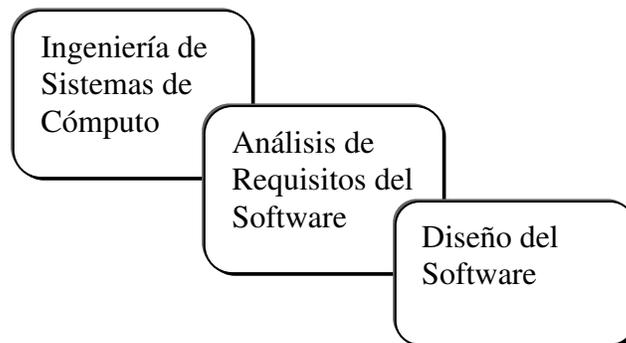
La **Ingeniería de Requerimientos** cumple un papel primordial en el proceso de producción de software, ya que enfoca un área fundamental: *la definición de lo que se desea producir*. Su principal tarea consiste en la generación de especificaciones correctas que describan con claridad, sin ambigüedades, en forma consistente y compacta, el comportamiento del sistema; de esta manera, se pretende minimizar los problemas relacionados al desarrollo de sistemas.

La razón principal de la ingeniería de requerimientos se fundamentó en la gran cantidad de proyectos de software que no llegan a cumplir sus objetivos. En nuestro país somos partícipes de este problema a diario, en donde se ha vuelto común la compra de sistemas extranjeros, para luego "personalizarlos" supuestamente a la medida de las empresas.

Estudios realizados muestran que más del 53% de los proyectos de software fracasan por no realizar un estudio previo de requisitos. Otros factores como falta de participación del usuario, requerimientos incompletos y el cambio a los requerimientos, también ocupan sitios altos en los motivos de fracasos.

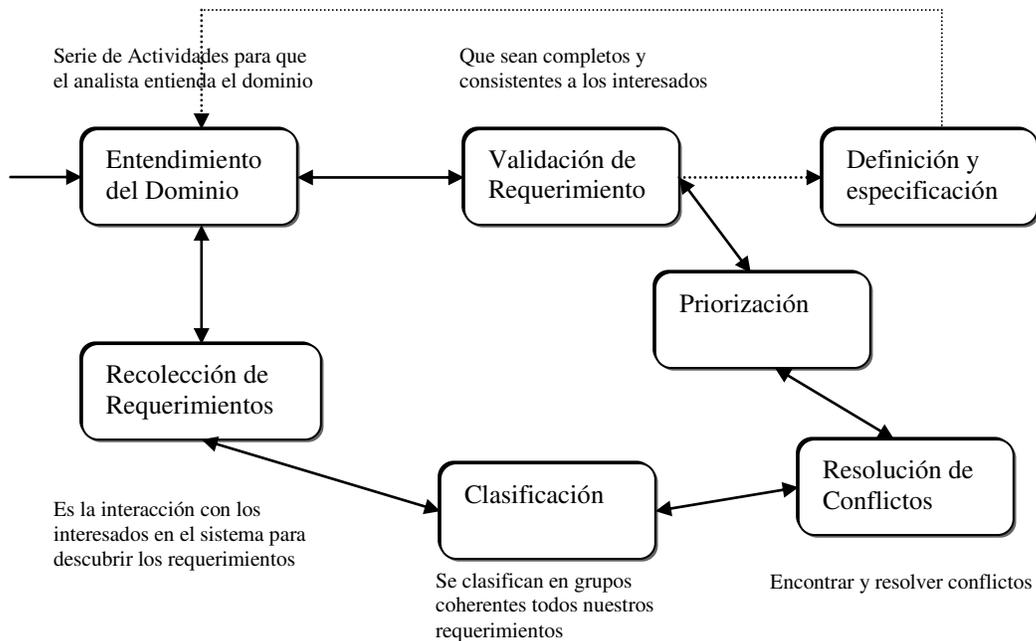
I.R. (Ingeniería de Requerimientos) Es la disciplina encargada de establecer los servicios que un sistema debe suministrar, así como las restricciones bajo las cuales debe operar.

A.R. (Análisis de Requerimientos) Es tarea de Ingeniería del Software que cubre el hueco entre definición del software a nivel sistema y diseño del software.



Proceso de Análisis de Requerimientos

Por lo tanto resumimos que dentro del proceso de análisis se deberán llevar las siguientes etapas y deberán culminar en un documento con las especificaciones.



Meta: Construir un documento con las especificaciones de requerimientos de software.

Importancia: Antes de diseñar un software, es necesario saber qué se quiere que haga dicho software.

Por medio de entrevistas a los usuarios se puede determinar el tipo de preguntas de negocio que el BI debe contestar.

- ❖ Prioridades, Atar obtención de la información con resultados concretos.

Casi la mitad del tiempo del proyecto debe ser adjudicado para analizar requerimientos de negocio y diseñando la solución tecnológica. No hacerlo sería una invitación al fracaso.

El Objetivo es el lograr establecer un conjunto de requerimientos que sean completos y consistentes.

Durante este proceso normalmente se pueden descubrir:

- ❖ Requerimientos faltantes.
- ❖ Conflictos en los requerimientos.
- ❖ Requerimientos ambiguos.
- ❖ Requerimientos traslapados.
- ❖ Requerimientos no realistas.

4.2.1 Propiedades de los Requerimientos.

¿Qué son Requerimientos?

(1) Una condición o necesidad de un usuario para resolver un problema o alcanzar un objetivo. (2) Una condición o capacidad que debe estar presente en un sistema o componentes de sistema para satisfacer un contrato, estándar, especificación u otro documento formal. (3) Una representación documentada de una condición o capacidad como en (1) o (2). Para nuestro caso de BI un requerimiento será aquella pregunta de negocio que se desea ser contestada.

Los requerimientos pueden dividirse en **requerimientos funcionales** y **requerimientos no funcionales**.

Los requerimientos funcionales definen las funciones que el sistema será capaz de realizar. Describen las transformaciones que el sistema realiza sobre las entradas para producir salidas.

Los requerimientos no funcionales tienen que ver con características que de una u otra forma puedan limitar el sistema, como por ejemplo, el rendimiento (en tiempo y espacio), interfaces de usuario, fiabilidad (robustez del sistema, disponibilidad de equipo), mantenimiento, seguridad, portabilidad, estándares, etc.

Un requerimiento para que cubra sus necesidades básicas deberá ser:

Necesario: Un requerimiento es necesario si su omisión provoca una deficiencia en el sistema a construir, y además su capacidad, características físicas o factor de calidad no pueden ser reemplazados por otras capacidades del producto o del proceso.

Conciso: Un requerimiento es conciso si es fácil de leer y entender. Su redacción debe ser simple y clara para aquellos que vayan a consultarlo en un futuro.

Completo: Un requerimiento está completo si no necesita ampliar detalles en su redacción, es decir, si se proporciona la información suficiente para su comprensión.

Consistente: Un requerimiento es consistente si no es contradictorio con otro requerimiento.

No ambiguo: Un requerimiento no es ambiguo cuando tiene una sola interpretación. El lenguaje usado en su definición, no debe causar confusiones al lector.

Verificable: Un requerimiento es verificable cuando puede ser cuantificado de manera que permita hacer uso de los siguientes métodos de verificación: inspección, análisis, demostración o pruebas.

Dificultades para definir los requerimientos

- ❖ Los requerimientos no son obvios y vienen de muchas fuentes.
- ❖ Son difíciles de expresar en palabras (el lenguaje es ambiguo).
- ❖ Existen muchos tipos de requerimientos y diferentes niveles de detalle.
- ❖ La cantidad de requerimientos en un proyecto puede ser difícil de manejar.
- ❖ Nunca son iguales. Algunos son más difíciles, más riesgosos, más importantes o más estables que otros.

- ❖ Los requerimientos están relacionados unos con otros, y a su vez se relacionan con otras partes del proceso.
- ❖ Cada requerimiento tiene propiedades únicas y abarcan áreas funcionales específicas.
- ❖ Un requerimiento puede cambiar a lo largo del ciclo de desarrollo.
- ❖ Son difíciles de cuantificar, ya que cada conjunto de requerimientos es particular para cada proyecto.

4.2.2 Personal involucrado en la Ingeniería de Requerimientos

Realmente, son muchas las personas involucradas en el desarrollo de los requerimientos de un sistema. Es importante saber que cada una de esas personas tienen diversos intereses y juegan roles específicos dentro de la planificación del proyecto; el conocimiento de cada papel desempeñado, asegura que se involucren a las personas correctas en las diferentes fases del ciclo de vida, y en las diferentes actividades.

No conocer estos intereses puede ocasionar una comunicación poco efectiva entre clientes y desarrolladores, que a la vez traería impactos negativos tanto en tiempo como en presupuesto. Los roles más importantes pueden clasificarse como sigue:

- ❖ **Usuario final:** Son las personas que usarán el sistema desarrollado. Ellos están relacionados con la usabilidad, la disponibilidad y la fiabilidad del sistema; están familiarizados con los procesos específicos que debe realizar el software, dentro de los parámetros de su ambiente laboral. Serán quienes utilicen las interfaces y los manuales de usuario.
- ❖ **Usuario Líder:** Son los individuos que comprenden el ambiente del sistema o el dominio del problema en donde será empleado el software desarrollado. Ellos proporcionan al equipo técnico los detalles y requerimientos de las interfaces del sistema.

- ❖ **Personal de Mantenimiento:** Para proyectos que requieran un mantenimiento eventual, éstas personas son las responsables de la administración de cambios, de la implementación y resolución de anomalías. Su trabajo consiste en revisar y mejorar los procesos del producto ya finalizado.
- ❖ **Analistas y programadores:** Son los responsables del desarrollo del producto en sí; ellos interactúan directamente con el cliente.
- ❖ **Personal de pruebas:** Se encargan de elaborar y ejecutar el plan de pruebas para asegurar que las condiciones presentadas por el sistema son las adecuadas. Son quienes van a validar si los requerimientos satisfacen las necesidades del cliente.

Otras personas que pueden estar involucradas, dependiendo de la magnitud del proyecto, pueden ser: administradores de proyecto, documentadores, diseñadores de base de datos, entre otros.

4.2.3 Técnicas y herramientas utilizadas en el análisis de requerimientos.

Se presentan las principales ventajas y desventajas de cada una de las técnicas utilizadas en las etapas del análisis de Requerimientos.

Técnica	Ventajas	Desventajas
Entrevistas y Cuestionarios	<ul style="list-style-type: none"> • Mediante ellas se obtiene una gran cantidad de información correcta a través del usuario. • Pueden ser usadas para obtener un pantallazo del dominio del problema. • Son flexibles. • Permiten combinarse con otras técnicas. 	<ul style="list-style-type: none"> • La información obtenida al principio puede ser redundante o incompleta. • Si el volumen de información manejado es alto, requiere mucha organización de parte del analista, así como la habilidad para tratar y comprender el comportamiento de todos los involucrados.

Lluvia de Ideas	<ul style="list-style-type: none"> • Los diferentes puntos de vista y las confusiones en cuenta a terminología, son aclaradas por expertos. • Ayuda a desarrollar ideas unificadas basadas en la experiencia de un experto. 	<ul style="list-style-type: none"> • Es necesaria una buena compenetración del grupo participante.
Prototipos	<ul style="list-style-type: none"> • Ayudan a validar y desarrollar nuevos requerimientos. • Permite comprender aquellos requerimientos que no están muy claros y que son de alta volatilidad. 	<ul style="list-style-type: none"> • El cliente puede llegar a pensar que el prototipo es una versión del software que será desarrollado. • A menudo, el desarrollador hace compromisos de implementación con el objetivo de acelerar la puesta del prototipo
Análisis Jerárquico	<ul style="list-style-type: none"> • Permite determinar el grado de importancia de cada requerimiento. • Ayuda a identificar conflictos en los requerimientos. • Muestra el orden en que deben ser implementados los requerimientos. 	<ul style="list-style-type: none"> • Debe construirse un estándar claro de evaluación, que incluya la participación del cliente.
Casos de Uso	<ul style="list-style-type: none"> • Representan los requerimientos desde el punto de vista del usuario. • Permiten representar más de un rol para cada afectado. • Identifica requerimientos estancados, dentro de un conjunto de requerimientos. 	<ul style="list-style-type: none"> • En sistemas grandes, toma mucho tiempo definir todos los casos de uso. • El análisis de calidad depende de la calidad con que se haya hecho la descripción inicial.

4.2.4 Definir los Límites del Sistema.

Los proveedores de requerimientos son frecuentemente inciertos acerca de lo que debería y no debería tener el sistema. Ellos pueden por lo tanto sugerir requerimientos inapropiados. Usted necesita dividir los pasos iniciales para eliminar requerimientos los cuales están claramente fuera del ámbito del sistema para que estos no confundan después el proceso de análisis. Al momento de eliminar requerimientos usted deberá decir a los usuarios con cuidado y sensibilidad del porque dichos requerimientos fueron descartados.

Interrogantes las cuales podrían ser de ayuda al tomar decisiones acerca de los límites del sistema son las siguientes:

- 1) ¿Un requerimiento implica la necesidad de realizar decisiones basada en información incompleta o inestable?
- 2) ¿La implementación de un requerimiento necesita información, la cual esta fuera de la base de datos definida para el sistema?
- 3) ¿Es un requerimiento concerniente con la funcionalidad principal del sistema.?

Para los requerimientos los cuales han sido asociados con el proceso operacional y los cuales han sido clasificados como fuera de los límites del sistema, usted debería preparar argumentos técnicos y/o económicos del porque estos han sido excluidos. Estos argumentos deberían estar basados en los objetivos de negocio definidos para la organización o en los resultados del estudio de la viabilidad del sistema.

Costos y Problemas

En la mayoría de los casos, usted podría ya tener alguna información de los procesos de valoración de requerimientos de modo que los costos de introducir estos requerimientos son simplemente el costo de definir la lista de control y los papeles para registrar formalmente las decisiones.

Los problemas más significativos son:

- 1) Las dificultades generales de realizar decisiones en la división de requerimientos y decisiones de comunicación para los orígenes de los requerimientos.

2) Decir a los usuarios del sistema que sus requerimientos son necesidades no válidas para ser realizadas, con cuidado y sensibilidad.

4.2.5. Uso de listas de control para el análisis de requerimientos.

Beneficio Principal: Análisis de requerimientos más rápidos y completos.

Una lista de control es una lista de cuestiones las cuales utiliza el analista para evaluar cada requerimiento.

- ❖ Un análisis basado en una lista de control (CheckList) es un método con el cual sistemáticamente se verifica cada requerimiento y acelera el proceso de análisis. Esto reduce la probabilidad de errores en el proceso de análisis y la omisión accidental de algunos requerimientos.
- ❖ Desarrollar una lista de control es una forma de rehusar el conocimiento del análisis de requerimientos a través de proyectos. El documento de lista de control conoce los problemas y evita el redescubrimiento de estos problemas en proyectos diferentes.

Costos y Problemas:

El principal costo involucrado en la introducción de estos requerimientos es establecer un acuerdo para el análisis de la lista de control. Usted puede comenzar con una propuesta o puede consultar analistas experimentados.

A pesar de que esto podría tomar pocos días para establecer una lista inicial, usted puede invertir tiempo en la validación de la lista con la gente quien esta involucrada en el proceso de análisis. Existen también costos involucrados en la administración de la lista de control.

Debería revisar la lista de control después de cada análisis para ver si algún cambio es requerido. Cuando nuevos problemas son descubiertos, la lista de control debería ser actualizada para incluir cuestiones las cuales reflejen estos problemas. Sin embargo, la lista de control es un recurso organizacional y no debería ser cambiada sin autorización.

Existen dos problemas potenciales con el análisis de la lista de control básica.

1) *Lista de control inapropiada.* Si la lista de control contiene preguntas erróneas, entonces la información resultante podría no ser útil. La lista de control debería ser revisada después de cada proyecto para decidir si se realizan cambios en ella.

2) *El enfoque de la lista de control.* Esto ocurre cuando el analista utiliza la lista de control exclusivamente para analizar los requerimientos y no se consideran por no estar incluidos en la lista.

4.2.6 Proveer Software de Apoyo en las Negociaciones.

Beneficio Principal: Proveer software para soportar las negociaciones.

Alentar el uso de sistemas electrónicos tales como e-mail para intercambiar información acerca de los requerimientos y para apoyar la negociación de requerimientos convenidos. La negociación de requerimientos usando e-mail y tablero de anuncios para el intercambio de información puede ser introducida en organizaciones de nivel de madurez básico, para las organizaciones con un nivel de madurez intermedio los sistemas de conferencia son más recomendables.

Muchos problemas de requerimientos son simplemente debido a la falta de entendimiento, donde un usuario no conoce o entiende como otros usuarios podrían usar el sistema y las facilidades del sistema que ellos requieren. Alentar una discusión electrónica de los requerimientos rompe las barreras entre los usuarios del sistema y hace más fácil y más rápido llegar a un acuerdo acerca de los requerimientos.

El intercambio de información electrónica hace posible la participación en discusiones acerca de los requerimientos a usuarios del sistema quienes pueden estar en organizaciones diferentes o geográficamente en lugares diferentes.

Factores personales son menos importantes en discusiones electrónicas que de cara a cara. Las personalidades fuertes no pueden imponer sus puntos de vista en los requerimientos.

Existen cuatro tecnologías básicas las cuales pueden ser usadas para la negociación de requerimientos electrónica.

- 1) Sistemas de correo electrónico donde los usuarios intercambian mensajes acerca de los requerimientos.
- 2) Sistemas de tablón de anuncios donde los problemas son mostrados en un tablón de anuncios común el cual es accesible para todos los usuarios.
- 3) Sistemas de base de datos compartidas tales como Lotus Notes donde un repositorio de requerimientos, problemas y comentarios pueden ser accedidos por todos los involucrados en el proceso.
- 4) Tecnologías Intranet donde software de Internet tal como browsers de la World-Wide-Web son usados para acceder datos internos en una organización.

Costos y Problemas

Los sistemas de correo electrónico y el tablón de anuncios son baratos. El costo de bases de datos compartidas y sistemas de conferencias están descendiendo rápidamente.

Puede haber algunos costos de entrenamiento involucrados en la capacitación para el uso de tableros de anuncios o bases de datos compartidas. Versiones modernas de estos sistemas tienen interfaces de usuario las cuales son relativamente fáciles de utilizar, de tal forma que la gente debería ser capaz de usar estos con la mínima capacitación.

4.2.7 Planear los conflictos y resolución de conflictos.

Beneficio Principal: Resolución más rápida de problemas en los requerimientos.

Habrán siempre conflictos, traslape y omisión en cualquier conjunto de requerimientos. Usted debería anticiparse a estos y planear la negociación de requerimientos conocidos para discutirlos y resolver los problemas descubiertos durante el análisis.

Beneficios:

- ❖ Las reuniones para la negociación de requerimientos organizados se concentran en resolver los problemas de los requerimientos reduciendo el tiempo requerido para llegar a un acuerdo del conjunto de requerimientos.
- ❖ Las reuniones para la negociación de requerimientos son un proceso abierto en el cual todos los usuarios del sistema pueden ser involucrados. Esto reduce la probabilidad de que requerimientos críticos de algunos

usuarios importantes, sean olvidados, y esto ayuda para convencerlos de que su visión ha sido considerada.

Las reuniones deberían ser conducidas en tres etapas:

- 1) Una etapa de información en donde la naturaleza de los problemas asociados con un requerimiento es explicado.
- 2) Una etapa de discusión donde los usuarios involucrados discutan como estos problemas podrían ser resueltos.
- 3) Una etapa de resolución donde las acciones concernientes a los requerimientos son puestas de acuerdo. Estas acciones podrían ser para eliminar los requerimientos, para sugerir modificaciones específicas al requerimiento o para conseguir más información acerca del requerimiento.

Costos y Problemas

No existen costos significativos involucrados en la introducción de esta directriz. El costo de aplicar esta directriz son los costos de llevar a cabo reuniones formales para la negociación. Sin embargo en algunos casos, esto puede reemplazar reuniones menos enfocadas y así no provocar costos adicionales.

Los problemas básicos de introducir esta directriz son los mismos problemas que existen para cualquier reunión, las dificultades de reunir a todos los participantes en el mismo lugar y al mismo tiempo y los problemas que surgen de las agendas ocultas de algunos participantes en la reunión.

4.2.8 Priorizar los Requerimientos.

Durante el proceso de análisis y negociación de los requerimientos, las prioridades deberían ser asignadas a los requerimientos individuales los cuales reflejan su importancia para los usuarios del sistema y para el éxito global del sistema.

Beneficios:

La asignación de prioridades ayuda a los usuarios a decidir los requerimientos principales para el sistema. La asociación explícita de las prioridades de requerimientos ayuda a los diseñadores para decidir en la arquitectura del sistema y ayuda a resolver conflictos de diseño los cuales pueden surgir.

Implementación:

Las prioridades deberían ser asignadas en discusiones entre los analistas de requerimientos y los usuarios. En algunos casos, usuarios diferentes podrían asignar una prioridad diferente para el mismo requerimiento. Esto puede reflejar necesidades reales o puede simplemente reflejar percepciones diferentes de los usuarios.

Usted debería decidir en un número relativamente pequeño de clasificación de prioridades tales como “esencial”, que significa que este debe ser incluido en el sistema. “útil”, que significa que el sistema podría ser menos efectivo sin este y “deseable”, que significa que esta facilidad no es una facilidad principal del sistema pero hace el sistema más atractivo a los usuarios de alguna forma.

Costos y Problemas:

Los costos no deberían ser significativos en la introducción y aplicación de esta directriz.

El problema principal que puede surgir en la asignación de prioridades es la asignación irreal de prioridades por parte de algunos usuarios. Ellos pueden simplemente decir que todos los requerimientos de interés para ellos son “esenciales” y por lo tanto subvertir la asignación de prioridades en el sistema. Esto causa dificultades particulares si los usuarios son clientes o gerentes en posiciones de autoridad.

Usted puede sólo oponerse a esto haciendo argumentos lógicos y económicos en contra de esto. Usted puede preguntar “Que pasa si”...

“Si esto dobla el costo del sistema al implementar este requerimiento”

4.2.9 Clasificación usando acercamiento dimensional.

Los requerimientos deben ser clasificados de modo que se identifiquen los requerimientos relacionados no solo clasificándolos en una sola clase, sino en varias.

Beneficios:

La clasificación de requerimientos es básica para descubrir relaciones entre ellos. Los conflictos y traslape son más probables entre requerimientos de la misma clase. La clasificación de requerimientos mejora la identificación de los documentos de requerimientos. Cuando se realizan cambios a los requerimientos en algunas clases, esto puede ser de ayuda para considerar si otro requerimiento en la misma clase es afectado por el cambio.

La clasificación puede ayudar a encontrar requerimientos perdidos. Si usted encuentra que no hay requerimientos en alguna clase la cual es parte de su esquema de clasificación normal, esto puede significar que requerimientos importantes han sido dejados fuera.

Implementación:

La forma más fácil para clasificar los requerimientos es usar la que es llamado método de facetas e identifica palabras clave las cuales describan cada una de estas.

Por ejemplo, posibles palabras clave las cuales podrían usarse para clasificar los requerimientos: Ejemplo: Sistema, Interfaz de Usuario, DataBase, Comunicaciones, Seguridad

Costos y Problemas:

Introducir esta directriz no es caro. Usted podría necesitar una sesión de capacitación pequeña que ilustre como los requerimientos pueden ser clasificados y como esto permite agrupar requerimientos relacionados para ser comparados. Un problema es que los requerimientos no siempre encajan exactamente en una clase y esto es una dificultad para decidir cual clasificación asignar.

4.2.10 Uso de Matrices de interacción para encontrar conflictos y traslapos.

Una matriz de interacción es una matriz donde los renglones y columnas son etiquetados con identificadores de requerimientos. Los elementos de la matriz son llenados con un valor el cual indica si existen o no conflictos o traslape en los requerimientos o si son independientes.

Beneficios:

Este es un método sistemático para comprobar requerimientos los cuales aseguren que las interacciones entre todos los requerimientos en el sistema o en una clase identificada de requerimientos son consideradas.

La matriz de interacción es una ayuda de entrada para la negociación, como esta es un resumen claro de los requerimientos los cuales pueden causar problemas y cuya forma final puede necesitar ser negociada.

Implementación:

La forma más fácil para construir una matriz de interacción es usar un programa de hoja de cálculo y etiquetar los renglones y las columnas con la identificación de los requerimientos. Cada requerimiento es entonces considerado y comparado con otros requerimientos. Usted puede entonces llenar los valores en las celdas de la hoja de cálculo como sigue:

- ❖ Cuales requerimientos tienen conflictos, llenar en un 1.
- ❖ Cuales requerimientos se traslapan, llenar con un 1000.
- ❖ Cuales requerimientos son independientes. Llenar con un 0.

La ventaja de usar valores numéricos es que usted puede sumar cada renglón y columna para encontrar el número de conflictos (por ejemplo el remanente cuando el total es dividido por 1000) y el numero de traslapes (El total dividido por 1000). Los requerimientos que tienen valores altos para una o ambas cifras debería ser cuidadosamente examinada como parte del proceso de análisis. Un número grande de conflictos o traslapes significa que cualquier cambio a esos requerimientos podría probablemente tener un impacto mayor en el resto del sistema.

Esta técnica sólo trabaja cuando usted tiene un pequeño número de requerimientos, ya que esta requiere que cada requerimiento sea comparado con cada uno de los otros requerimientos en el sistema. El límite superior es probablemente aproximadamente 200 requerimientos.

Costos y Problemas:

Este método es fácil de entender y relativamente barato para introducir en el proceso de análisis de requerimientos. Esta es una forma efectiva y sistemática para descubrir conflictos en los requerimientos, pero esto es bastante caro para implementar. Asumiendo que usted usa un programa de hoja de cálculo estándar para el análisis, usted necesita preparar las matrices, y comparar cada requerimiento con cada uno de los otros requerimientos para llenar los valores en las celdas.

El mayor problema con este método es que este es sólo conveniente para analizar un número relativamente pequeño de requerimientos.

4.2.11 Evaluar el riesgo del requerimiento

Para cada requerimiento o conjunto de requerimientos relacionados, acarrea un análisis de riesgo en el cual se sugieren posibles problemas los cuales pueden surgir en la implementación de esos requerimientos, la probabilidad de que surjan esos problemas y los efectos que podrían tener.

Beneficios:

La evaluación de los riesgos explícitos es un medio de identificar los requerimientos los cuales son probablemente la causa particular de las dificultades

para los desarrolladores del sistema. Si esto puede ser identificado en esta fase, puede posiblemente reducir los riesgos en el proceso de desarrollo.

La asignación de riesgos a los requerimientos frecuentemente revela que la información es insuficiente.

Implementación:

El análisis de riesgos es un proceso difícil y no es un método general que sea aplicable a todos los tipos de requerimientos. Es necesario involucrar la experiencia de la gente que esta involucrada en el proceso de análisis. Ellos usan sus juicios para decidir los riesgos de los requerimientos.

Identificar estos riesgos en una fase temprana significa que usted puede buscar más información para reducir los riesgos antes de que los requerimientos son finalizados.

Los tipos de riesgos que podrían considerarse son los siguientes:

- ❖ Riesgos de desempeño.
- ❖ Riesgos de procesos.
- ❖ Riesgos en las Bases de Datos.
- ❖ Riesgos de programación.
- ❖ Riesgos Externos.
- ❖ Riesgos de estabilidad.

Costos y Problemas:

Los costos de la introducción de esta directriz son los costos de identificar los tipos de riesgos que son más aplicables a su organización y el costo de idear algún método sistemático para la evaluación de los riesgos. Aplicar esta directriz

es bastante caro, así como la evaluación de riesgos es técnicamente difícil. La experiencia de la gente debe ser involucrada en el proceso y debe gastar tiempo evaluando los riesgos de cada requerimiento. El tiempo requerido deberá ser proporcional al número de requerimientos.

4.3 Evaluar la calidad de los Datos.

4.3.1. Introducción

Los sistemas de información actuales necesitan integrar grandes cantidades de información de múltiples fuentes de datos para resolver requerimientos complejos de los usuarios. Un desafío en este tipo de sistemas es proveer al usuario con información adaptada a sus requerimientos de calidad. La calidad se expresa como un conjunto de factores de calidad que miden ciertos aspectos relevantes de los resultados, como la frescura de los datos, la completitud o el tiempo de respuesta.

Los avances tecnológicos de los últimos años en materia de comunicaciones han permitido el desarrollo de sistemas de información de gran porte que brindan acceso a grandes volúmenes de información. La necesidad de acceder en forma uniforme a la información disponible en múltiples fuentes de datos, ya sean internas a una organización o accesibles a través de Internet, es cada vez más fuerte y generalizada. Dichos requisitos de información son generalmente resueltos implementando complejos procesos de manipulación de datos que implican vistas o consultas sobre fuentes de datos heterogéneas y autónomas. A medida que aumenta la cantidad de datos potencialmente recuperados, los usuarios se interesan más y más en la calidad de los resultados.

Debido a la heterogeneidad de las fuentes de datos resulta difícil evaluar la calidad de los datos para brindar a los usuarios respuestas uniformes y de alta calidad.

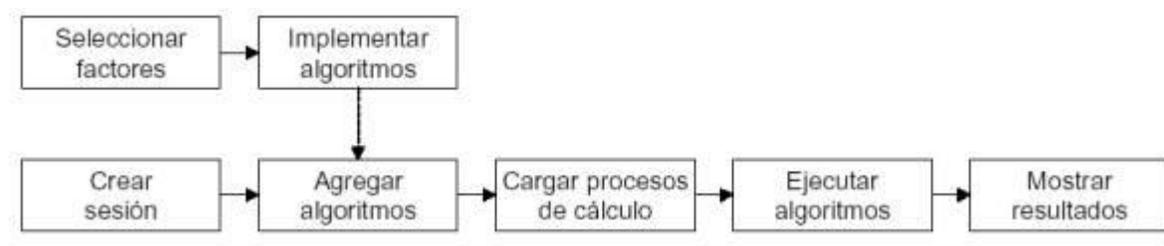
La calidad de la información devuelta al usuario depende principalmente de la calidad de las fuentes de datos y de las características del proceso de cálculo que construye dicha información a partir de las fuentes. Más concretamente, la calidad depende de la calidad interna de las fuentes (la coherencia, la completitud, la frescura, etc.), de la confianza sobre quién produce los datos de esas fuentes, y también de la forma de producir la información devuelta al usuario. En un contexto en donde la información es producida por algoritmos sofisticados de agregación, la evaluación de la calidad requiere un conocimiento fino del proceso de producción. Además, la heterogeneidad de las fuentes de datos (por ejemplo diferentes formatos o semántica de los datos) agrega complejidad a la evaluación.

La información devuelta al usuario puede ser diferente dependiendo de la forma de producirla, es decir, de las fuentes de donde se extraen los datos y de las operaciones realizadas para integrar dichos datos. Los diferentes procesos de cálculo pueden ser estudiados y comparados para seleccionar la mejor implementación del sistema de integración de acuerdo a las necesidades de los usuarios. La calidad de los datos producidos por cada proceso es un elemento importante para realizar dicha comparación.

La calidad se expresa mediante un conjunto de factores de calidad, los cuales miden ciertos aspectos del resultado que son de importancia para los usuarios, como por ejemplo, la frescura de los datos, el tiempo de respuesta o la disponibilidad de las fuentes.

4.3.2 Proceso de Evaluación de la Calidad.

El primer paso es crear una sesión, en la cual se definen las fuentes de datos a las que se accede y las clases de consultas de usuarios que obtienen datos del sistema. A continuación se agregan a la sesión los algoritmos que se quiere utilizar para evaluar la calidad, una vez configurada la sesión con información sobre las fuentes, las clases de consultas, los algoritmos de evaluación (y por lo tanto los factores de calidad) y los procesos de cálculo, se ejecutan los algoritmos.



4.3.3 Ejemplo ilustrativo

Consideremos un sistema de consultas sobre pasajes aéreos partiendo o llegando a la ciudad de Montevideo. Los usuarios están interesados en tres clases de consultas: (Q1) obtención de precios de pasajes por destino y por aerolínea; (Q2) listado de diferentes vuelos para llegar a un destino dentro de un periodo de tiempo; y (Q3) obtención del mínimo precio de pasaje para llegar a un destino.

Para obtener la información se cuenta con tres posibles fuentes de datos: (R1) con información de vuelos internacionales de las principales aerolíneas (pero no contiene información de aerolíneas pequeñas que realizan vuelos cortos); (R2) con datos de una agencia de viajes local sobre aerolíneas que salen y llegan a Montevideo y las posibles combinaciones; y (R3) con información sobre vuelos de Pluna (aerolínea local), actualizada en tiempo real.

Las consultas de usuarios pueden ser resueltas de diferentes formas, siguiendo diferentes procesos de cálculo. Para resolver las consultas de precios de pasajes se puede extraer datos de cualquiera de las fuentes o combinar datos de varias de ellas (promedio de precios, mínimo, máximo). Sucede lo mismo con las consultas de obtención de los vuelos hacia un destino, el resultado puede ser la unión de los vuelos encontrados en las fuentes, la intersección o sólo los que aparecen en alguna fuente en particular. Por lo tanto, se obtienen diferentes resultados dependiendo de la forma de resolver la consulta y dichos resultados tendrán diferentes valores de calidad. Si un vuelo de Pluna cambia de horario sería actualizado de inmediato en la fuente R3. Las consultas resueltas a partir de R3 devolverán datos actualizados, pero si se utilizan las otras fuentes los resultados podrían no estar al día. Sin embargo, los resultados sobre vuelos devueltos por R3 pueden no ser completos, ya que R3 no tiene información del resto de las aerolíneas.

Los valores de calidad más relevantes (o a los que se pretenda dar más peso) pueden ser muy diferentes dependiendo de los usuarios y sus aplicaciones. Por ejemplo, un usuario que está planeando un viaje y busca ideas de destinos (perfil tipo navegador) no se interesa tanto por la frescura o la precisión de los horarios de los vuelos, sino que el tiempo de respuesta es más relevante para él. Por el contrario, para un agente de viajes que está reservando o vendiendo un boleto la frescura de la información sobre las plazas disponibles, promociones y precios es fundamental y puede permitirse esperar más por dichos datos.

4.3.4 Modelando el sistema de integración de datos.

Para modelar los algoritmos de evaluación de la calidad es necesario estudiar y modelar las características del sistema, incluyendo propiedades de los datos fuentes utilizados como materia prima y las propiedades de los procesos de cálculo. Además es necesario tener en cuenta las preferencias de los usuarios,

tanto en los factores de calidad que son más relevantes para sus aplicaciones como en los valores de calidad esperados.

4.3.5. Modelo de DAG

Un sistema de integración de datos es un sistema de información que integra datos de diversas fuentes y provee a los usuarios de un acceso uniforme a dichos datos por medio de un modelo global. Las consultas de los usuarios se expresan en términos del modelo global. Algunos ejemplos de SID son los *Sistemas de la Mediación*, los cuales extraen e integran información de varias fuentes de datos para realizar consultas, los *Sistemas de Data Warehousing*, que extraen, transforman y resumen datos de varias fuentes (posiblemente heterogéneas) y la dejan disponible para análisis estratégico y toma de decisiones, las Federaciones de Bases de Datos, donde la autonomía de las fuentes es una característica clave y los Portales Web que proporcionan acceso a información temática, adquirida y sintetizada de fuentes Web, generalmente utilizando técnicas de Caching.

Un SID puede verse como un flujo de trabajo (workflow) compuesto de actividades que representan las diversas tareas de extracción, transformación y retorno de datos a los usuarios. Cada actividad toma como entrada datos de las fuentes u otras actividades y produce datos que pueden ser utilizados como entrada para otras actividades. De esta forma, los datos siguen un camino desde las fuentes hasta los usuarios, siendo transformados y procesados según la lógica del sistema. Los datos producidos por una actividad pueden ser inmediatamente consumidos por otra actividad o pueden materializarse para ser consultados más tarde. Observe que esta noción de actividad puede representar procesos de diversa complejidad; desde simples operaciones SQL hasta complejos procedimientos de transformación que pueden ejecutarse autónomamente.

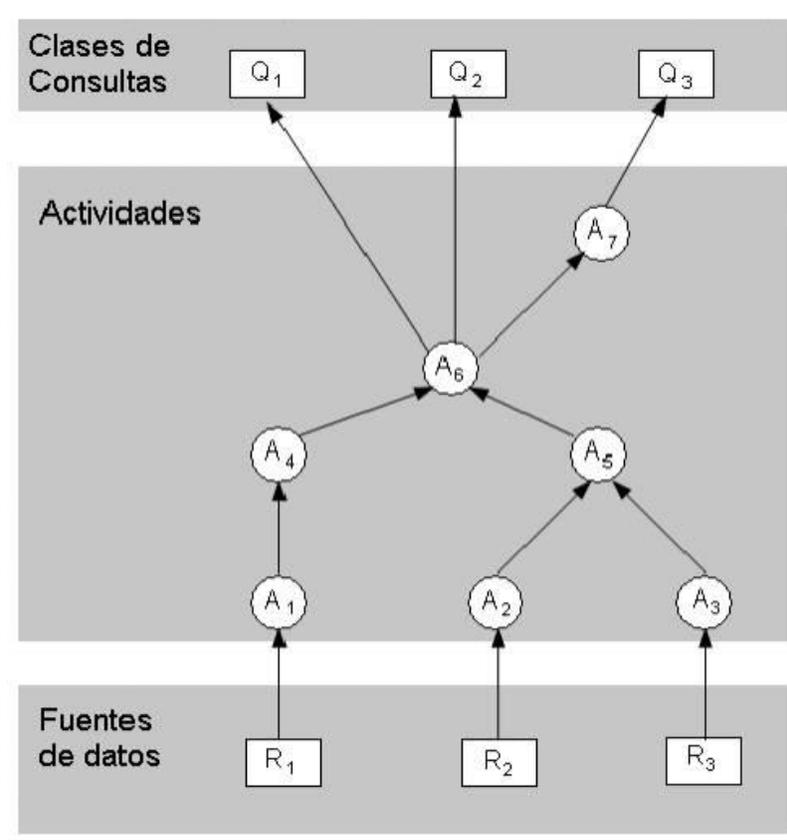
La figura de abajo muestra la representación de un SID como un flujo de trabajo. En la parte inferior se muestran las fuentes (R_i). En la parte central se muestran las diversas actividades (A_i) que ejecutan a partir de los datos fuentes. Las flechas indican que el nodo de la salida utiliza los datos producidos por el nodo de la entrada. Las actividades que toman los datos directamente de relaciones fuente son los extractores de datos (wrappers). Las otras actividades toman datos de entrada directa o indirectamente de los wrappers. En la parte superior se muestran las clases de consultas de usuario (Q_i) que representan las familias de consultas que pueden responderse usando los datos producidos por las actividades.

Formalmente, representamos el flujo de trabajo del SID por medio de un grafo acíclico dirigido (DAG) que describe las actividades implicadas, sus entradas y salidas. El DAG muestra el flujo de datos desde las fuentes hasta las clases de consultas pasando por las diversas actividades.

Un DAG de cálculo (CDAG) G , es un grafo acíclico dirigido definido de la siguiente manera: Los nodos de G son de tres tipos: nodos fuente (sin ninguna arista entrante) que representan las fuentes, nodos destino (sin ninguna arista saliente) que representan las clases de consultas, y nodos de actividades (con aristas entrantes y salientes) que representan las diferentes actividades que calculan el conjunto de nodos destino a partir de los nodos fuente. Las aristas de G indican que un nodo es calculado a partir de otro (los datos transitan en el sentido de la flecha).

Ejemplo 1. En la figura se muestra un CDAG que representa un proceso de cálculo para resolver las consultas del ejemplo de la sección 4.3.3. Las actividades A_1 , A_2 y A_3 extraen la información de las fuentes de datos. La actividad A_4 filtra los datos de las compañías que pueden ser operadas desde Uruguay. A_5 integra información de A_2 y A_3 . A_6 realiza la unión de la información

de A4 y A5. Por último, A7 computa una función agregada sobre los datos de A6 obteniendo el mínimo precio por cada destino.



Para la realización correcta evaluación de calidad deberá elabora un modelo de todos sus requerimientos (Indicadores clave de Proceso) a desarrollar esquematizado como el ejemplo anterior u otro modelo seleccionado.

4.3.6 Propiedades asociadas al cálculo de la calidad.

En esta sección describimos las características y propiedades del SID necesarias para expresar la calidad del sistema. Para realizar la evaluación se necesita, en primer lugar, identificar qué factores de calidad se van a evaluar. La elección de los factores de calidad más apropiados para un determinado SID

depende de los requerimientos del usuario y sus aplicaciones. Varios trabajos estudian los factores de calidad que son más relevantes para diversos tipos de sistemas, por ejemplo. La selección de los factores de calidad implica la selección de las métricas apropiadas y la implementación de los algoritmos de la evaluación para esos factores.

Los algoritmos de evaluación necesitan como entrada cierta información que describe las características del sistema, por ejemplo, el tiempo que una actividad necesita para ejecutar o un descriptor que indica si una actividad materializa datos o no. Las propiedades pueden ser de dos tipos: (i) descripciones, indicando una cierta característica del sistema (costos, retardos, políticas, estrategias, restricciones, etc.), o (ii) medidas, indicando un valor que corresponde a un factor de calidad, que puede ser un valor real adquirido de una fuente, un valor calculado por un algoritmo de evaluación o un valor esperado por un usuario.

La elección de las propiedades adecuadas depende de los factores de calidad y del tipo de sistema. A continuación se describe como ejemplo el factor de calidad frescura y tres propiedades necesarias para su evaluación:

Frescura: La frescura de los datos (data freshness) representa la “edad” de los datos, indicando qué tan “viejos” son dichos datos y si son “apropiados” para un determinado requerimiento. Se mide como el tiempo transcurrido entre el momento en que los datos fueron producidos y el momento en que dichos datos son retornados al usuario en respuesta a una consulta.

En la *medición de la frescura* de los resultados son importantes: la frescura de los datos en las fuentes (*frescura real*), los tiempos de ejecución de las actividades (*costo de procesamiento*) y los tiempos de espera entre la ejecución de las actividades (*demora de sincronización*). También es importante conocer las

expectativas de frescura de los usuarios (*frescura esperada*) para comparar con las mediciones. A continuación se describen dichas propiedades:

Costo de procesamiento: El costo de procesamiento de una actividad es el lapso de tiempo, en el peor caso, que la actividad necesita para leer los datos de entrada, ejecutar y producir el resultado.

Hay varios retardos relacionados con el costo de procesamiento de la actividad. Para los wrappers incluye el tiempo necesario para comunicarse con la fuente (enviar de la consulta y recibir la respuesta), realizar la extracción y materializar los resultados (en caso necesario). Para las otras actividades incluye el tiempo necesario para leer los datos de entrada, procesar los datos y materializar los resultados (en caso necesario).

Demora de sincronización: Cuando dos actividades consecutivas en un camino del CDAG ejecutan con diferentes frecuencias (por ejemplo una ejecuta una vez al día y otra una vez por semana), los datos producidos por la primera se deben materializar para ser consultados más adelante por la segunda. En tal caso, hay una demora de sincronización. Dicha demora es la cantidad de tiempo transcurrido entre el final de la ejecución de una actividad y el comienzo de la otra. Las demoras de sincronización son muy importantes en la evaluación de la frescura porque introducen tiempos de espera suplementarios y por lo tanto disminuyen la frescura de los datos.

Frescura real: La frescura real es la medida de la frescura de datos en una fuente. La misma puede ser proporcionada por la fuente, puede ser estimada o acotada por el sistema o puede ser estimada por usuarios expertos.

Frescura esperada: La frescura esperada es el valor de frescura máximo esperado o tolerado por los usuarios. La misma puede ser proporcionada por los

usuarios o estimada a través de su comportamiento (logs, históricos de requerimientos, etc.).

Una propiedad se relaciona con los ciertos nodos o aristas del CDAG. Por ejemplo, podemos asociar el costo de procesamiento a los nodos de actividades, la frescura real a los nodos fuente y la demora de sincronización a las aristas.

4.3.7 Algoritmos de Evaluación

El cálculo de la calidad es realizado por algoritmos de evaluación. En este capítulo se dio un overview al algoritmo de evaluación de la frescura que calcula la frescura que tendrán los datos devueltos a los usuarios por un proceso de cálculo. El objetivo de esta tesis no es mostrar a detalle los algoritmos de evaluación de la calidad de los datos, por lo que no se entrará en mayor detalle lo importante es hacer énfasis en la importancia de la calidad de los datos de los cuales se hará toma de decisión. Actualmente existen herramientas para realizar de una forma sencilla esta actividad.

4.4 Modelo de Datos de Negocio.

4.4.1 Introducción

Un archivo es un elemento de información conformado por un conjunto de registros. Estos registros a su vez están compuestos por una serie de caracteres o bytes. Los archivos, alojados en dispositivos de almacenamiento conocidos como memoria secundaria, pueden almacenarse de dos formas diferentes: archivos convencionales o bases de datos.

Los archivos convencionales, pueden organizarse como archivos secuenciales o archivos directos. Sin embargo, el almacenamiento de información a través de archivos convencionales presenta una serie de limitaciones que

restringen de manera importante la versatilidad de los programas de aplicación que se desarrollan.

4.4.2 Definición de Base de Datos

Se define una base de datos como una serie de datos organizados y relacionados entre sí, los cuales son recolectados y explotados por los sistemas de información de una empresa o negocio en particular.

Las bases de datos proporcionan la infraestructura requerida para los sistemas de apoyo a la toma de decisiones y para los sistemas de información estratégicos, ya que estos sistemas explotan la información contenida en las bases de datos de la organización para apoyar el proceso de toma de decisiones o para lograr ventajas competitivas. Por este motivo es importante conocer la forma en que están estructurados las bases de datos y su manejo.

4.4.3 Los modelos de datos

En el proceso de abstracción que conduce a la creación de una base de datos desempeña una función prioritaria el modelo de datos. El modelo de datos, como abstracción del universo de discurso, es el enfoque utilizado para la representación de las entidades y sus características dentro de la base de datos, y puede ser dividido en tres grandes tipos:

1. Modelos lógicos basados en objetos: los dos más extendidos son el modelo *entidad-relación* y el *orientado a objetos*. El modelo entidad-relación (E-R) se basa en una percepción del mundo compuesta por objetos, llamados entidades, y relaciones entre ellos. Las entidades se diferencian unas de otras a través de atributos. El orientado a objetos también se basa en objetos, los cuales contienen valores y métodos, entendidos como órdenes que actúan sobre los valores, en niveles de anidamiento. Los objetos se agrupan en clases,

relacionándose mediante el envío de mensajes. Algunos autores definen estos modelos como "modelos semánticos".

2. Modelos lógicos basados en registros: el más extendido es el *relacional*, mientras que los otros dos existentes, *jerárquico* y *de red*, se encuentran en retroceso. Estos modelos se usan para especificar la estructura lógica global de la base de datos, estructurada en registros de formato fijo de varios tipos. El modelo relacional representa los datos y sus relaciones mediante tablas bidimensionales, que contienen datos tomados de los dominios correspondientes. El modelo de red está formado por colecciones de registros, relacionados mediante punteros o ligas en grafos arbitrarios. El modelo jerárquico es similar al de red, pero los registros se organizan como colecciones de árboles. Algunos autores definen estos modelos como "modelos de datos clásicos".

3. Modelos físicos de datos: muy poco usados, son el modelo unificador y el de memoria de elementos. Algunos autores definen estos modelos como "modelos de datos primitivos".

4.4.4 Objetivos del modelo de datos.

Los objetivos del modelo de datos son dos:

1. **Formalización:** definir formalmente las estructuras permitidas y las restricciones a fin de representar los datos de un sistema de información.
2. **Diseño:** el modelo resultante es un elemento básico para el desarrollo de la metodología de diseño de la base de datos.

Los diferentes modelos de datos comparten, aunque con diferentes nombres y notaciones, unos elementos comunes, componentes básicos de la representación de la realidad que realizan. Estos componentes se identifican

gracias a la clasificación, y pueden identificarse conceptos *estáticos* y conceptos *dinámicos*.

Los conceptos *estáticos* corresponden a:

- 1. Objeto:** cualquier entidad con existencia independiente sobre el que almacenan datos. Puede ser simple o compuesto.
- 2. Relación:** asociación entre objetos.
- 3. Restricción estática:** propiedad estática del mundo real que no puede expresarse con los anteriores, ya que sólo se da en la base de datos; suele corresponder a valores u ocurrencias, y puede ser sobre atributos, entidades y relaciones.
- 4. Objeto compuesto:** definidos como nuevos objetos dentro de la base de datos, tomando como punto de partida otros existentes, mediante mecanismos de agregación y asociación.
- 5. Generalización:** se trata de relaciones de subclase entre objetos, es decir, parte de las características de diferentes entidades pueden resultar comunes entre ellas.

Por su parte, los conceptos *dinámicos* responden a:

- 1. Operación:** acción básica sobre objetos o relaciones (crear, modificar, eliminar...).
- 2. Transacción:** conjunto de operaciones que deben ejecutarse en su conjunto obligatoriamente.
- 3. Restricción dinámica:** propiedades del mundo real que restringen la evolución en el tiempo de la base de datos.

4.4.5 Modelo de entidades y relaciones (MER)

- Introducido por Peter Chen en 1976.

- Fundamento formal
- Componentes básicos:
 - ❖ Entidades
 - ❖ Atributos (identificadores y descriptores)
 - ❖ Relaciones

Un entidad es...

- ❖ Puede ser tanto un ítem encontrado en la realidad como un concepto abstracto.
- ❖ Unidad Semántica
- Persona: EMPLEADO, ESTUDIANTE, PACIENTE
- Lugar: CIUDAD, AULA, DELEGACION
- Objeto: MAQUINA, AUTO
- Evento: ORDEN DE PEDIDO, FACTURA
- Concepto : CUENTA, CURSADA

Las representaremos...

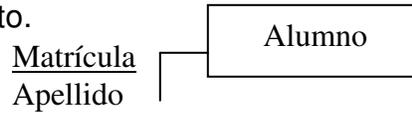
Con un rectángulo con el nombre de la misma en su interior.



Un atributo es...

- A los atributos se los denomina también propiedades son características de una entidad o relación.
- Contexto.
 - ❖ Atributos simples, por ej. Número de cliente, Apellido, Nombre, Número de Paciente.
 - ❖ Atributos Compuestos, por ej. Dirección la que puede estar compuesta por por Calle, Número, Ciudad, Código Postal.

- ❖ Atributos almacenados vs atributos derivados, existen situaciones en las que un atributo puede ser calculado, por ejemplo la edad a partir de su fecha de nacimiento.



Una instancia es...

Una instancia es una ocurrencia en una entidad.

La entidad es un molde que tiene un conjunto de propiedades comunes a todos los elementos que pertenecen a ella. Una instancia es una ocurrencia única de una entidad.

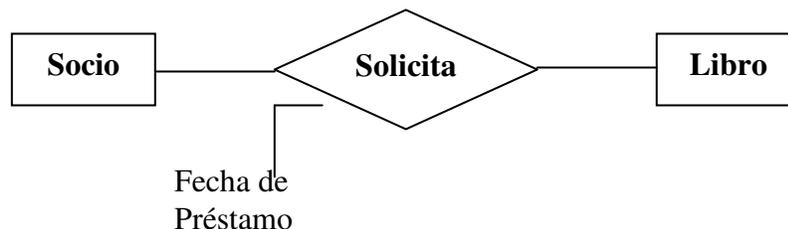


Cada instancia de una entidad debe poder ser identificada de manera única, por un atributo (o conjunto de ellos) que se lo denomina identificador.

Una relación es...

Una relación es una asociación entre instancias de una o más entidades.

- Se identifican con un rombo.
- Se utiliza algún verbo para describir la relación.
- Puede tener atributos



- Grado u orden: cantidad de entidades que participan en una relación.
- Cardinalidad o multiplicidad: cantidad de instancias de las entidades involucradas que participan en una relación.

- Opcionalidad/obligatoriedad Relaciones

4.4.6 Ejemplo

Una industria fabrica varios PRODUCTOS, y cada uno a su vez en distintas PRESENTACIONES. Cada presentación de cada producto tiene un código de identificación único. Debe considerarse la información para:

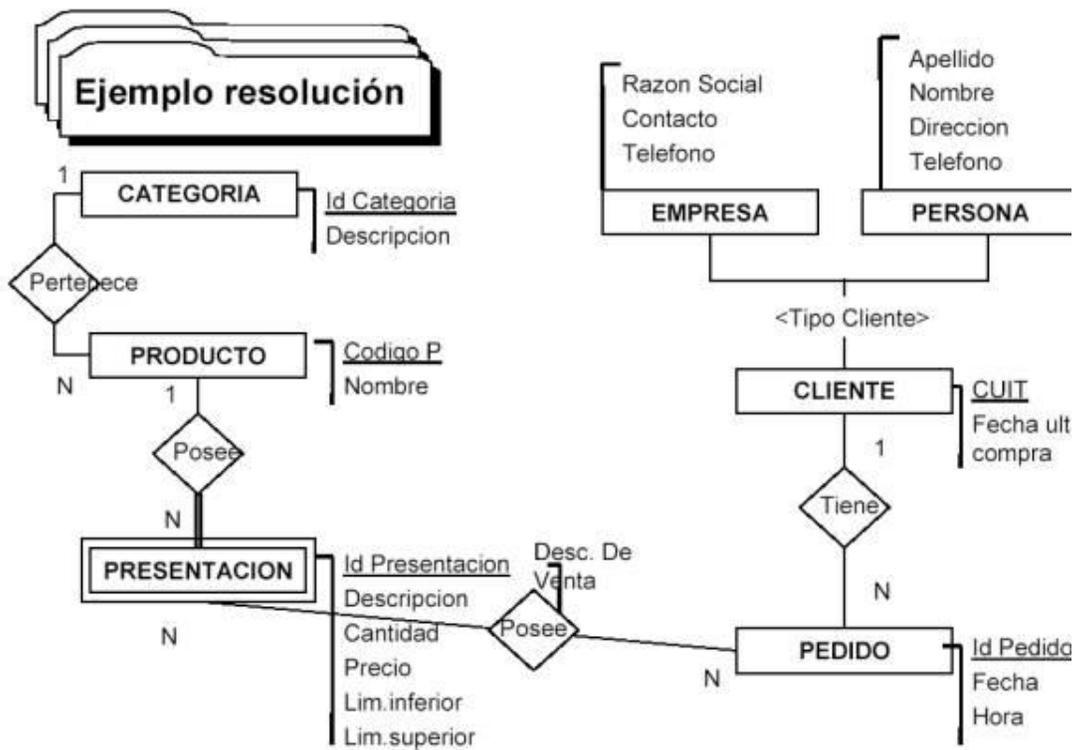
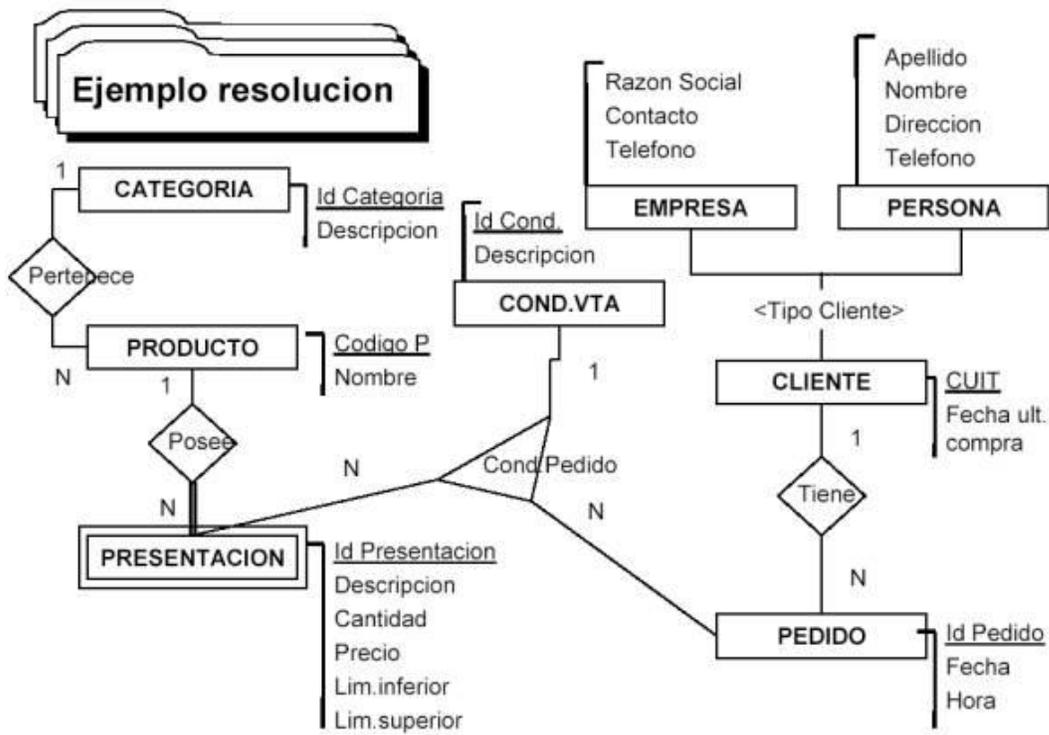
Manejo de Stocks. Los datos de stock deben comprender no sólo la cantidad en existencia de cada presentación de un producto sino también los límites tolerados de variación y el precio de venta actual. Cada producto pertenece a una CATEGORÍA.

Clientes. Los CLIENTES se identifican unívocamente por su código único de identificación tributaria CUIT que puede pertenecer a PERSONAS físicas o EMPRESAS; se debe registrar además para cada uno la fecha de la última compra que efectuara a la fábrica.

Pedidos. Los PEDIDOS de los clientes poseen un identificador que los individualiza y deben incluir, igual que las facturas, la fecha en que se produjeron.

Condiciones de venta. Los clientes efectúan pedidos sobre determinadas presentaciones de productos con CONDICIONES DE VENTA específicas.

Resolución:



Capítulo 5 Implementación

5.1 Modelo Físico de Datos (Creación del Datawarehouse).

Es la aplicación del Modelo Lógico de Datos en sistema de base de datos que se debe optimizar de acuerdo a las capacidades del motor de base de datos seleccionado. Idealmente la estructura del modelo lógico podrá ser transferida al físico sin mayores cambios.

Para aquellos que desarrollan y mantienen los sistemas de datos haciéndolos disponibles para los directivos de las empresas, el término *DataWareHouse* o también conocido como *Almacén de Datos*, ofrece la solución como ubicación central para que todos puedan acceder a la información con los reportes necesarios, dando respuesta a necesidades de diferentes tipos de usuarios.

El Datawarehouse surgió con el objetivo de hacer consultable la información que se tiene de una empresa tanto de meses como de años anteriores, organiza y orienta los datos desde la perspectiva del usuario final, mientras que los sistemas operacionales organizan sus datos desde la perspectiva de la aplicación, para lograr eficiencia en el acceso a datos.

Los principales objetivos de un Datawarehouse son:

- Comprender las necesidades de los usuarios por áreas dentro del negocio.
- Determinar qué decisiones se pueden tomar con la ayuda del DW .
- Seleccionar un subconjunto del sistema de fuentes de datos que sea el más efectivo y procesable para presentar el DW.

- Asegurar que los datos sean precisos, correctos y confiables y que mantengan la consistencia.
- Monitorear continuamente la precisión y exactitud de los datos y el contenido de los reportes generados.
- Publicar los datos.

5.1.1. ¿Por qué se justifica construir un Datawarehouse?

Generalmente, los sistemas transaccionales o OLTP usan estructuras normalizadas, en las cuales se optimizan las inserciones y actualizaciones de artículos e incluso algunas selecciones, pero es menos probable que el sistema se organice de forma tal que produzca reportes eficientes para datos resumidos con cierta jerarquía. Y es aquí donde debería usarse el DW, que usa los datos relevantes de fuentes existentes y los combina en una estructura que ha sido optimizada para las selecciones.

Esta es la razón por la cual se construye un datawarehouse para solucionar la problemática de tener un sistema fuente transaccional. Esta información se necesita que esté consultable para los clientes de la empresa de forma remota y sin embargo, por problemas de seguridad no puede estar directamente disponible desde el mismo sistema fuente.

El datawarehouse ha sido la solución propuesta para que la información sea utilizada por una aplicación cliente de acceso remoto. De esta forma se aprovecha la forma en que se organizan los datos en el almacén, en el modelo dimensional y se brinda a la gerencia un grupo de informaciones organizadas en cubos multidimensionales que les permite profundizar en el análisis de la información y ver su variación en el tiempo.

Un DW debe tener cuatro características que son primarias . Es una colección de datos ***orientada a un tema, integrada, variable en el tiempo y no volátil*** que sea útil para la toma de decisiones.

Es ***orientada a un tema*** porque tiene en cuenta los procesos de negocio de la empresa que se deseen priorizar. Es ***integrado*** porque agrupa a todos los sistemas operacionales en un sistema de información con formatos y códigos consistentes. Es ***variante en el tiempo*** porque los datos se organizan y almacenan en jerarquías en el tiempo, lo que permiten análisis comparativos de estados actuales y de períodos anteriores. Es ***no volátil*** porque se usa principalmente para operaciones de recuperación de información y no para actualizaciones.

Los DW están en la categoría de los sistemas para el soporte de decisiones (DSS) que tienen como objetivos medir y controlar el desarrollo de las variables importantes del negocio, buscando identificar, proyectar y predecir tendencias a partir de los datos acumulados.

Los datos que se manejan en el DW son informacionales, esto significa que son datos resumidos y periódicos a diferencia de los datos operacionales.

5.1.2. Comenzando a construir un DW.

Para llevar a cabo con éxito un proyecto Datawarehouse, es vital considerar al inicio de su construcción tres factores esenciales: RH, tecnología y disciplina.

La disciplina es fundamental para el desarrollo del DW. Estas disciplinas son usadas para asegurar calidad, lograr sinergia, y mejorar el equipo de trabajo

durante todo el proceso de desarrollo. Así los siguientes factores resultan ser imprescindibles para llevar a cabo la implementación de un DW:

- ❖ Prácticas de trabajo efectivo en el equipo de trabajo participante en el proyecto para lograr metas compartidas.
- ❖ Estándares, convenciones de calidad y en general la forma de resultados.
- ❖ Una metodología de desarrollo la cual defina formalmente los pasos y resultados del desarrollo.

5.1.3. Análisis del Problema

La complejidad en el desarrollo se ha presentado como la principal desventaja de un DW. Esto se debe a que la realidad para cada negocio es distinta, y un DW debe responder a las características particulares que presenta cada uno de ellos, tanto de configuración como del conjunto de requisitos a satisfacer; por lo cual no es fácil estandarizar la forma de desarrollar este tipo de proyectos.

El empleo de una forma de trabajo ordenada es un factor de importancia en el desarrollo e implantación de proyectos de Datawarehousing, y la tendencia en general busca lograr a través del uso de una metodología, recortar los tiempos de desarrollo y programar la inversión de recursos de manera eficiente; además proporciona un lenguaje común logrando que exista comunicación, permitiendo la incorporación de nuevos miembros al equipo de trabajo siendo productivos inmediatamente.

En la actualidad no podemos asegurar cuál estrategia de implementación es mejor o peor, sin embargo al analizar las tendencias generales del mercado se encuentra que la estrategia de desarrollo de datamarts está siendo adoptada con mayor frecuencia en los últimos tiempos. A esta tendencia general se le ha identificado como la aproximación que garantiza la probabilidad de éxito más

grande en la implantación de datawarehousing, tanto por la rapidez en la obtención de resultados en períodos cortos con inversiones moderadas como por la modularidad posible de alcanzar con este enfoque considerando cada datamart como un incremento del sistema final (datawarehouse).

5.1.4. Construcción de un DW.

Al iniciar un proyecto Datawarehousing no debemos olvidar establecer un marco de referencia de construcción del DW. Podemos distinguir en dicha construcción dos etapas principales: *la definición de una arquitectura DW y la construcción de los incrementos DW.*

5.1.4.1 Definición de una arquitectura.

Arquitectura DW establece el marco de trabajo, estándares y procedimientos para el DW a un nivel empresarial. Los objetivos de las actividades de la arquitectura son simples, integrar al DW las necesidades de información empresarial.

El DW necesita de tres áreas bien diferentes entre sí pero que se integran para lograr un buen funcionamiento:

1. Los sistemas origen o fuentes legados .
2. El área de transformación de los datos (ETL la cual se profundizará en el punto 5.2)
3. El Servidor de presentación del DW

Las *fuentes legados* son los sistemas de procesamiento que capturan las transacciones del negocio. Deben ser sistemas confiables y consistentes , aunque entre ellos hay marcadas diferencias en los formatos y las estructuras de los datos. Quedan fuera del DW por lo que no tenemos el control sobre el contenido de sus datos.

El área de transformación de los datos (ETL) consta tanto del área de almacenamiento como del conjunto de procesos que se usan frecuentemente para la extracción, transformación y carga de los datos . Es generalmente la parte más compleja de esta arquitectura. Es todo lo que se presenta entre la fuente externa y el área de presentación de los datos.

El servidor de presentación es la fuente consultable de datos de la empresa. Es donde se organizan los datos del DW, se almacenan y están disponibles para hacer consultas, reportes y acceder mediante aplicaciones.

El DW se soporta sobre el modelo dimensional a diferencia de los sistemas de bases de datos que están basados en el modelo Entidad-Relación. Este modelo contiene la misma información que el modelo E/R pero empaqueta los datos en un formato simétrico cuyo objetivo es ganar una mayor comprensión del usuario y garantizar la ejecución rápida y eficiente de las consultas. A diferencia del modelo E/R, el modelo dimensional no necesita anticipar las consultas que se van a realizar y es muy elástico a los cambios que se produzcan en los patrones de los usuarios.

5.1.4.2. Construcción en Incrementos, Datamarts.

Para la implementación en incrementos el warehouse desarrolla y genera un subconjunto del DW total. La implementación incremental es una aproximación pragmática para construir un warehouse a un nivel empresarial en forma evolutiva.

El catálogo de Datamarts(metadatos) juega un papel decisivo en la arquitectura del DW , ya que suministra los parámetros y la información que

permite que la aplicación realice sus tareas; contiene un conjunto de información de control del DW, su contenido, sus fuentes de origen y sus procesos.

Hasta este punto, el catálogo de datos es sólo un concepto lógico y cualquier cambio en él se reflejará en toda la arquitectura y estará disponible para todos los servicios a la vez.

En la mayoría de los casos, no es muy práctico traer toda la información a un solo lugar. Los Datamarts(metadatos) existen en varias herramientas, programas y utilitarios que hacen que el DW funcione.

Los resultados más significativos de la implementación de un incremento DW, incluyen:

- ❖ Las bases de datos warehouse.
- ❖ Programas y procedimientos para extraer, transformar y cargar datos.
- ❖ Instalar herramientas de acceso a los datos.
- ❖ Poblar el DW con los datos necesarios.
- ❖ Poblar el catálogo de metadatos con los datos necesarios.
- ❖ Técnicas de uso y soporte el almacén.

5.1.5. Consideraciones de Implementación mediante DataMarts.

- ❖ La arquitectura Datawarehouse se debe desarrollar al principio del proyecto.
- ❖ El primer incremento se desarrolla basándose en la arquitectura.
- ❖ La construcción del primer incremento puede causar cambios en la arquitectura.
- ❖ La operación del DW puede implicar la realización de cambios en la arquitectura.
- ❖ Cada incremento adicional puede extender el datawarehouse.

- ❖ Cada nuevo incremento puede causar ajustes en la arquitectura.
- ❖ La operación continua puede causar ajustes en la arquitectura.

La siguiente figura, intenta esquematizar el proceso de construcción del Datawarehouse.

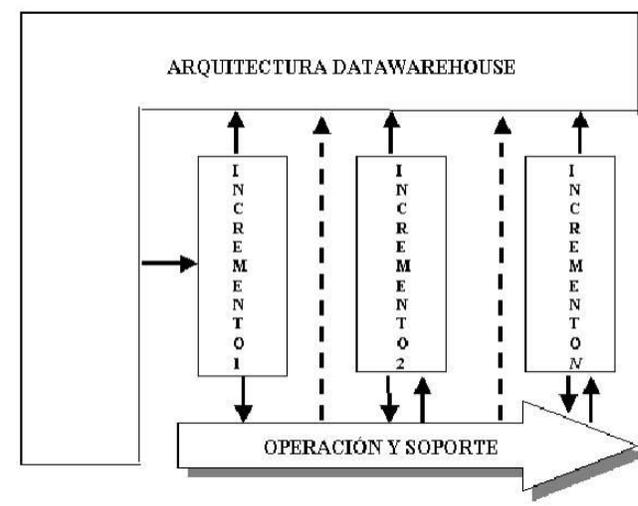


Figura 1 : Esquema de Construcción del DW[MicroSt96]

5.1.6. Metodología de desarrollo de DW

De acuerdo a la forma de implementación analizada, se deben considerar y asociar metodologías congruentes con el desarrollo de incrementos dirigidos a grupos específicos en las organizaciones.

Al acercarse a la implementación de un DW con un conjunto de proyectos pequeños, altamente enfocados para implementar partes DW, resulta natural pensar en una metodología incremental para abordar su desarrollo. Pero no debemos olvidar la integración de cada incremento a la arquitectura Datawarehouse, así entonces el desarrollo evolutivo resulta ser una aproximación práctica de construcción de un DW.

De esta manera, estos proyectos pueden aprovechar los beneficios de la implementación incremental, que incluyen la contención de riesgos, oportunidades para aprender a desarrollar datamarts, entrega frecuente y la minimización del impacto sobre la comunidad de usuarios, y a la vez pueden ser organizados en forma secuencial, paralela o en una combinación de estructuras en serie y paralelo.

En tanto, el desarrollo evolutivo implica que cada vez que un incremento sea entregado, se debe operar y desarrollar simultáneamente el DW. De esta forma se logra integrar cada incremento a la estructura final DW.

El costo del desarrollo evolutivo queda dado por el incremento en la frecuencia de los cambios y la necesidad intensificada de realizar soporte del DW.

5.1.7. Evolución del DW

5.1.7.1 Evolución de la Implementación

La aproximación del desarrollo incremental es por naturaleza evolutiva. El primer incremento libera un subconjunto del DW el cual satisface un conjunto limitado de necesidades de información. Con cada incremento que es agregado el DW se vuelve más completo, quedando habilitado para satisfacer un mayor conjunto de necesidades de información.

5.1.7.2 Evolución de la Arquitectura

El desarrollo incremental también ofrece una oportunidad para aprender y minimizar el impacto de cometer errores en el proceso de construcción; es poco probable que algún desarrollo de arquitectura de DW sea perfecto antes de construir su primer incremento. Ambas, las actividades de construcción de

incrementos y las actividades de operación del DW, proveen retroalimentación valiosa que ayuda a refinar la arquitectura.

5.1.8. Conclusiones y comentarios del DW.

En general, el factor que más incide en la necesidad de implementar un DW es la excesiva complejidad de los modelos relacionales que soportan los sistemas operacionales como lo podrías haber visto al analizar las fuentes de información necesarias para responder las preguntas de negocio clave, lo cual dificulta en extremo la extracción de información para llevar a cabo labores de gestión de negocios. Y también es precisamente esta complejidad en los sistemas operacionales el factor que hace más complejo y tedioso el desarrollo de un Datawarehouse.

En el caso de tener que decidir las herramientas de desarrollo, esto pasa no sólo por análisis de precio, soporte y las clásicas consideraciones de hardware y software disponibles. También son factores claves a considerar el objetivo final (¿se desea construir DataMart independientes o a futuro se espera llegar a una integración de ellos?) como la facilidad de uso que otorgue al usuario.

Con las características que poseen las herramientas de desarrollo la labor de construcción en el ámbito multidimensional se agiliza enormemente. El esfuerzo principal entonces está en el diseño del modelo a implementar. Un buen diseño no sólo optimiza el tiempo de respuesta y el espacio físico utilizado, sino que permite además un crecimiento en el análisis (es decir, consultas no consideradas al inicio) a través de fórmulas y cálculos definidos en tiempos de ejecución.

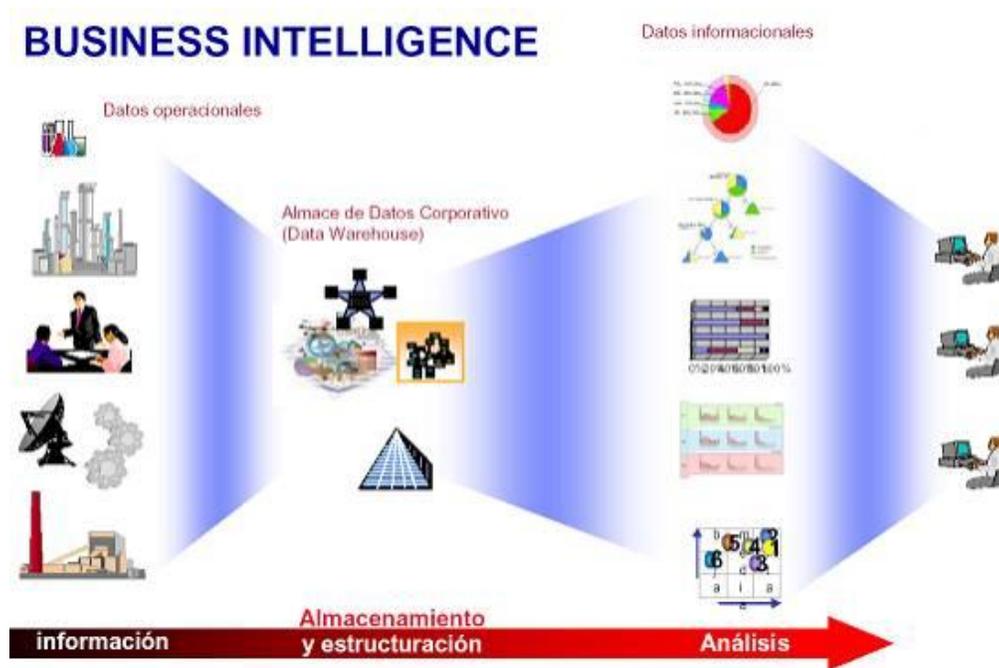
Es imposible evaluar con eficacia cualquier diseño de datamart hasta que se carguen los datos y se muestren a los usuarios. Diseñar las bases de datos para los datamarts es totalmente diferente del diseño para los sistemas

operacionales. El diseño de base de datos operacionales captura los requisitos del negocio y el proceso de diseño se considera un éxito si representa exactamente estos requisitos. En cambio para los datamarts, el proceso de diseño puede ser considerado acertado sólo si el diseño de base de datos refleja exactamente al usuario junto a sus requisitos procedurales, acomodando para ello datos de unos o más sistemas operacionales. La única forma de alcanzar una confiabilidad en el diseño del datamart es crear un ciclo de retroalimentación temprano en el proyecto. Este ciclo de retroalimentación utiliza a usuarios finales para evaluar los prototipos construidos, en un proceso iterativo que continúa a lo largo del proyecto. De esta manera el equipo de proyecto puede vigilar la aceptación de usuario y descubrir defectos de diseño en forma temprana.

Los sistemas DSS/EIS son limitados en el alcance de información que ellos proveen. El valor completo de un DW está dado cuando los sistemas DSS/EIS son usados en combinación con destrezas de acceso directamente al DW para encontrar sus necesidades especiales de información.

Finalmente, no se puede dejar de mencionar el gran aporte que una tecnología como esta significa para los usuarios que son, al fin y al cabo quienes tienen bajo su responsabilidad la labor de gestión de la empresa, decidiendo no sólo el futuro de su empresa sino también del propio desarrollo del país y porque no decirlo, de la humanidad.

Es posible prever la implementación de esta tecnología a nivel masivo por parte de las grandes empresas. El hecho de hacer hincapié respecto a las grandes empresas se debe básicamente a dos factores: debido al alto costo de inversión que tiene esta tecnología y por otra parte, debido al volumen de datos alcanzado para estas empresas que resultan necesarios para desempeñar las labores de gestión.



5.2 Extracción de la Información

5.2.1 Antecedentes.

Los datos se encuentran en su mayoría en la organización, normalmente estos están aislados en diferentes fuentes y formatos, en modelos estructurados o simples archivos.

Estos datos deben ser extraídos, integrados y sincronizados permanentemente en el modelo de Inteligencia de Negocio: el “Datawarehouse” o almacén de conocimiento organizacional. El proceso de creación y enriquecimiento del “Datawarehouse” se conoce como “Datawarehousing”.

Este proceso obtiene de las diferentes fuentes, a través de reglas de actualización, validación y seguridad; la información base del modelo. Puede realizarse de muchas formas y niveles de complejidad según el proyecto o fuente de datos: *acceso directo o ETL*.

El ETL es un lenguaje para el entendimiento y sincronización de datos entre diferentes plataformas tecnológicas, el cual respeta los esquemas de seguridad de la fuente e interpreta las reglas de acceso necesarias.

El ETL alimenta los Datamart según las reglas de vigencia y el diseño establecido por el equipo.

Un subconjunto del DataWarehouse, que involucre datos y reglas de negocio específicos para cada área o proyecto de Inteligencia de Negocio se denomina “Datamart”. Cada área, por lo tanto tiene un “Datamart” que constituye su base de datos particular.

Sobre los Datamart se realizan consultas y reportes resumidos y consolidados de acuerdo a las reglas de integración, simplificando muchos procesos de detalle, consolidación y enriquecimiento de los datos en el origen. Sin embargo, el “Datawarehouse” o sus Datamart, en si no son el objetivo final de los proyectos de Inteligencia de Negocio, constituye una herramienta de soporte a las aplicaciones en este sentido.

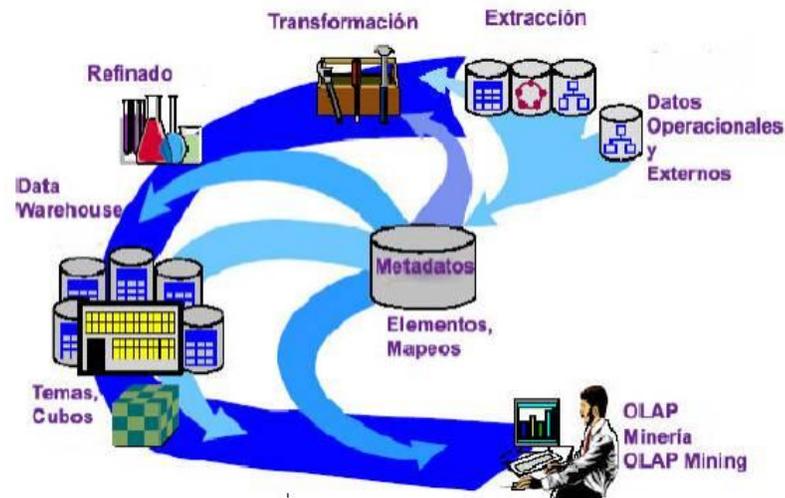
5.2.3. ¿Qué es un Datamart?

- Es una colección de datos a la medida de las necesidades para toma de decisiones de un departamento particular en una institución.
- Es un subconjunto de un datawarehouse que ha sido hecho a la medida para satisfacer las necesidades de alguna área.
- Alternativamente, un datamart puede ser un recurso compartido por varios departamentos donde existen necesidades analíticas (por ejemplo: Análisis de rentabilidad).

5.2.4. Metadatos.

Un metadato es un almacén de atributos de los datos o datos de los datos.

Metadatos son los datos que describen el origen, significado, derivación, etc. de los datos operacionales.



5.2.5 Definición de ETL.

Las herramientas de extracción, transformación y carga de datos (ETL) son la base de cualquier estrategia de una organización que quiera descubrir el valor oculto de sus recursos en forma de información. Poder acceder al gran volumen de información en bruto que hoy crean la mayoría de las organizaciones no es una tarea nada fácil. A medida que las organizaciones crecen, éstas adquieren diferentes aplicaciones, sistemas operativos, plataformas de hardware y bases de datos. Distribuidas a través de diferentes departamentos, divisiones o líneas de negocio, estas “islas” de información van quedándose más aisladas y haciéndose más inutilizables por el resto de la empresa.

Las herramientas ETL, también llamadas herramientas de integración, trabajan para traspasar información desde diversos sistemas hacia un datawarehouse central o un grupo de data marts de tal manera que se pueda distribuir por toda la empresa una única versión de la información. Las

herramientas ETL apoyan también el intercambio de información entre sistemas empresariales, ayudando a crear organizaciones en tiempo real, un entorno que maximiza la agilidad de la organización y en el que la información más actualizada está disponible para aquellos que más la necesitan.

Además de una mayor demanda de información actualizada, relevante y detallada, el crecimiento del tamaño de los datawarehouse está aumentando la necesidad por parte de las organizaciones de instalar herramientas ETL con mayores capacidades. Las bases de datos de varios terabytes de tamaño son ya moneda común. Mientras el volumen de la información empresarial aumenta y sus fuentes se multiplican, mientras las “batch windows” disminuyen o incluso desaparecen en un mundo de acceso las 24 horas al día y los siete días de la semana, las organizaciones buscan soluciones de integración que sean capaces de consolidar rápidamente toda su información en bruto y convertirla en datos útiles que puedan ser utilizados para el análisis, la toma de decisiones o en la demanda de información por parte de usuarios finales.

5.2.6 Una solución universal para la integración de información.

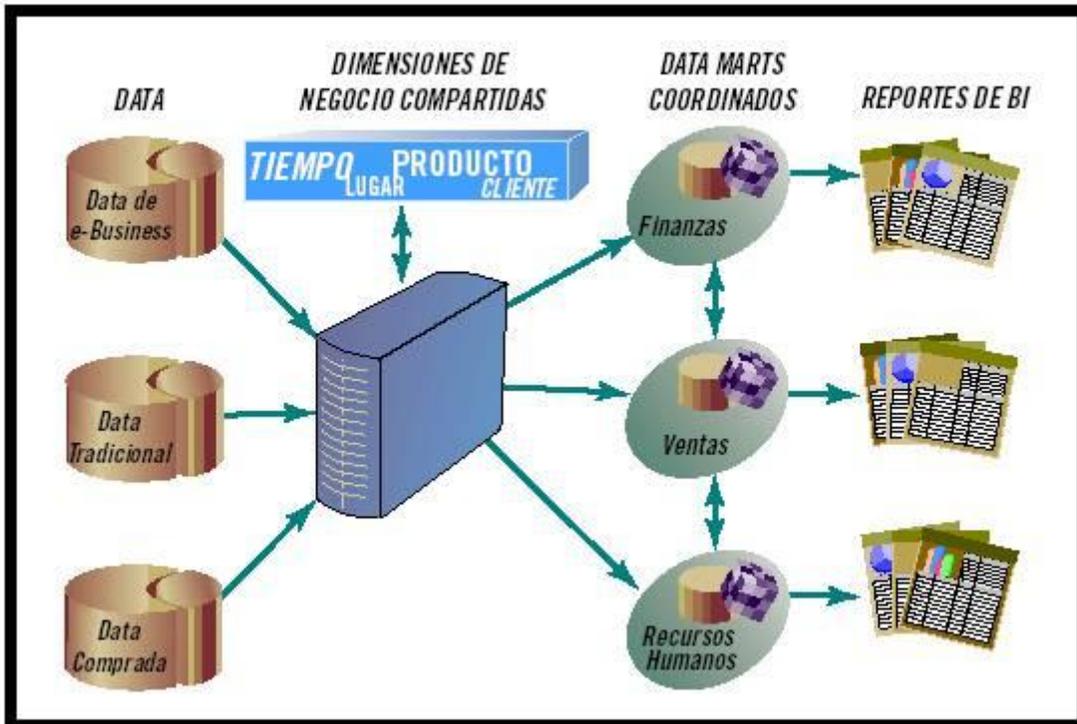
La gran mayoría de las herramientas ETL que se comercializan incluyen una suite de herramientas que pueden gestionar información a través de toda la cadena de suministro de información. Esto supone la extracción de datos desde una variedad de sistemas y de aplicaciones, su transformación, su estandarización si es necesaria, su transporte allí donde sea menester, y su carga en tablas de bases de datos en los formatos deseados. Casi todas pueden llevar a cabo todos los intercambios de información críticos independientes del formato, sintaxis, fuente o destino de esta información, desde soporte al XML hasta conectividad de mainframe, desde bases de datos relacionales hasta el proceso analítico on-line multidimensional sin necesidad de programación. Las organizaciones pueden por tanto seguir utilizando los entornos heterogéneos de tecnologías de la información

que ya tienen instalados sin desperdiciar recursos, tiempo y dinero en “estandarizar” sus infraestructuras.

Las herramientas de ETL constan de un poderoso motor que controla el flujo de información desde varias fuentes en una arquitectura de intercambio de datos, eliminando la necesidad de personalización, de procesos de intercambio de datos de punto a punto, y de programación manual de las transformaciones de los datos. Además, maximizan el intercambio de datos mediante diferentes modalidades, aprovechando las capacidades de transformación de bases de datos nativas para llevar a cabo las tareas de integración de información. ¿El resultado? minimizan el tráfico de la red, reducen los cuellos de botella y aceleran los procesos de toma de decisiones mediante el envío de información personalizada a aquellos que la necesitan en el momento que la necesitan.

Son también una herramienta para la gestión de metadatos totalmente integrada en la suite, asegurando la coherencia e integridad de los datos a través de toda la organización. La herramienta recoge, organiza y documenta la información tanto estructurada como no estructurada para optimizar su implementación y su utilización a través de la organización.

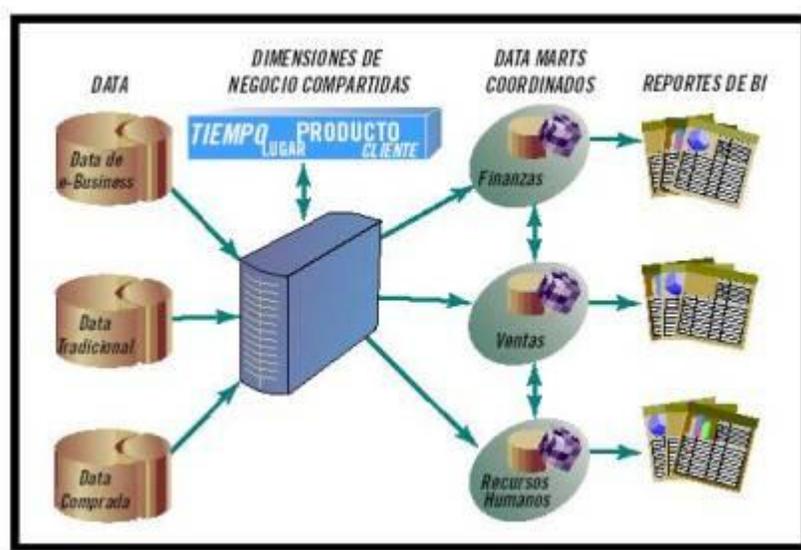
Por último pueden ofrecer una conectividad transparente con las herramientas existentes de BI mencionadas en el punto 3.6. Dichos sistemas BI cuentan ya con una herramienta ETL que facilita su implementación.



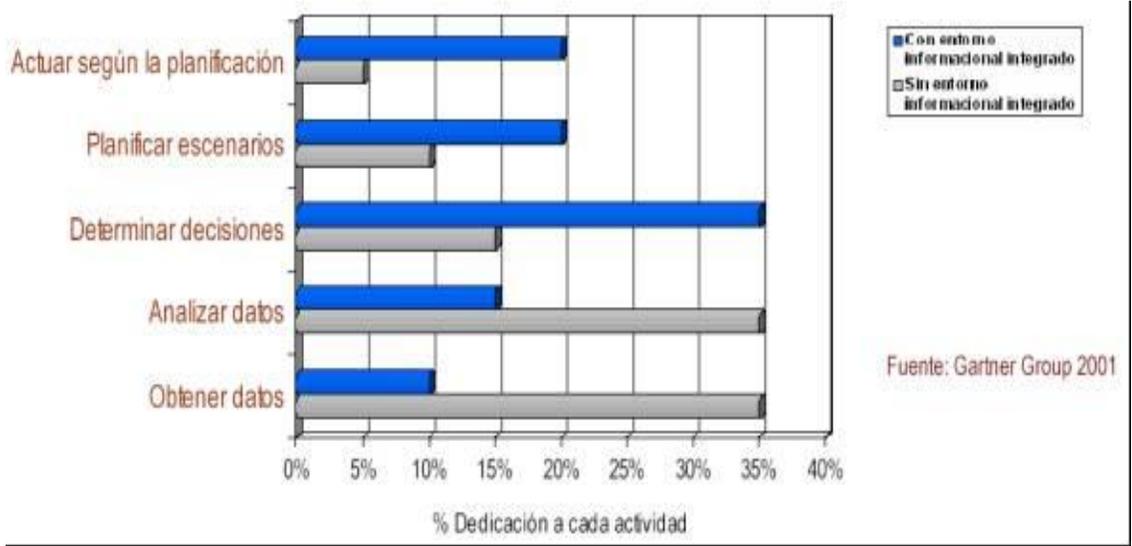
5.3 Transformación de la Información en BI.

5.3.1 Introducción.

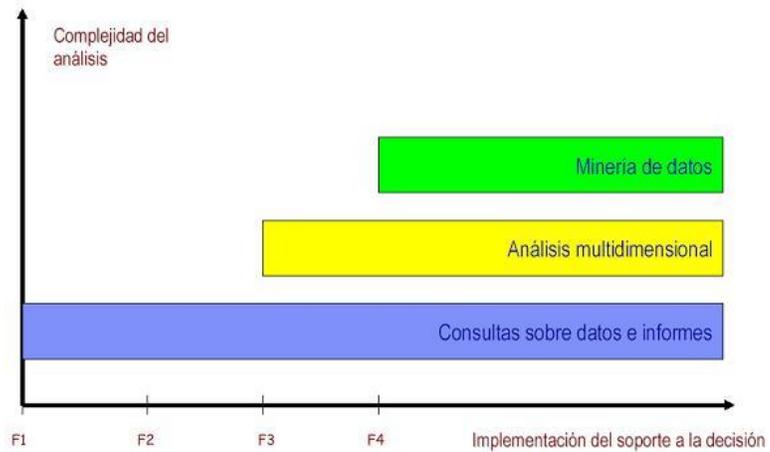
Se necesita una plataforma consolidada para desarrollar aplicaciones analíticas de valor para el negocio.



Desde el punto de vista estratégico, la implementación de un entorno informacional beneficia a todos los aspectos relacionados con la gestión de datos y toma de decisiones.



Básicamente existen 3 aproximaciones para la explotación de la información, todas ellas complementarias y con objetivos distintos.



5.3.2 Minería de Datos.

Es la explotación de grandes volúmenes de datos con el fin de extraer información valiosa para el negocio y previamente conocida. El Data Mining es un término usado para describir el conocimiento descubierto en bases de datos, la extracción del conocimiento, la arqueología de los datos, la exploración de los datos, el proceso de modelo de los datos, la información generada y el software.



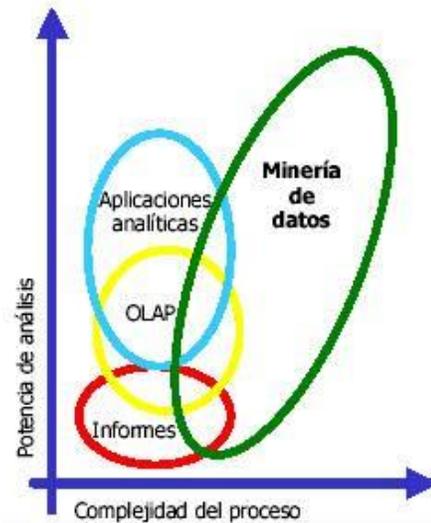
- ❖ Transacciones
- ❖ Sociodemográficos
- ❖ Respuesta a campañas
- ❖ Compra de productos
- ❖ Hábitos de comportamiento
- ❖ Clasificaciones externas
- ❖ Segmentación de Clientes
- ❖ Afinidad entre productos
- ❖ Patrones secuenciales de compra
- ❖ Modelos de predicción
- ❖ Rendimiento de campañas
- ❖ Tratamiento diferenciado a clientes
- ❖ Venta cruzada de productos
- ❖ Efectividad de promociones
- ❖ Incremento de cuota de mercado
- ❖ Retención de Clientes
- ❖ Recuperación de Clientes

5.3.2.1. Qué es y qué no es la minería de datos.

La minería de datos puede definirse como la extracción no trivial de información implícita, previamente desconocida y potencialmente útil, a partir de los datos. Para conseguirlo hace uso de diferentes tecnologías que resuelven problemas típicos de agrupamiento automático, clasificación, asociación de atributos y detección de patrones secuenciales. La minería de datos es, en principio, una fase dentro de un proceso global denominado descubrimiento de conocimiento en

bases de datos (Knowledge Discovery in Databases o KDD), aunque finalmente haya adquirido el significado de todo el proceso en lugar de la fase de extracción de conocimiento.

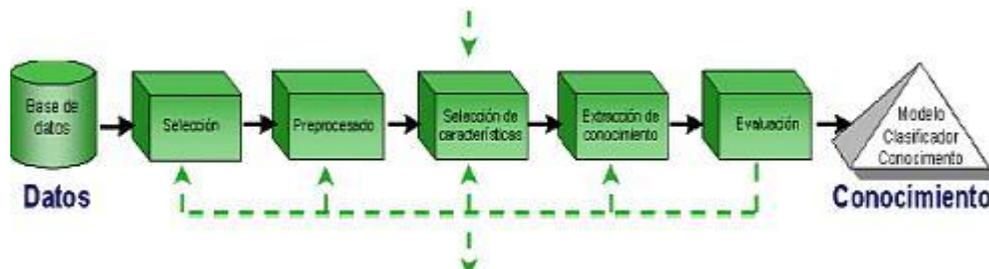
Es habitual que los expertos en estadística confundan la minería de datos con un análisis estadístico de éstos (afirmaciones de este tipo pueden encontrarse en documentación de empresas dedicadas al procesamiento estadístico que venden sus productos como herramientas de minería de datos).



En el caso de la minería de datos el proceso es muy distinto: la consulta que se realiza a la base de datos (al Data Warehouse) busca relaciones entre parejas de productos que son adquiridos por una misma persona en una misma compra. De esa información, el sistema deduce, junto a otras muchas, la afirmación anterior. Como podemos ver, en este proceso se realiza un acto de descubrimiento de conocimiento real, puesto que no es necesario ni siquiera sospechar la existencia de una relación entre estos dos productos para encontrarla.

5.3.2.2. Fases de un Proyecto de Minería de Datos

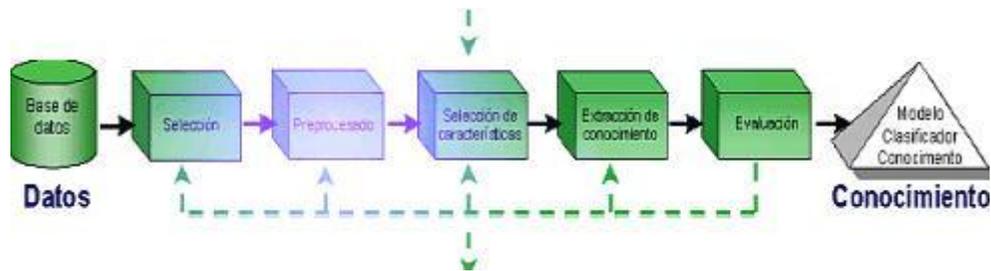
Los pasos a seguir para la realización de un proyecto de minería de datos son siempre los mismos, independientemente de la técnica específica de extracción de conocimiento usada.



El proceso de minería de datos pasa por las siguientes fases:

- ❖ Filtrado de datos
- ❖ Selección de Variables
- ❖ Extracción de Conocimiento
- ❖ Interpretación y Evaluación

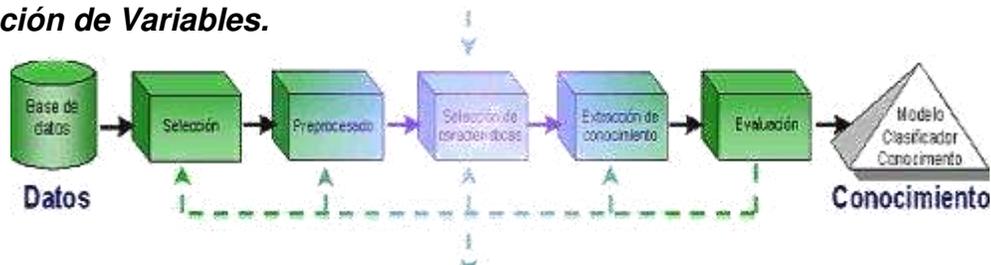
Filtrado de datos.



El formato de los datos contenidos en la fuente de datos (base de datos, DataWarehouse...) nunca es el idóneo, y la mayoría de las veces no es posible ni siquiera utilizar ningún algoritmo de minería sobre los datos "en bruto".

Mediante el preprocesado, se filtran los datos (de forma que se eliminan valores incorrectos, no válidos, desconocidos... según las necesidades y el algoritmo a usar), se obtienen muestras de los mismos (en busca de una mayor velocidad de respuesta del proceso), o se reducen el número de valores posibles (mediante redondeo, clustering,...).

Selección de Variables.

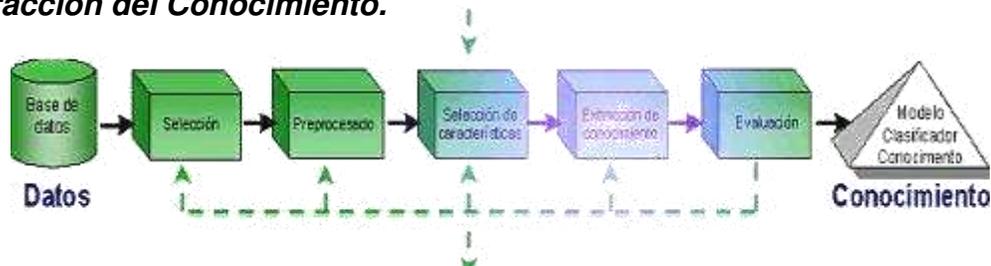


Aún después de haber sido preprocesados, en la mayoría de los casos se tiene una cantidad ingente de datos. La selección de características reduce el tamaño de los datos eligiendo las variables más influyentes en el problema, sin apenas sacrificar la calidad del modelo de conocimiento obtenido del proceso de minería.

Los métodos para la selección de características son básicamente dos:

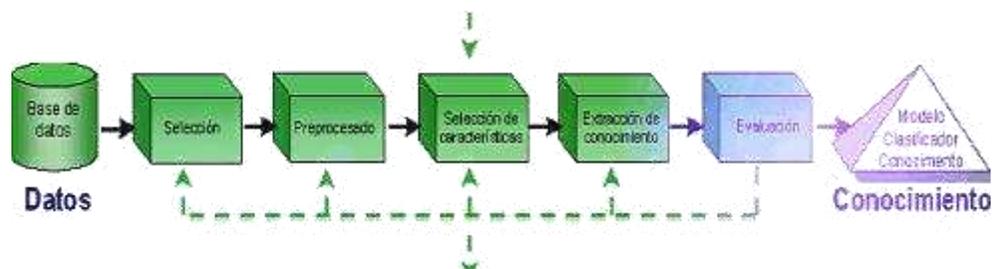
- ❖ Aquellos basados en la elección de los mejores atributos del problema,
- ❖ Y aquellos que buscan variables independientes mediante tests de sensibilidad, algoritmos de distancia o heurísticos.

Extracción del Conocimiento.



Mediante una técnica de minería de datos, se obtiene un modelo de conocimiento, que representa patrones de comportamiento observados en los valores de las variables del problema o relaciones de asociación entre dichas variables. También pueden usarse varias técnicas a la vez para generar distintos modelos, aunque generalmente cada técnica obliga a un preprocesado diferente de los datos.

Interpretación y Evaluación.



Una vez obtenido el modelo, se debe proceder a su validación, comprobando que las conclusiones que arroja son válidas y suficientemente satisfactorias. En el caso de haber obtenido varios modelos mediante el uso de distintas técnicas, se deben comparar los modelos en busca de aquel que se ajuste mejor al problema. Si ninguno de los modelos alcanza los resultados esperados, debe alterarse alguno de los pasos anteriores para generar nuevos modelos.

5.3.3 Análisis Multidimensional.

OLAP (On Line Analytical Process) es una forma de acceder a la información mediante dimensiones, de forma que un mismo dato puede ser analizado bajo distintos puntos de vista.



Los datos corporativos son complejos. Para aprender de ellos, necesitamos herramientas de análisis que puedan darle sentido a la información. Procesamiento Analítico en Línea (OLAP) es un medio poderoso para extraer el máximo valor de los datos corporativos a través de análisis multidimensional.

Nosotros pensamos de forma natural acerca de nuestros negocios en términos de dimensiones, tales como periodos de tiempos, productos, clientes y ubicaciones. Igualmente le damos seguimiento a medidas tales como ingresos, rentabilidad, retención de clientes. Al organizar información dentro de dimensiones y medidas, OLAP nos permite medir tendencias reales de clientes, resaltar anomalías a través de productos; comparar ventas anuales en una región por producto o por tipo de cliente; incluso probar un plan estratégico.

De esta manera tenemos:

Soluciones Analíticas por áreas funcionales

- ❖ Exploración de los datos en términos de negocio.
- ❖ Análisis de la misma información bajo distintos puntos de vista.

Exploración de los datos según necesidades

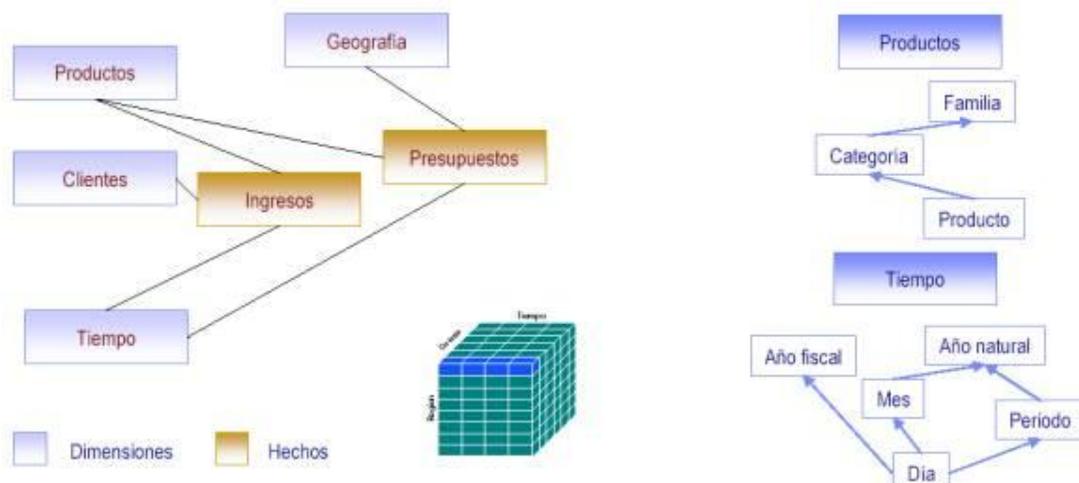
- ❖ Comprensión del negocio mediante: informes (¿Qué ha pasado?), planificación (¿Qué pasaría si?) y predicción (¿Qué vendrá después?).

Acceso a la información consolidado y bajo distintas formas

- ❖ Todos los usuarios acceden a la misma información, y en tiempo real.
- ❖ Flexible para el análisis sin límite en el número de dimensiones o agregaciones.

5.3.3.1 Hechos y dimensiones.

El análisis multidimensional permite la exploración de los datos a través de varias dimensiones, compuestas a su vez por miembros, por las que se requiere analizar los hechos del negocio:



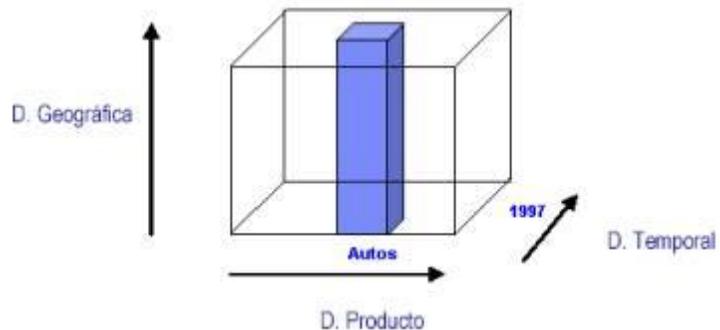


5.3.3.2 Múltiples vistas de negocio.

El análisis de datos, a lo largo de varias dimensiones (cubo) y a través de distintas categorías, permite detectar comportamientos y reglas de negocio difíciles de cubrir.

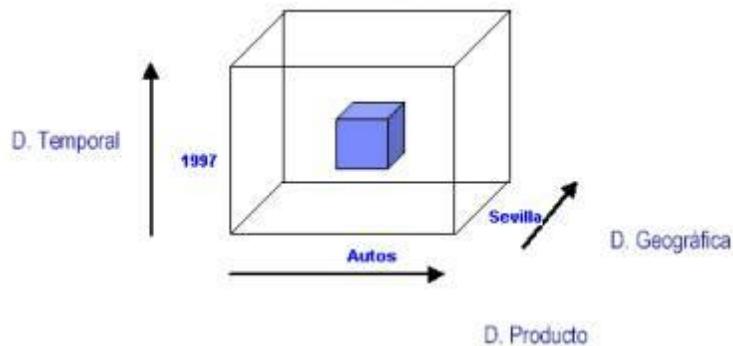
Sección del cubo

- Vista específica de un producto a lo largo del tiempo y en relación a las distintas geografías

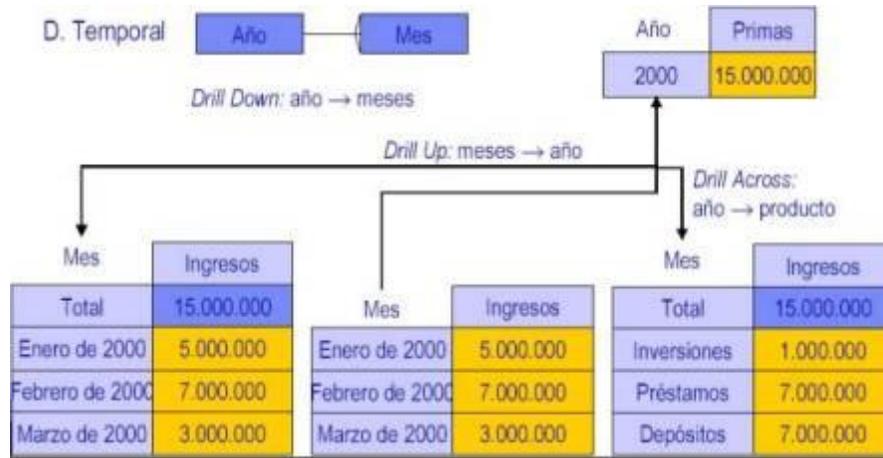


Pivotado del cubo

- El cubo puede ser rotado, obteniéndose otra vista representativa de los datos
- Supone un arreglo de las dimensiones en el resultado



El análisis multidimensional, permite consolidar y agregar información a través de la jerarquía de miembros de cada una de las dimensiones, facilitando la navegación sobre los datos.



Mediante operaciones de corte y rotación (slice & dice), es posible modificar en tiempo real la vista del cubo, cambiando la perspectiva de la información en cada momento.



5.3.3.3. ¿Cuándo es necesario OLAP?

- ❖ Las consultas de usuario sobrecargan al servidor de datos departamental.

- ❖ Los informes tardan mucho en ejecutarse.
- ❖ Los usuarios emplean mucho mas tiempo en adquirir y transformar los datos y no queda tiempo para el análisis, siendo este además difícil de implementar.
- ❖ Los medios de consulta son poco flexibles para soportar cambios. Los usuarios reclaman capacidades de análisis dinámicas, no solo estáticas.
- ❖ Existe problemas de unificación y actualización de las fuentes de información para todos los usuarios y analistas de negocio. No todos los usuarios pueden acceder a los datos.
- ❖ Es necesario un sistema que sea capaz de modelizar el impacto de un precio en los ingresos y beneficios, permitiendo plantear múltiples escenarios de simulación y cálculos.
- ❖ Es necesario un mecanismo que soporte múltiples herramientas de usuario final.

5.4. Generación de informes.

Actualmente, un gran número de organizaciones recurren a soluciones múltiples para responder a sus necesidades en materia de elaboración de informes. Implementadas generalmente a nivel departamental, estas soluciones debe ser utilizadas simultáneamente y deben ofrecer una sola versión de la verdad.

Paralelamente en los esfuerzos por recortar los costos y la complejidad de la estructura, los responsables de los servicios informáticos buscan una solución única para la elaboración de informes, basados en web y capaz de responder a todas sus demandas. Se requiere de una solución flexible que pueda funcionar en la infraestructura informática ya existente y que sea evolutiva.

Englobar todas las exigencias de elaboración de informes en una única solución constituye una verdadera revolución en el universo del software del reporting.

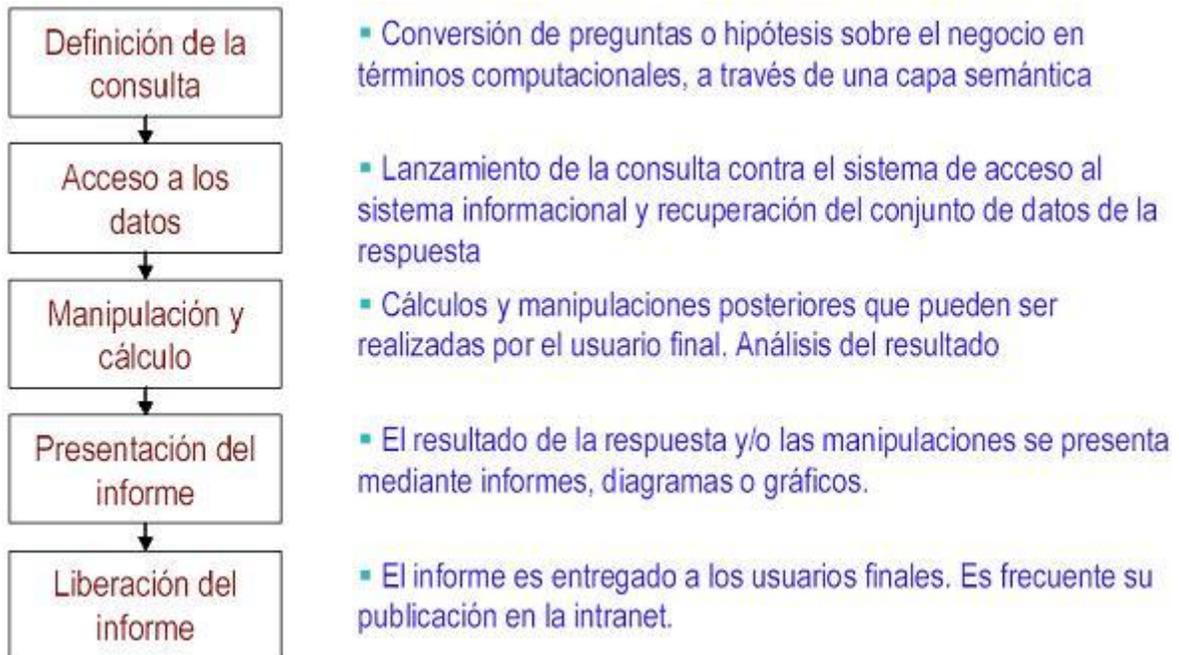
5.4.1 Informes completos y decisionales

La herramienta de reporte deberá responder a todas las necesidades de creación de informes, tales como las consultas compuestas (aquellas que buscan información sobre varias bases de datos diferentes), consultas personalizadas, generación de informes predefinidos con destino a un gran número de usuarios, informes específicos de actividad y, asimismo, informes de producción por lotes.

Una interfaz web 100% suprime las operaciones de instalación y mantenimiento de las aplicaciones cliente y aligera la carga de los servicios informáticos. Los informes deberían poder generarse fácilmente con sólo seleccionar y arrastrar los elementos y organizarlos a partir de una plantilla. El entorno de creación aligera considerablemente los procesos tediosos y rígidos de los sistemas convencionales.



5.4.2. Etapas en la creación de informes



Existen 3 aproximaciones básicas.

Consultas empaquetadas

- ❖ Consultas e informes predefinidos.
- ❖ Ejecuciones periódicas.
- ❖ Carga de trabajos estables.

Consultas Ad-Hoc.

- ❖ Consultas recurrentes y aleatorias.
- ❖ La carga de trabajo es variables.

Consultas de Usuario Final.

- ❖ Carga de trabajo variables.

Capítulo 6

Conclusiones y trabajo futuro

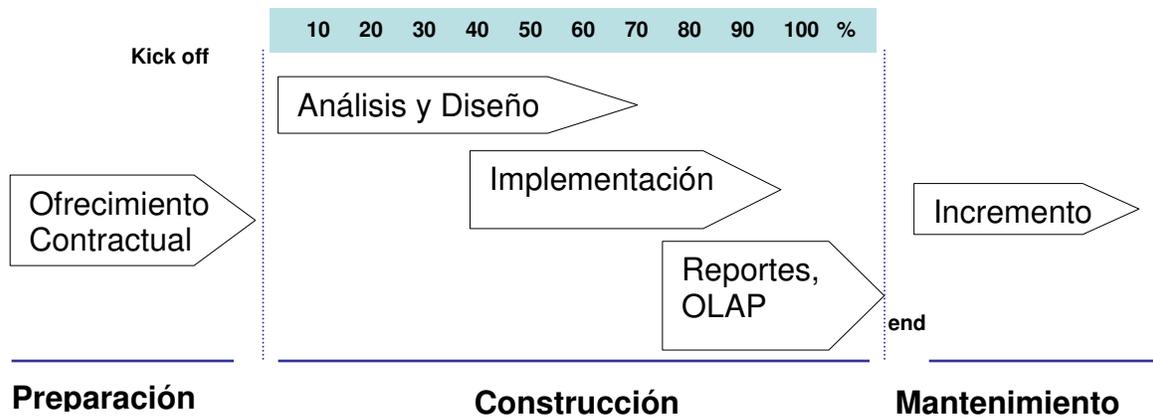
6.1. Conclusiones.

Muchas empresas utilizan el sentido común y la intuición para tomar decisiones, la información que se traduce en conocimientos acerca la visión interna a la realidad y esa diferencia existente es la que puede representar miles o millones de pesos. Lo que se pretende es acercar el mundo real a la visión interna para generar ganancias, para convertir la información en utilidades, para darle un valor a la información.

Inteligencia de negocios es información sobre eventos pasados y sus tendencias con el propósito de realizar acertados procesos de toma de decisiones en el futuro. La mayor parte de las organizaciones tienen toda la información necesaria en sus bases de datos operacionales para realizar decisiones de negocio inteligentes. Sin embargo, estos datos no están organizados de tal forma que puedan ser consultados de forma eficiente para obtener indicadores, estadísticas y tendencias con diferentes criterios. Este trabajo ofrece una metodología para ayudar a las empresas en su proceso de Inteligencia de Negocio.

6.1.1. La receta

Tres macro tareas en general nos permiten desarrollar BI. Hay una tarea de preparación, una de construcción y por último una de mantenimiento.



13 Pasos a seguir.

Preparación

1. **Defina al equipo de trabajo.** Del equipo de trabajo depende el éxito del proyecto.
2. **Identifique a los beneficiarios.** Quien puede ganar mas con la implementación de BI? Quien esta ya trabajando con bases de datos o emplea metodologías estadísticas
3. **Defina Metas.** Las decisiones de BI deben ser basada en medidas cuantificables
4. **Examine la Situación Actual.** Cual es el nivel de satisfacción de los usuarios con la información disponible?
5. **Preguntas Estratégicas de Negocio que el BI va a servir para aclarar.**
6. **Identifique Posibles Fuentes de Información.** Es suficiente las fuentes que tengo para responder a las preguntas de negocio planteadas?

Construcción-Análisis y Diseño

7. **Análisis de los Requerimientos.** Casi la mitad del tiempo del proyecto debe ser adjudicado para analizar requerimientos del negocio y diseñando la solución tecnológica.
8. **Evaluar la Calidad de los Datos.** Con frecuencia la calidad de los sistemas es sobreestimada. Requiere de un análisis detallado.

9. **Modelo de Datos del Negocio.** Crear la representación lógica de los datos a través de un modelo que represente los objetos y sus relaciones.

Construcción-Implementación

10. **Modelo Físico de Datos.** Es la aplicación del Modelo Lógico de Datos en sistema de base de datos que se debe optimizar de acuerdo a la capacidad del motor de base de datos seleccionado.
11. **Extracción de la información.** El análisis de procesos de negocio y fuentes operativas de información, nos permite obtener reglas para la extracción de información de los distintos sistemas. La información extraída debe encajar en el modelo de BI. Por último la información deberá ser cargada y establecidos los periodos de frecuencia de las cargas.
12. **Transformar la información en BI.** Montar las herramientas OLAP, diseñar los tipos de reportes, capacitar a usuarios finales.

Mantenimiento

13. **Incrementos.** Crecer el BI con los nuevos requerimientos del negocio y mantener los ya operantes.

6.2. Trabajos futuros.

Hablar de BI es un tema muy extenso por lo que se puede trabajar en proyectos futuros en profundizar los temas de:

- DatawareHouse
- Herramientas ETL
- Modelado de Datos
- Evaluación de la Calidad de los Datos.
- Herramientas OLAP
- Software BI

Referencias Bibliográficas

Abukari, Kobana; Job, Vigía. (2003). "Business Intelligence in action". Proquest. CMA Management, (Mayo 07, 2003).

AlQasem, Iyas (2003). Moving onto a Business Footing. Published May 16, 2003. [WWW] <http://www.datawarehouse.com/article/?articleId=3154&searchTerm=Business%20Intelligence%20&%20Sales%20Benefits>

Aylmer V. Nichols., "SISTEMA MODERNO DE PROCESAMIENTO DE DATOS"., Editorial Limusa., Primera Edición., p. 381

Betts, Mitch. "The future of business intelligence". Computerworld. Tupson Technologies. <http://www.tupson.com/busintel.htm>

Buytendijk, F., Dresner, H. J., Linden, A., Tiedrich A., Hostmann B., Herschel, G., Gassman, B. (2004). Hype Cycle for Business Intelligence, 2004, Gartner [on-line database]

Burch Jhon G. (1993) "Diseño de sistemas de información"., Editorial Limusa., p. 985

Cognos (2005), Cognos posicionado como líder en el último estudio de Gartner sobre el mercado del Business Intelligence, <http://www.cognos.com/es/pressfolder/2004/2712.html>

Gelbart, Frank. BI Trends (2003). Looking for New Ways to Cut Costs and Boost Revenues?. April 4, 2003. [WWW]. http://www.dmreview.com/article_sub.cfm?articleId=6574.

Churchman C. West., "EL ENFOQUE DE SISTEMAS"., Editorial Diana., Primera Edición., p. 270

Dresner, H.J., Hostmann B., Buytendijk F. (Noviembre, 2004). Magic Quadrants for Enterprise BI Suites and Platforms, Gartner [on-line database]

Dresner H.J., Buytendijk F., Friedman T. (Noviembre, 2003), Predicts 2004: Business Intelligence Markets, Gartner [on-line database]

Elliott, T.(2004) Implementing Business Intelligence Standards. Business Objects. [electronic text], Febrero 2004.

Elmasri, R., Navathe S. (2000). Fundamentals of Database Systems, Addison Wesley. Third Edition.

Fajardo, F.; Crispino I.(2004). “Implementación de una Plataforma para Análisis de Calidad de Datos”. Informe de proyecto de grado, Facultad de Ingeniería, Universidad de la República, Uruguay, 2004.

Friedman, T. (2004). Management Update: The Cornerstones of Business Intelligence Excellence. Gartner [on-line database]

Graham, C. (Marzo, 2005). BI Software Becoming a Top Priority for CIOs, Gartner [on-line database]

Guarente, Matthew.(2002) HOW is your business doing?. Sunday Business. London (UK): pg. 1 [electronic journal]. Jan 20, 2002.

Havenstein, H. (2005). Users turn to operational business intelligence tools: Emerging capability promises real-time access to most data,
<http://www.computerworld.com/databasetopics/businessintelligence/story/0,10801,100506,00.html?SKC=businessintelligence-100506>

InformationBuilders (2004).Business Intelligence Goes Operational: On the Front Lines Making a Bottom-Line Difference,
http://www.informationbuilders.com/solutions/operational_bi.html

Kaplan, Robert y Norton (2000), David, Cuadro de Mando integral, Ediciones Gestión 2000, 1997.

Kimball R. (2003), " Build your DataWarehouse one piece at a time" , Datawarehouse Designer, Octubre 2002. Extraído el 4 de marzo del 2003 de <http://www.iemagazine.com/021030>

Kimball R. (2003), "Is Data Staging Relational?" DBMS, April 1998 . Extraído el 26 de febrero de 2003 de <http://www.dbmsmag.com>

Kimball R. (1998), Laura Reeves, Margy Ross & Warren Thornthwaite, The [Data Warehouse](#) Life Cycle Toolkit, [New York](#) : John Wiley & Sons, 1998.

Laudon, Kenneth C.; Laudon, Jane P. (2004) Management Information Systems. Editorial Pearson Prentice Hall, New Jersey U.S.A.
Lokken, B. (2001), Business Intelligence: An Intelligent Move or Not?,
<http://bi.ittoolbox.com/browse.asp?c=BIPeerPublishing&r=%2Fpub%2FAO031202%2Epdf>.

Martínez, J. MicroStrategy (2002) ¿Qué es la Inteligencia de Negocios y qué beneficios ofrece?, [electronic text], MicroStrategy México. 2002.

Mendelzon, A. y Ale, Juan (1999). Introducción a las bases de datos relacionales. Prentice Hall.

Mika Hannula, Virpi Pirttimaki.(2003) “Business intelligence empirical study on the top 50 Finish companies”. Journal of American Academy of Business, Cambridge [electronic journal] Vol. 2, Iss. 2; pg. 593, 7 pgs Mar 2003.

McNamara, B. (1990). An Appraisal of Executive Information and Decision Support Systems. Journal of Systems Management [electronic journal]; Mayo 1990

Moss, L.T., Atre, S. (2003) .Business Intelligence Roadmap: The Complete project Lifecycle for decision support applications, Editorial.Addison-Wesley, EUA.

Naumann, F.; Leser, U.(1999): “Quality-driven Integration of Heterogeneous Information Systems”. En anales de 25th International Conference on Very Large Databases (VLDB’99), Scotland, 1999.

Pressman Roger S. ., "INGENIERIA DEL SOFTWARE"., Editorial McGrawHill.,

Quinn, K. (2003). Establishing a Culture of Measurement: A Practical Guide to BI, informationbuilders, www.informationbuilders.com/cgi-shell/products/whitepaper/whitepaper_form.pl?Whitepaper_Code=WHTBI_Philosophy

Peralta, V.; Ruggia, R.; Kedad, Z.; Bouzeghoub, M. (2004). “A Framework for Data Quality Evaluation in Data Integration Systems”. Aceptado para el 19º Simposio Brasileiro de Bases de Datos (SBBD’2004), Brasil, 2004.

Rob, P. y Coronel, C. (2000) Database System Design, implementation, and management. Course Technology.

SAS Institute (2004). Business Intelligence. Insurance and Technology. [electronic journal]. Vol. 29 Issue 1, p12, 2p. Published January 2004.

Seddon, Peter. Staples, Sandy. Patnayakuni, Ravi. Bowtell, Matthew (1999). Dimensions of information systems success. Communcation of the AIS. [electronic journal]. Volume2, Issue 3es (November 1999). Article 5.

Tiedrich Alan (Junio, 2003). Business Intelligence Tools: perspective, *Gartner* [on-line database]

Timo E., Rowland, S., Byrne, K., Surak B., (2003). Implementing a Business Intelligence Strategy: A Practical Guide to Business Intelligence Standardization, Bussines Objects, www.dmreview.com/whitepaper/WID537.pdf, (Consultada en mayo de 2005).

Teorey, T. (1990) Database Modeling and Design.Morgan Kauffman Pub.

Wang, R.; Strong, D. (1996) "Beyond accuracy: What data quality means to data consumers". Journal on Management of Information Systems, Vol. 12, 4:5-34, 1996.

Wittschen, Lee (2004). Why Business Intelligence?. January, 2004. [WWW] .
<http://www.businessintelligence.com/ex/asp/code.29/xe/article.htm>. June 2004

Wu, Jonathan (2001). The Value of Business Intelligence □ Applications: Part 1-3. Published August 13, 2001. [WWW].
http://www.dmreview.com/article_sub.cfm?articleId=3887

<http://www.microstrategy.com.mx>

<http://www.sybase.com.mx/>

<http://www.cognos.com>

Glosario

BI: Inteligencia de Negocios.

DW: Datawarehouse. Repositorio de información integrada de varias fuentes que sea fácilmente accesada por los usuarios.

OLAP: On-line Analytical Process (Proceso analítico en línea). Define el comportamiento de un sistema de análisis de datos y elaboración de información

Modelo: Proyecto o reproducción a escala.

ODS: Almacenes de datos operacionales.

data marts: Es una aplicación de un DATA WAREHOUSE construida rápidamente para soportar una línea de negocio simple. Los DATA MARTS, tienen las mismas características de integración, no volatilidad y orientación temática que el DW. Representan una estrategia de “divide y vencerás” para ámbitos muy genéricos de un DATA WAREHOUSE.

EIS: Sistemas de Información Ejecutiva.

DSS: Sistemas de Soporte de Decisiones.

DM: data mining. Minería de Datos.

RT: Herramientas de reporte.

CMI: Cuadro de Mando Integral.

BS: Balanced Scorecard. Balanceo de Indicadores de Operación.

OLTP: On-line Transaction Process (Proceso transaccional en línea).

KPIs: Indicadores de proceso clave.

DFM: Administrador de flujo de datos.

I.R.: Ingeniería de Requerimientos.

A.R.: Análisis de Requerimientos.

Query: Consulta.

SQL: Lenguaje estructurado .

4GL: Lenguaje de 4ª. generación.

Statu quo: Estado del momento actual.

CheckList: Lista de chequeo.

Sistema: es un conjunto integrado, real o abstracto, de componentes o partes que se interrelacionan.

DataBase: Base de datos.

Workflow: Flujo de trabajo.

KDD: Knowledge Discovery in Databases