

Sistemas de Información

Unidad 7
INTELIGENCIA de NEGOCIOS



ÍNDICE DE CONTENIDOS

BUSINESS INTELLIGENCE (BI)	1
CUADRO DE MANDO INTEGRAL	3
OBJETIVOS	5
UTILIDAD.....	5
TIPOS DE TABLEROS	5
PERSPECTIVAS	6
DATA WAREHOUSING Y DATA MINING.....	8
DATA WAREHOUSE (DW)	9
COSTOS VS. VALOR DE DW	10
COSTOS DE UN DW	11
VALOR DEL DW	12
DATA MINING O MINERÍA DE DATOS	12
ALCANCE	13
¿CÓMO TRABAJA EL DATA MINING?	13
BIG DATA.....	15
BIG DATA Y LA INTELIGENCIA DE NEGOCIOS	15
CARACTERÍSTICAS DE BIG DATA	16

BUSINESS INTELLIGENCE (BI)



En la actualidad, prácticamente todas las empresas poseen sistemas de información estructurados alrededor de bases de datos que almacenan datos sobre sus actividades y sus colaboradores (clientes, proveedores, etc...) es decir que disponen de una gran cantidad de datos históricos, fiables y rigurosos de todas las actividades realizadas.

Es lógico pensar que dichos datos podrían ser refinados, agrupados, tratados y analizados para intentar extraer información que permitiera ayudar en la toma de decisiones de la empresa.

Encontrar patrones de conducta en la compra de los clientes, presentar información en tiempo real sobre el rendimiento de las distintas sucursales de una empresa a su dirección, o identificar los clientes que no nos son rentables (el costo de su gestión es superior al beneficio que dejan) son ejemplos de ello.

La conversión de los datos operativos de las empresas en información que dé soporte a la toma de decisiones, es lo que se conoce como inteligencia de negocio o Business Intelligence.

El contexto de la sociedad de la información, esto es las tecnologías que permiten la generación, distribución y uso de información en cualquier actividad, ha puesto de manifiesto la necesidad de tener mejores, más rápidos y más eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de las actividades que generan valor en ella (cadena de valor).



La **inteligencia de negocios (BI)** responde a dicha necesidad ya que **es el conjunto de metodologías, aplicaciones y prácticas enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización.**

Las situaciones en las que resulta necesario un sistema de Business Intelligence pueden ser:

- La toma de decisiones se realiza de forma intuitiva en la organización.
- Identificación de problemas de calidad de información.
- Uso de herramientas de ofimática como repositorios de información corporativos o de usuario (Ej.: Excel).
- Necesidad de cruzar información de forma ágil entre departamentos.
- Evitar silos de información.
- Existe demasiada información para ser analizada de la forma habitual.
- Es necesario automatizar los procesos de extracción y distribución de información.

Los sistemas de Business Intelligence buscan responder a las preguntas:

- ¿Qué pasó?
- ¿Qué pasa ahora?
- ¿Por qué pasó?
- ¿Qué pasará?

Entre las tecnologías que forman parte de Business Intelligence, serán objeto de estudio el Cuadro de Mando Integral, el Data Warehouse y Data Mining (Minería de datos).



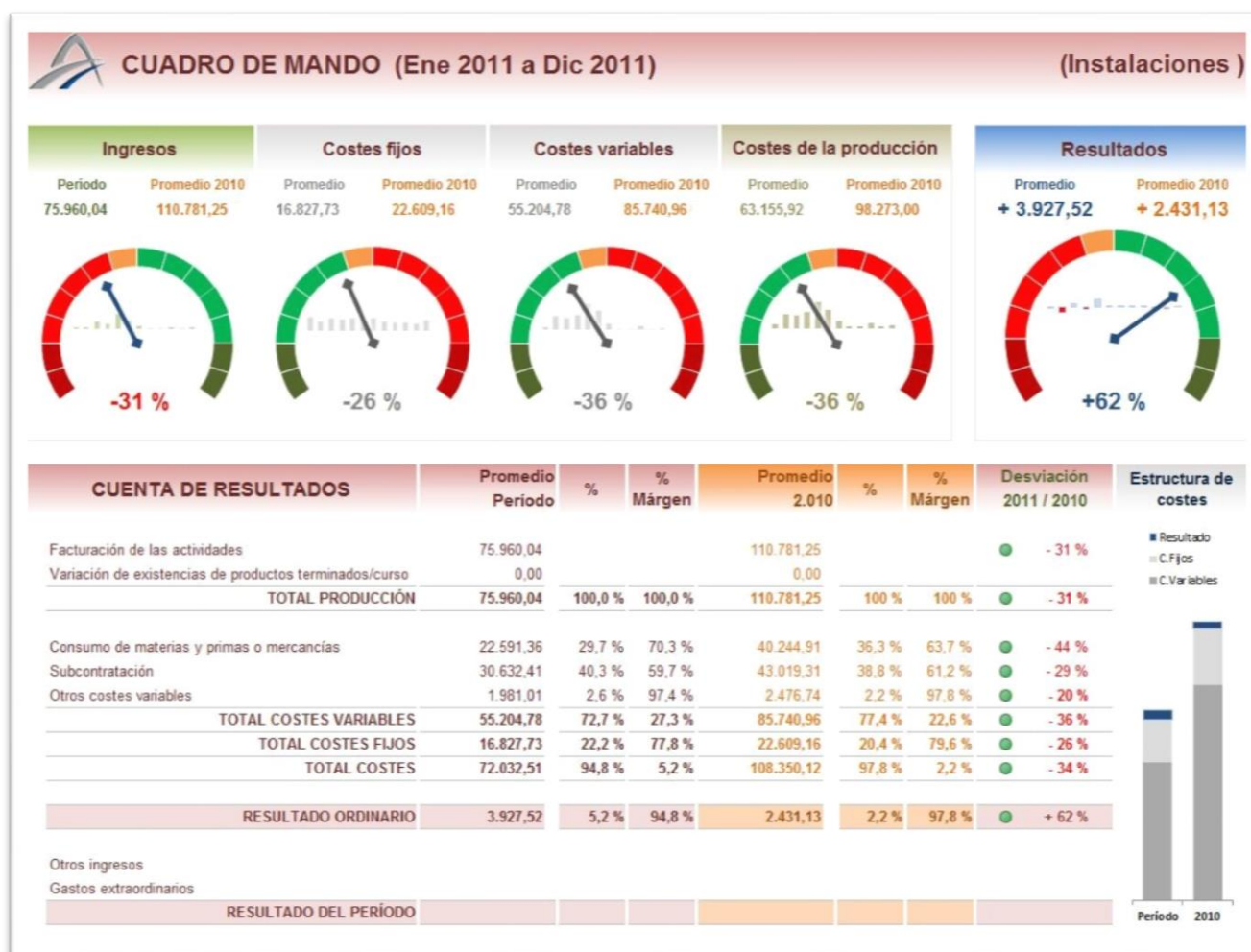
CUADRO DE MANDO INTEGRAL

El **cuadro de mando integral** también conocido como tablero de comando o tablero de control, es un modelo de gestión que traduce la estrategia institucional en objetivos relacionados entre sí, los cuales son medidos a través de indicadores y vinculados a planes de acción que permiten alinear el comportamiento de los miembros de la organización con la estrategia de la empresa.

Es una herramienta aplicable a cualquier organización y nivel de la misma, cuyo objetivo y utilidad principal es diagnosticar adecuadamente una situación. Para ello utiliza un conjunto de indicadores cuyo seguimiento y evaluación periódica permitirá contar con un mayor conocimiento de la situación de la empresa o sector apoyándose en tecnologías informáticas.

El diagnóstico y monitoreo permanente de determinados indicadores e información es la base para mantener un buen control de situación cualquiera sea la disciplina de que se trate.

El tablero permite a través del color de luces y alarmas ser el disparador para la toma de decisiones, por lo que se torna fundamental definir los indicadores a monitorear. Hay indicadores genéricos para todas las empresas especialmente en áreas como las económicas financieras, donde cada empresa o sector requiere definiciones de parámetros y de quién y cómo va a monitorear esa información.





El Tablero de Comando ayuda a la alta dirección a visualizar en tiempo real el avance en el logro de las metas prefijadas.



Logra el balance entre indicadores financieros y no financieros, de resultados y de proceso, entre el corto y el largo plazo, y un balance entre el entorno y el interior de la organización.

Es una herramienta estratégica porque cuenta como se implementarán las estrategias por medio de un mapa estratégico de enlaces causa-efecto. Se debe definir:

- **Reporte o Pantalla:** que muestre la información clave para el diagnóstico, con su formato y configuración. El reporte debe reflejar la necesidad de información de la empresa y debe ser estándar a distintos niveles de la organización.
- **Período del Indicador:** día, mes, acumulado del ejercicio, proyectado a fin del período fiscal o para los próximos meses, etc.
- **Apertura:** forma en la cual se podrá abrir y clasificar la información y sus sucesivos niveles de desagregación en tablas o matrices multidimensionales por producto, sector geográfico, dimensión de análisis, etc.
- **Frecuencia de actualización:** tiempo que transcurre entre distintas actualizaciones de los datos. Ejemplo: on-line, diaria, semanal, mensual.
- **Referencia:** base sobre la que se calculan las desviaciones (el mes anterior, el promedio de los últimos meses, el presupuesto inicial, un objetivo o una meta, etc).
- **Parámetros de alarmas:** niveles por encima o por debajo de los cuales el indicador es preocupante, por ejemplo más o menos 5% sobre una base de referencia.
- **Gráfico:** representación gráfica de la información (tortas, barras, líneas, etc).
- **Responsable de monitoreo:** es quien debe informar al nivel superior los valores no deseados de los indicadores.



OBJETIVOS

- Que la gerencia conozca el estado presente de la organización desde varias perspectivas.
- Conocer las tendencias del futuro, para poder alcanzar el éxito.
- Establecer un sistema de gestión estratégica.
- Conseguir el incremento de la productividad y de la eficiencia.
- Utilizar feedback de procesos como guía continua de mejora.
- Computar información cuantitativa como método de pronóstico y soporte del modelo de decisión.

UTILIDAD

Según Peter Drucker *“sólo tendrán éxito en los próximos años las organizaciones capaces de diagnosticar su salud global”*. Conocer el estado de situación es el primer paso clave en todo proceso directivo para lo cual deberán tener incorporado sistemas permanentes de diagnóstico en los diferentes niveles. Los desafíos que enfrentaban históricamente las empresas y que llevaron al desarrollo de Sistemas de Mediciones de Performance son permanentes y parecidos a los que tiene la empresa actual: **la necesidad de diagnosticar la situación, concretar objetivos y bajar consignas claras**. En economías en constante proceso de globalización se debe disponer permanentemente de información interna y externa que permita estar constantemente actualizado. Esa información, de no ser organizada adecuadamente, corre el riesgo de volverse inerte e incluso constituir un obstáculo.

TIPOS DE TABLEROS

Existe la posibilidad de implementar cuatro tipos genéricos de Tableros:

- 1. Tablero de Control Operativo:** Permite hacer un seguimiento del estado de situación de un sector o proceso de la empresa, para poder tomar a tiempo las medidas correctivas necesarias. El Tablero debe proveer la información que se necesita para entrar en acción y tomar decisiones operativas en áreas como las finanzas, compras, ventas, precios, producción, logística, etc.
- 2. Tablero de Control Directivo:** Permite monitorear los resultados de la empresa en su conjunto y de los diferentes temas claves en que puede segmentarse. Está más orientado al seguimiento de indicadores de los resultados internos de la empresa en su conjunto en el corto plazo. Su monitoreo es mensual. Puede incluir indicadores de todos los sectores para los directivos claves o sectorizado para un directivo. Permite diagnosticar la situación global de la empresa hacia adentro, conteniendo mediciones de resultados con cifras, coeficientes, porcentajes y todo tipo de indicadores. Se dice que lo que no se mide no se controla, por lo que en el tablero se debe intentar medir todos los aspectos relevantes, cualitativos o cuantitativos, de corto o largo plazo que hagan al objetivo del control. De esta forma se puede ampliar la tradicional visión financiera de los sistemas de información, que con una perspectiva exclusivamente cortoplacista y económica sólo presentan a fin de mes los informes de gestión financiera, generando el manejo de la empresa por el presupuesto, inversiones, costos, ventas, activos, rentabilidad, deuda, pasivos y flujo de caja.



- 3. Tablero de Control Estratégico:** Brinda la información interna y externa necesaria para conocer la situación y evitar sorpresas desagradables importantes respecto al posicionamiento estratégico y a largo plazo de la empresa. "La dirección por objetivos funciona si se conocen los objetivos y hoy en día el 90% de las veces se los ignora." En un entorno tan cambiante, para los directivos de una empresa el Tablero de Control podría pasar a ser el centro del proceso de dirección estratégica. Sus usuarios suelen ser tanto los accionistas, directores o los número uno de las empresas, así como los analistas externos que precisen hacer análisis de la situación de la compañía. Estos usuarios se manejan con los indicadores más relevantes, agregándoles información seleccionada del entorno y del sector para poder diagnosticar correctamente el futuro de la empresa.
- 4. Tablero de Control Integral:** Información relevante para que la alta dirección de una empresa pueda conocer la situación integral de su empresa. ***"Información es aquello que contesta una pregunta."*** El Tablero de Control Integral es la consolidación en una "unidad integrada" de los tres tipos de tableros vistos para ser usada por los primeros niveles directivos en orden a avanzar en una dirección estratégica sin papeles. Esto implica que la información más relevante para dirigir se encuentre organizada y sistematizada en una sola herramienta tecnológica e incorporada en el proceso directivo. Como fortaleza del TCI debe remarcarse que la metodología de área e indicador clave ha resultado muy útil para vincular las definiciones de negocio y las bases de datos e información.

PERSPECTIVAS

Las perspectivas son áreas de actuación enfocadas a un determinado ámbito de la organización. Son claves para entender la estructura y la estrategia de una organización para el logro de sus objetivos. Estas son:

- 1. Perspectiva del Cliente:** las organizaciones que conozcan las preferencias de los clientes a quienes desean llegar pueden sobrevivir y después tener la posibilidad de triunfar. Esto no significa que los procesos internos o la innovación hayan dejado de ser importantes, sino que cuando se conoce lo que desea el cliente se puede trabajar mejor en esa perspectiva.

Indicadores:

- Satisfacción del cliente.
- Rentabilidad del cliente: Utilidad producida por el cliente / Gastos para obtenerlo.
- Incremento de clientes.
- Retención de clientes (tiempo de permanencia).
- Precio (respecto a otras ofertas).
- Reconocimiento de la marca.
- Tiempo de entrega, demora en las entregas, % entregas demoradas respuestas.
- Crecimiento anual de compras.

- 2. Perspectiva Procesos Internos:** basada en la perspectiva del cliente busca la forma en que los procesos internos puedan satisfacer los requerimientos detectados, desarrollando los productos y servicios que sean necesarios. Procesos internos excelentes permitirán alcanzar ventajas competitivas sostenibles en el tiempo. Se debe trabajar sobre procesos que tienen valor para el cliente (cadena de valor).



Indicadores:

- Tasa de Fallas (reclamos) clientes (cantidad de reclamos sobre total de clientes).
- Tasa de repetición de reclamos o visitas.
- Tiempo de demora en la entrega o reparación.
- Costo por errores de diseño.
- Nuevas cuentas de clientes.
- Reconocimiento de la marca.
- Eficiencia de proceso de venta o reparación (tiempo con el cliente).
- Tiempos de producción.

- 3. Perspectiva Formación y Crecimiento:** No importa donde se encuentre una empresa ni que productos elabore o los servicios que ofrezca, deberá competir con otras de distintas culturas organizacionales. Esta perspectiva se orienta a las capacidades y herramientas que son necesarias para que el personal de la empresa logre conseguir los objetivos estratégicos. Los objetivos están relacionados con la mejora en la capacitación de los empleados y el desarrollo de sus habilidades, el incremento de la motivación y coordinación o mejorar la utilización de herramientas (como sistemas de información, que mejoren la gestión de la cadena de valor o que faciliten la prestación del servicio). Los indicadores se relacionan con el nivel de satisfacción de los trabajadores, los niveles de productividad del personal, los niveles de cualificación o la inversión realizada en tecnologías y sistemas de información.

Indicadores:

- Utilidad / Cantidad de Empleados.
- Utilidad / Costo Laboral.
- Porcentaje de aumento de capacitación.
- Medición de satisfacción y retención del personal.
- Números de sugerencias.
- Números de sugerencias puestas en prácticas.
- Porcentaje del personal que tiene vinculada su remuneración a incentivos.

- 4. Perspectiva Financiera:** vincula los objetivos relacionados con la contabilidad y demás datos económicos de cada unidad del negocio de la organización, como el incremento de la facturación, la reducción de los costos directos o el aumento de la rentabilidad. Estos objetivos son el resultado de acciones pasadas y son fácilmente medibles, ya que pueden obtenerse de la contabilidad de la organización. Sirve de enfoque para todos los objetivos e indicadores de todas las demás perspectivas.

Indicadores:

- Crecimiento de ventas.
- Crecimiento de clientes.
- Crecimiento de mercados.
- Nuevos productos y servicios.
- Potencial de crecimiento.
- Margen de Utilidad Neta = $\text{Utilidad Operativa Neta} / \text{Ventas}$.
- Rentabilidad Bruta = $(\text{Ventas} - \text{Costos}) \times 100 / \text{Ventas}$.
- Retorno de la Inversión = $\text{Utilidad Neta} \times 100 / \text{Ventas}$.
- Rendimiento de los Activos = $\text{Utilidad Operativa Neta} \times 100 / \text{Patrimonio Neto}$.
- Solvencia = $\text{Activo Total} / \text{Pasivo Exigible Total}$.
- Liquidez = $\text{Activo corriente} / \text{Pasivo Corriente}$.



DATA WAREHOUSING y DATA MINING

La orientación al cliente es un planteo estratégico en las empresas comerciales. La materialización de esta estrategia encuentra un fuerte respaldo en dos manifestaciones de la tecnología informática: el **Data Warehousing** y el **Data Mining**.

El concepto de **Data Warehouse** (DW o almacén de datos) se refiere a la adecuada administración de la información actual e histórica registrada en una organización, a fin de suministrar a cada ejecutivo los elementos apropiados para la adopción de sus decisiones. Su recurso fundamental es la información y su objetivo es realizar un manejo eficiente de esa información.

Un DW no busca producir información nueva sino organizar los datos generados a lo largo del tiempo por todos los subsistemas de información de la empresa, a fin de:

- ✓ Facilitar la manipulación de los datos por los usuarios.
- ✓ Ofrecer información histórica para el análisis de tendencias.
- ✓ Integrar los datos para permitir análisis corporativos.
- ✓ Transformar los datos en información útil y oportuna.

El DW puede permitirle a una empresa:

- ✓ Conocer mejor a cada cliente y entender más claramente las necesidades del mismo.
- ✓ Evaluar los aspectos financieros de cada uno de los productos y servicios ofrecidos.
- ✓ Acelerar la etapa de lanzamiento de nuevos productos.
- ✓ Realizar análisis de tendencias.
- ✓ Elaborar modelos de utilidades.
- ✓ Analizar y administrar el riesgo.

Puede decirse que el DW proporciona al directivo de un supermercado el mismo conocimiento que tenía el almacenero, quien sabía que determinado cliente compraba arroz todos los fines de semana. Bajo el concepto del DW el cliente ya no es una persona indiferenciada, sino un individuo concreto e identificado que tiene determinadas preferencias, una cierta capacidad de compra, una familia integrada de un modo conocido, una determinada vivienda, un cierto automóvil, etc.

La tecnología de DW se complementa con la de Data Mining.

El Data Mining consiste en la búsqueda de patrones o relaciones ocultas en la información contenida en un DW. Debajo de la montaña de información de las transacciones de una empresa, hay un conjunto de vínculos y patrones globales como, por ejemplo, la relación entre un cliente y los productos que compra habitualmente. Para descubrir esta “información escondida” a fin de utilizarla en estrategias de marketing, el Data Mining emplea herramientas de inteligencia artificial, tales como modelización predictiva, inducción a reglas, patrones secuenciales, agrupamiento de series de tiempo, segmentación, etc.

Con una implementación de DW y Data Mining cualquier empresa podría encarar estrategias de comercialización basadas en el conocimiento individual y concreto de cada cliente.



DATA WAREHOUSE (DW)

En el pasado, las organizaciones eran estructuradas en forma piramidal con la información generada en su base fluyendo hacia lo alto, lugar donde se tomaban decisiones a partir de ella, con un bajo aprovechamiento del potencial de esa información. Hoy en día cuentan en su mayoría con procesos automatizados, manejando gran cantidad de datos en forma centralizada y manteniendo sus sistemas en línea. Sin embargo los sistemas tradicionales presentan dificultades para satisfacer las necesidades de información, lo que dio lugar al surgimiento del Data Warehouse (en adelante DW) como solución a la necesidades globales de la empresa.

El término **DW** se traduce literalmente como **Almacén de Datos** y es una colección de datos de una organización, integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones.

El **objetivo** del **DW** es **satisfacer los requerimientos de información interna de la empresa para una mejor gestión**. El DW es el lugar donde la gente puede acceder sus datos, permitiendo así un mejor control de la información que se está abarcando. Al manejar eficientemente la información de cada área de la empresa, se pueden tomar mejores decisiones, efectuar acciones apropiadas y finalmente conseguir un mejor control sobre la producción empresarial.

Su ventaja principal se basa en el almacenamiento de información homogénea y fiable en una estructura basada en la consulta y el tratamiento jerarquizado de la misma y en un entorno diferenciado de los sistemas operacionales.

El DW **se caracteriza** por ser:

- ✓ **Integrado:** los datos almacenados en el DW se integran en una estructura consistente, por lo que las inconsistencias existentes entre los sistemas operacionales deben ser eliminadas. La información suele estructurarse en distintos niveles de detalle para adecuarse a las distintas necesidades de los usuarios.
- ✓ **Temático:** sólo los datos necesarios para el proceso de generación del conocimiento del negocio se integran desde el entorno operacional. Los datos se organizan por temas para facilitar su acceso y entendimiento por parte de los usuarios finales. Por ejemplo, todos los datos sobre clientes pueden ser consolidados en una única tabla del DW. De esta forma, las peticiones de información sobre clientes serán más fáciles de responder dado que toda la información reside en el mismo lugar.
- ✓ **Histórico:** en los sistemas operacionales, los datos reflejan el estado de la actividad del negocio en el momento presente. Por el contrario, la información almacenada en el DW sirve para realizar análisis de tendencias ya que se carga con los distintos valores que toma una variable en el tiempo para permitir comparaciones.
- ✓ **No volátil:** el almacén de información de un DW existe para ser leído, y no modificado. La información es por tanto permanente, significando la actualización del DW la incorporación de los últimos valores que tomaron las distintas variables contenidas en él sin ningún tipo de acción sobre lo que en él ya existía.



El énfasis no está sólo en llevar la información hacia lo alto sino a través de la organización para que todos los empleados que la necesiten la tengan a su disposición.

El DW puede verse como una bodega donde están almacenados todos los datos necesarios para realizar las funciones de gestión de la empresa, de manera que puedan utilizarse según se necesiten.

Los sistemas transaccionales son dinámicos, en el sentido que constantemente se encuentran actualizando datos. Analizar esta información puede presentar resultados distintos en cuestión de minutos, por lo que se deben extraer y almacenar fotografías de datos (snapshots) para estos efectos, con la implicancia de un consumo adicional de recursos de cómputo.

Además, llevar a cabo un análisis complejo sobre un sistema transaccional puede resultar en la degradación de su performance con el consiguiente impacto en la operación del negocio, lo que se evita usando el DW puesto que éste último se construye con los datos a ser analizados que existen en los sistemas transaccionales.

Los DW generan bases de datos con una perspectiva histórica, utilizando datos de múltiples fuentes que se fusionan en forma congruente. Estos datos se mantienen actualizados, pero no cambian al ritmo de los sistemas transaccionales.

Así un DW resulta ser un **recipiente de datos transaccionales para proporcionar consultas operativas, y la información para poder llevar a cabo el análisis multidimensional.** De esta forma, dentro de un almacén de datos existen dos tecnologías complementarias, una relacional para consultas y una multidimensional para análisis.

Se denomina **Data Warehousing** al **proceso ETL** (Extract, Transform and Load) **de extraer y filtrar datos** de las operaciones comunes de la organización, procedentes de los distintos sistemas de información operacionales y/o sistemas externos, para **transformarlos, integrarlos y almacenarlos** en un DW con el fin de acceder a ellos para dar soporte en el proceso de toma de decisiones de una organización.

Un **Data Mart** es un **subconjunto de los datos del DW** cuyo **objetivo es responder a un determinado análisis, función o necesidad de un área específica dentro del negocio** y ayudarla a que pueda tomar mejores decisiones.

COSTOS vs. VALOR de DW

En todo proyecto es esencial realizar siempre un análisis desde la perspectiva Costo/Valor.

Los costos asociados a un proyecto DW incluyen el costo de construcción y el de mantenimiento y operación una vez que está construido.

En cuanto al valor, éste considera, el valor de mejorar la entrega de información, el valor de mejorar el proceso de toma de decisiones y el valor agregado para los procesos empresariales.



COSTOS DE UN DW

COSTOS DE CONSTRUCCIÓN

Los costos de construir un DW son similares para cualquier proyecto de tecnología de información. Estos pueden ser clasificados en tres categorías:

- **RRHH:** de su desarrollo participan los responsables del negocio y los especialistas tecnológicos quienes deben trabajar juntos compartiendo su conocimiento y destrezas.
- **Tiempo:** para la planeación del proyecto, la definición de la arquitectura y para la construcción y entrega de resultados.
- **Tecnología:** costo de la tecnología a adquirir para la realización del proyecto.

COSTOS DE OPERACIÓN / MANTENIMIENTO

Una vez que está construido y entregado un DW debe ser soportado para que tenga valor empresarial. Son justamente estas actividades de soporte, la fuente de continuos costos operacionales de un DW. Se pueden distinguir tres tipos de costos de operación:

- **Evolutivos:** ajustes continuos del DW a través del tiempo, como cambios de expectativas y cambios producto del aprendizaje del RRHH.
- **De Crecimiento:** Incrementos en los volúmenes de datos, del número de usuarios del DW, lo cual conllevará a un incremento de los recursos necesarios como a la demanda de monitoreo, administración y sintonización del DW (evitando así, un incremento en los tiempos de respuesta y de recuperación de datos, principalmente).
- **De Cambios:** El DW requiere soportar cambios que ocurren tanto en el origen de datos que éste usa como en las necesidades de la información que éste soporta. Cuando se implementa un DW el impacto de los cambios es compuesto y existen dos orígenes:
 - **Cambios en el ambiente empresarial:** Un cambio en el ambiente empresarial puede cambiar las necesidades de información de los usuarios. Así, el contenido del DW se puede ver afectado y las aplicaciones pueden requerir cambios.
 - **Cambios en la tecnología:** Un cambio en la tecnología puede afectar la manera que los datos operacionales son almacenados, lo que implicaría un ajuste en los procesos de Extracción, Transporte y Carga para adaptar las variaciones presentadas.



VALOR DEL DW

El valor de un DW queda descrito en tres dimensiones:

1. **Mejorar la Entrega de Información:** información completa, correcta, consistente, oportuna y accesible. Información que la gente necesita, en el tiempo que la necesita y en el formato que la necesita.
2. **Mejorar el Proceso de Toma de Decisiones:** con un mayor soporte de información se obtienen decisiones más rápidas; el decisor adquiere mayor confianza en sus propias decisiones y las del resto, y logra un mayor entendimiento de los impactos de ellas.
3. **Impacto Positivo sobre los Procesos Empresariales:** cuando a la gente se le da acceso a una mejor calidad de información, la empresa puede lograr por sí sola:
 - ✓ Eliminar los retardos de los procesos empresariales que resultan de información incorrecta, inconsistente y/o no existente.
 - ✓ Integrar y optimizar procesos empresariales a través del uso compartido e integrado de las fuentes de información.
 - ✓ Eliminar la producción y el procesamiento de datos que no son usados ni necesarios, producto de aplicaciones mal diseñados o ya no utilizados.

DATA MINING o MINERÍA de DATOS

Data Mining o minería de datos es el conjunto de tecnologías avanzadas, susceptibles de analizar la información de un Data Warehouse para obtener tendencias, para segmentar la información o para encontrar correlaciones en los datos.

Desde un punto de vista orientado al proceso, lleva adelante la búsqueda de patrones ocultos, válidos y potencialmente útiles en grandes conjuntos de datos sin una idea predeterminada, o hipótesis acerca de cuáles pueden ser esos patrones. Consiste en descubrir relaciones insospechadas o previamente desconocidas entre los datos que permitan la extracción de información.

Los patrones descubiertos en la base de datos son utilizados para predecir futuras tendencias y comportamientos, permitiendo tomar decisiones proactivas y conducidas por un conocimiento acabado de la información.

Las técnicas de Data Mining son el resultado de un largo proceso de investigación y desarrollo de productos. Esta evolución comenzó cuando los datos de negocios fueron almacenados por primera vez en computadoras, y continuó con mejoras en el acceso a los datos, y recientemente con tecnologías generadas para permitir a los usuarios navegar a través de los datos en tiempo real. Data Mining toma este proceso de evolución más allá del acceso y navegación retrospectiva de los datos, hacia la entrega de información prospectiva y proactiva.



Las técnicas de Data Mining pueden ser implementadas rápidamente en plataformas ya existentes de software y hardware para acrecentar el valor de las fuentes de información.

Una vez implementadas, pueden analizar bases de datos masivas para brindar respuesta a preguntas tales como, "¿Cuáles clientes tienen más probabilidad de responder al próximo mailing promocional, y por qué?" y presentar los resultados en formas de tablas, con gráficos, reportes, texto, hipertexto, etc.

Los componentes esenciales de la tecnología de Data Mining han estado bajo desarrollo por décadas, en áreas de investigación como estadísticas, inteligencia artificial y aprendizaje de máquinas. Hoy la madurez de estas técnicas junto con los motores de bases de datos relacionales de alta performance hizo que estas tecnologías fueran prácticas para los entornos de DW actuales.

ALCANCE

Al examinar bases de datos de suficiente tamaño y calidad la tecnología de Data Mining puede generar nuevas oportunidades de negocios al proveer estas capacidades:

- ✓ **Predicción automatizada de tendencias y comportamientos.** automatiza el proceso de encontrar información predecible en grandes bases de datos. Preguntas que requerían un intenso análisis, pueden ser contestadas rápidamente desde los datos.
Ej.: pronosticar problemas financieros futuros y otras formas de incumplimiento, identificar segmentos de población que respondan similarmente a eventos pasados.
- ✓ **Descubrimiento automatizado de modelos previamente desconocidos.** Las herramientas de Data Mining barren las bases de datos e identifican modelos previamente escondidos en un sólo paso. Otros problemas de descubrimiento de modelos incluye detectar transacciones fraudulentas de tarjetas de créditos e identificar datos anormales que pueden representar errores de tipeado en la carga de datos.

¿CÓMO TRABAJA el DATA MINING?

¿Cuán exactamente es capaz Data Mining de predecir cosas importantes que se desconocen o que van a pasar? La técnica usada para realizar estas hazañas en Data Mining se llama Modelado.

El **Modelado** es el acto de construir un modelo. Los modelos se originan, ya sea como un conjunto de ejemplos o como una relación matemática, en base a los datos de situaciones sobre las que se conoce la respuesta, para luego aplicar el modelo a otros eventos donde existen cuestiones por resolver.

Las técnicas de modelado no son de nueva aparición, sino que han estado disponibles durante décadas, aunque sólo recientemente se ha logrado alcanzar la capacidad de almacenamiento de datos y de comunicación necesarios para recoger y guardar grandes volúmenes de información. A ellos hay que añadir la potencia de cálculo necesaria para automatizar las técnicas de modelado, que permite trabajar directamente sobre los datos.



La **creación de modelos** de Data Mining aplicada a los negocios **permite**:

- ✓ **Extraer conocimiento de los datos.**
- ✓ **Exprimir todo el valor que encierra ese conocimiento.**
- ✓ **Perfeccionar la estrategia empresarial.**

Ejemplo de ello sería un modelo que defina a un grupo específico de clientes, los mejores, los que más consumen y además lo hacen con más frecuencia.

Para construir el modelo de data mining sería necesario conocer mejor a los clientes y para ello habría que acudir a la base de datos, profundizando en la información que se tiene de cada uno (edad, sexo, ocupación, código postal, ingresos... todos los datos disponibles).

El análisis de estos datos se completaría con un estudio que permita identificar características comunes de los clientes objetivo. De esta forma, se habría detectado un segmento, un modelo de clientes de alto valor sobre los que trabajar de muchas formas; entre ellas:

- Ofreciendo promociones personalizadas.
- Diseñando campañas de marketing específicas.
- Aplicando técnicas de retención.

Esta construcción de un modelo es algo que se vino haciendo desde antes del auge de las computadoras y de la tecnología de Data Mining.

Lo que ocurre en las computadoras, no es muy diferente de la manera en que la gente construye modelos. Las computadoras son cargadas con mucha información acerca de una variedad de situaciones donde una respuesta que es conocida y luego el software de Data Mining debe correr a través de los datos y distinguir las características de los datos que llevarán al modelo.

Una vez que el modelo se construyó, puede ser usado en situaciones similares donde no se conoce la respuesta.

BIG DATA



BIG DATA y la INTELIGENCIA de NEGOCIOS

La inteligencia de negocios es un modelo que antecede en el tiempo al surgimiento de Big Data, orientado a ayudar a la toma de decisiones y que está basado en reportes, consultas y cuadros de mando o tableros de control que contienen información de todas las áreas de la organización, donde los directivos de la organización buscan respuestas a las principales cuestiones utilizando la analítica objetiva, es decir, resolver preguntas como ¿qué pasó ayer?, ¿cuántos clientes fueron los que compraron en los últimos n meses? entre otros, para descubrir ideas más esclarecedoras respecto a rendimiento del negocio.



En los últimos años ha cambiado el hábito de las personas, quienes hoy en día poseen un teléfono móvil o smartphone, una tablet, una computadora portátil o de escritorio en su casa, su trabajo y hasta en el automóvil, un Smart TV, lo que se ha adoptado y asumido como algo natural pues aporta comodidad. Cada dispositivo conectado permite además recabar datos sobre la ubicación de las personas, el rendimiento de máquinas, estado del clima, del tráfico, la salud de pacientes y hasta sentimientos, gustos, preferencias y comportamientos expresados en publicaciones realizadas en redes sociales.

La consecuencia de ello es la generación y toma de datos, en su mayoría no estructurados, en cantidades exorbitantes producto de la infinidad de interacciones que tiene el ser humano con las diferentes tecnologías en el mundo digital.



Se puede definir Big Data como “grandes volúmenes de datos” o “datos masivos”, y todo lo que se puede hacer con estos datos pero a gran escala. Aunque no solo se trata de un montón de datos, sino también de los procesos que se agrupan alrededor de estos, desde su recolección, su almacenamiento y su posterior procesamiento y análisis con el objetivo de extraer valor de los mismos.

Big Data es la capacidad de generar valor alrededor de grandes volúmenes de datos, tanto estructurados como no estructurados, y a partir de su procesamiento poder analizarlos en forma descriptiva, proyectiva y prescriptiva para tomar decisiones.

CARACTERÍSTICAS de BIG DATA

Las características que diferencian a Big Data frente a otros conceptos como el Data Mining o el Business Intelligence, se pueden identificar como las **5V**:

- un gran **V**olumen de datos (cantidad),
- generados y utilizados a gran **V**elocidad,
- provenientes de una gran **V**ariiedad de fuentes, formatos y estructuras,
- que no siempre son **V**eraces (intrínsecamente imprecisos) y que
- no todos tienen **V**alor (eficiencia y rentabilidad) por lo que es necesario discernir esto.



Volumen: a cada instante se generan cantidades de datos imposibles de imaginar hace unos años, pasando en poco tiempo de los Gigabytes a los Terabytes, a los Petabytes y recientemente a los Zettabytes (1.099.511.627.776 Gigabytes). Quizás en un futuro no muy lejano se alcance a medirlos en Yottabytes. La cantidad de datos almacenados crece de forma exponencial cada año.

Velocidad: en la obtención de los datos. Cada vez es más fácil obtener datos o mediciones con una frecuencia mayor, teniendo en cuenta que se generan datos cada segundo, incluidos entre ellos transacciones, fotos y video, por lo que también quedan desfasados y pierden su valor rápidamente.

Ej.: con el avance de los chips GPS y el alcance de las redes inalámbricas es posible medir la posición de un vehículo con alta precisión y cada pocos segundos, y almacenarla con suma sencillez. De igual forma la velocidad es sumamente importante en la transmisión, en el procesamiento y el análisis de los datos, donde unos pocos minutos pueden llegar a ser excesivos para detectar un posible fraude en una transacción. La no disponibilidad de la información en el momento oportuno no solo resta valor a la misma sino que inutiliza su aplicación.



Variedad: Big Data se nutre de fuentes de datos muy heterogéneas (sensores de calor, cámaras de seguridad, smartphones, medios de pago, automóviles, sistemas de navegación y GPS, redes sociales, viajes, etc.). La capacidad de análisis de datos hasta el momento limitada solo a datos estructurados y almacenados en bases de datos relacionales, evoluciona con Big Data a través de su capacidad de combinar datos de orígenes distintos y de formatos no homogéneos ni predefinidos, aumentando exponencialmente las posibilidades de análisis. Por ejemplo en un análisis se puede combinar las trazas de dispositivos móviles y los datos meteorológicos para localizar, medir y pronosticar el tráfico de vehículos en una ciudad.

Veracidad: se refiere al nivel de fiabilidad asociada a los datos, puesto que la mayoría llegan con campos faltantes o son incorrectos, o en diferentes formatos. Limpiar estos datos puede ser la actividad creadora más desafiante para generar valor en ellos. Obtener datos de calidad es indispensable para la correcta obtención de información.

Valor: es la característica más importante y el propósito final de los datos. Los datos en su estado natural carecen de valor, siendo el hecho de combinarlos, transformarlos y analizarlos lo que permite extraer el valor latente que hay en ellos. Las organizaciones buscan la forma de convertir los datos en información confiable, de manera que su explotación sea coherente, pensada y alineada con la estrategia de la empresa para que pueda ser aprovechable, rentable y eficiente.

La evolución de las tecnologías de la información y las comunicaciones y su impacto en el aumento en forma exponencial de la capacidad de procesamiento, la velocidad y capacidad de las memorias RAM, la capacidad de almacenamiento, como la disminución también exponencial de los costos de producción de todo ello, ha permitido afrontar la potencia de proceso necesaria para analizar las ingentes cantidades de datos del Big Data, aunque el salto cualitativo ha venido de la mano del almacenamiento y procesamiento distribuido, uniendo las capacidades de muchos computadores o servidores. El Cloud Computing o computación en la Nube es la plataforma que puede alojar toda la infraestructura de almacenamiento, proceso, gestión, seguridad de datos e información con acceso desde internet, y el Software as a Service (SaaS) o Servicio de Software provisto por una organización dedicada a ello como si se tratase de un outsourcing, son los soportes sobre los cuales se erigen las soluciones de Big Data.

Su aplicación en marketing y ventas permite llegar al público objetivo de forma mucho más precisa con herramientas ofrecidas como servicios de facebook y google, usado por ellos mismos como empresas.