



UTN-FRBA

GESTION DE DATOS

INTELIGENCIA DE NEGOCIOS - TECNOLOGÍAS OLAP

DIRECTOR CATEDRA: ING. ENRIQUE REINOSA

INTRODUCCION

- ▶ La inteligencia del negocio consiste en la Transformación de datos en información y, ésta, en conocimiento, con la intención de mejorar al máximo el proceso de toma de decisiones de la organización.
- ▶ Desde la perspectiva de la tecnología de la información, la inteligencia del negocio se define como “el conjunto de metodologías, herramientas y estructuras de almacenamiento” que permiten la reunión, depuración y transformación de los datos en una información integrada que se pueda analizar y convertir en conocimiento para la optimización del proceso de toma de decisiones.

DATOS, INFORMACION Y CONOCIMIENTO

- ▶ En general, estos términos se consideran sinónimos; sin embargo, entre ellos existen grandes diferencias: **los datos** son valores ya conocidos que se encuentran diseminados en diferentes partes; **la información**, en cambio, es un dato asociado a una relación que, vinculado y acumulado luego de un proceso de análisis, se transforma en **conocimiento**.

DATOS

- ▶ Los datos son la mínima unidad semántica que se corresponden con los elementos primarios de la información que en sí mismos no tienen ningún valor. Para brindar algún tipo de información necesitan que se los vincule con alguna relación. También se pueden tomar como un conjunto discreto de valores absolutos que no pueden aportar nada que contribuya con la toma de decisiones. Por ejemplo: un número, un nombre, etc. Asimismo, un CD, un DVD o un disco rígido pueden almacenar físicamente una colección de datos.
- ▶ Los datos, en general, provienen de diferentes orígenes: internos; es decir, de la propia organización, o externos, extraídos del contexto; por ejemplo, de la competencia. A la vez, pueden ser objetivos, subjetivos y de tipo cualitativo o cuantitativo.

INFORMACION

- ▶ Se puede definir a la información “como el conjunto de datos procesados o relacionados con un significado específico”.
- ▶ En consecuencia, si se le añade alguna relación, los datos se pueden convertir en información. A continuación, se describe de qué manera un dato se puede transformar en información:
 - ▶ **Contextualizando:** se sabe en qué contexto y para qué propósito se generaron.
 - ▶ **Categorizando:** se conocen las unidades de medida que ayudan a interpretarlos.
 - ▶ **Calculando:** los datos fueron procesados matemática o estadísticamente.
 - ▶ **Corrigiendo:** habiendo eliminando errores o inconsistencias de los datos.
 - ▶ **Condensando:** resumiendo los datos de forma más concisa (agregación de datos).

CONOCIMIENTO

- ▶ El conocimiento es la fusión de valores, información y experiencia, es el marco conceptual adecuado para la incorporación de nueva información. A medida que se va incorporando más información se generan nuevos conocimientos que contribuirán con la toma de decisiones.
- ▶ Para que la información se convierta en conocimiento se deben llevar adelante las siguientes acciones:
 - ▶ Comparación con otros elementos.
 - ▶ Predicción de consecuencias.
 - ▶ Búsqueda de conexiones.
 - ▶ Conversación con otros portadores de conocimiento.

TECNOLOGIAS OLAP

- ▶ La aparición y evolución de la informática a finales de la década de los 90 trajo una acumulación masiva de datos, en el entorno empresarial. Esta gran cantidad de datos proviene en su mayor parte de la aplicación de la informática en las actividades transaccionales de la empresa, contabilidad, gestión de almacén, facturación.
- ▶ Hace falta un modo de estructurar la información y los datos que aporte una nueva perspectiva de los mismos.
- ▶ Nace de este modo la tecnología O.L.A.P. “On Line Analytical Processing” basada en la utilización de tecnología de Bases de Datos Multidimensionales, para diferenciarse de OLTP “On Line Transaction Processing”, que se fundamentan en Bases de Datos Relacionales.

CONCEPTOS

- ▶ **OLAP:** “On Line Analytical Processing” también llamado **Modelo Relacional**, debido a que **analiza y relaciona la información analizada**.
- ▶ **OLTP:** “On Line Transaction Processing”, también llamado **Modelo Transaccional**, debido a que **se basa en la ejecución de un conjunto de transacciones** para obtener el resultado esperado.

CARACTERISTICAS OLTP

- ▶ Su ejecución se basa en transacciones
- ▶ Conforman el 99% de los sistemas existentes
- ▶ Son sistemas “operativos”
- ▶ Procesan datos
- ▶ Los datos se almacenan normalizados
- ▶ Registran datos nivel de detalle de cada transacción
- ▶ Los datos volátiles

CARACTERISTICAS OLAP

- ▶ Su ejecución se basa en el análisis
- ▶ Conforman el 1% de los sistemas existentes
- ▶ Son sistemas para la toma de decisiones
- ▶ Procesan información
- ▶ La información se almacenan desnormalizados
- ▶ Registran información global por patrones de interés también conocidos como “dimensiones”
- ▶ La información es persistente o “no volátil”

OLTP vs OLAP

- ▶ **OLAP:** Las aplicaciones en OLAP, en cambio, son usadas por analistas y gerentes que frecuentemente quieren vistas de alto nivel de los datos, tales como las ventas de una línea de productos, por región, etc. La base de datos OLAP es usualmente actualizada por bloques, generalmente de múltiples fuentes, y provee poderosas aplicaciones multiusuario de poder analítico. Por lo tanto, las bases de datos OLAP son optimizadas para el análisis.
- ▶ **OLTP:** Las aplicaciones con OLTP están caracterizadas en que muchos usuarios crean, actualizan, o retienen registros individuales. Entonces, las bases de datos con OLTP son optimizadas para las actualizaciones de las transacciones

CARACTERISTICAS OLAP

- ▶ Las herramientas OLAP que se han ido desarrollando han buscado mantener una compatibilidad con otras herramientas de análisis que ya se utilizaban en la empresa, como son las hojas de cálculo tradicionales.
- ▶ La mayoría de los usuarios coincide en la facilidad y beneficio de utilizar aplicaciones multidimensionales pero presentan una actitud expectante ante los problemas de construcción y mantenimiento.

DUDA (???)
Cuales son esos
problemas de
construcción y
mantenimiento???

ESTRUCTURA OLAP

- ▶ La gran mayoría de los datos que se usan en aplicaciones OLAP son originarios de otros sistemas y aplicaciones. De cualquier modo, en casi la totalidad de las aplicaciones OLAP, los datos son capturados directamente por la aplicación OLAP. Cuando los datos proceden de otras aplicaciones es necesario duplicarlos y almacenarlos separadamente de los originales de los cuales proceden, para poder ser utilizados de manera activa por la aplicación OLAP de manera independiente.

ESTRUCTURA OLAP

- ▶ Algunas de las razones que obligan a duplicar los datos para formar el MODELO OLAP son:
 - ▶ **Ejecución:** Las aplicaciones OLAP son con cierta frecuencia de un gran tamaño y se suelen usar para realizar análisis interactivos inciertos. Esto requiere que se pueda acceder a los datos de manera muy rápida, lo cual obliga a que se guarden separados, y a disponer de una estructura de datos optimizada que pueda ser accedida sin perjudicar la respuesta operativa del sistema.
 - ▶ **Múltiples fuentes de datos:** Muchas aplicaciones OLAP requieren de datos que son originados desde distintos y diversos sistemas. El proceso para unir y combinar estos datos procedentes de distintas aplicaciones o sistemas (bases de datos, hojas de cálculo...) puede ser extremadamente complejo, porque estas aplicaciones o sistemas suelen usar sistemas de codificación diferentes y además pueden disponer de periodicidades distintas.

ESTRUCTURA OLAP

- ▶ **Filtrado de datos.:** En la gran mayoría de sistemas transaccionales nos encontramos con mucha frecuencia gran cantidad de datos que necesitan ser filtrados antes para poder realizar un buen análisis que nos permita generar informes adecuados. a tomar una decisión equivocada.
- ▶ **Ajuste y modificación de datos:** Hay varias razones por la cuales los datos deben ser ajustados antes de realizar el análisis. Algunas de ellas, podrían ser:
 - ▶ Sucursales situadas en otros países operan con contabilidades distintas y los datos puede que necesiten ser modificados antes de usarse en el análisis.
 - ▶ Las distintas estructuras de la compañía no siempre son iguales. Existe diferencias en los modos de trabajar de las direcciones departamentales, en las estructuras operativas...
 - ▶ Se pueden realizar análisis que no parten de datos operativos como pueden ser los que se obtienen de las características demográficas, publicidad televisiva...

ESTRUCTURA OLAP

- ▶ **Actualización y consistencia de datos:** Si los datos que van a ser usados por una aplicación OLAP proceden de distintas fuentes es muy probable que no se actualicen todos al mismo tiempo. Es decir, las aplicaciones de las cuales proceden los datos pueden estar, y de hecho seguro que ocurrirá así, en diferentes estados de actualización. El análisis que realiza un OLAP depende en gran medida de la consistencia de los datos y por lo tanto es necesaria una plataforma que garantice esa consistencia.
- ▶ **Historia de los datos:** La gran mayoría de aplicaciones OLAP incluyen el tiempo como una dimensión. El uso del tiempo como una dimensión permite obtener resultados muy provechosos en cuanto a análisis temporales cuando se dispone de datos de varios años atrás. La adquisición de datos años anteriores supone un importante esfuerzo, ya que es necesario migrar estos datos de aplicaciones antiguas y ajustarlos para que puedan ser utilizados en la base de datos del OLAP.

ESTRUCTURA OLAP

- ▶ **Distintas perspectivas o vistas:** Los datos operacionales tienen que ser necesariamente muy detallados, pero muchas de las actividades de toma de decisiones requieren una visión a más alto nivel, no tan estructurada. Interesa, por lo tanto, combinar almacenes de datos, ajustar la información según el nivel de resumen o el nivel de visión que se quiere alcanzar.
- ▶ **Actualización de datos:** Si la aplicación dispone de varias entradas de datos es obvio que es necesario separar la base de datos de OLAP para que no se sobrescriban los datos operacionales que se están usando en un determinado momento.

BASES DE DATOS MULTIDIMENSIONALES

- ▶ Las dimensiones determinan la estructura de la información almacenada y definen adicionalmente caminos de consolidación. La información almacenada se presenta como variables que a su vez están caracterizadas por una o más dimensiones



BASES DE DATOS MULTIDIMENSIONALES

- De este modo la información puede analizarse dentro del cubo formado por la intersección de las dimensiones de la variable particular



A) Visión Regional



B) Visión del Producto



C) Visión Temporal



D) Vista A-d-Hoc

DISPERSION DE DATOS

- ▶ A medida que se agregan dimensiones a una Base de Datos Multidimensional, el número de puntos de datos o “celdas” crece rápidamente. Por ejemplo, considerando que no se venden todos los productos en todas las sucursales todos los días, si se considerara que las sucursales más pequeñas solo pueden manejar el 20% de todos los productos, el 80% de las celdas estarán vacías. En la práctica, muchas Bases de datos tienen el 95% de las celdas vacías o en cero. Esto es conocido como “sparsely populated” ó poblados dispersos ó **dispersión de datos** ó simplemente sparse.

DUDA (???)

xq quedarian vacias? xq se cargan con el resto de la informacion, pero esa faltante serian columnas vacias?

DISPERSION DE DATOS

- ▶ **Hipercubo:** La información se guarda implícitamente en un gran y único cubo, presentando los datos al usuario en un formato de hipercubo, donde todos los datos en la aplicación aparecen como una sencilla estructura multidimensional.
- ▶ **Multicubo:** la información se almacena dividiendo los datos en grupos más pequeños y densos (**objetos multidimensionales**), donde la base de datos multidimensional consiste en un número de objetos separados normalmente con diferentes dimensiones.

BDM vs BDR

	Base de Datos Relacional	Base de Datos Multidimensional
Deposito de datos, acceso y visión	Relacional Tablas de Columnas e hileras Lenguajes SQL con ampliaciones Herramientas de terceros que usan API	Dimensional Arreglos: Hipercubo, Multicubo Tecnología de matriz dispersa Propietario de hoja de calculo
Utilización e incorporación	OLTP Motor RBDMS Profundización a nivel de detalle Desempeño de consultas: rango amplio	OLAP Motor multidimensional Profundización a nivel de resumen/adición Desempeño de consultas: rápido
Tamaño y actualización de bases de datos	Gigabytes a terrabytes El deposito de indices y el retiro de normas que incrementan tamaño Consulta y cargas paralelas Actualizacion durante uso	Gigabytes Compresion y adicion de datos dispersos Dificil actualizar durante uso; los cambios pequeños pueden requerir reorganizacion