



Inteligencia de Negocios



Universidad Iberoamericana Plantel León

Conceptos de Base de Datos



Dato:

- Es una representación simbólica (numérica, alfabética, algorítmica, etc), un atributo o una característica de una entidad.
- Parte mínima de la información.

Información:

- Es un conjunto organizado de **datos** procesados, que constituyen un mensaje que cambia el estado de conocimiento del sujeto o sistema que recibe dicho mensaje.



Conceptos de Base de Datos



Base de Datos:

Colección organizada de datos almacenados en forma estructurada y organizada para dar servicio.

Sistema:

Conjunto de elementos relacionados entre sí, interdependientes, con un fin en común.



Finalidad de los Sistemas de Información

La información *reduce nuestra incertidumbre* (sobre algún aspecto de la realidad) y, por tanto, nos permite tomar mejores decisiones



DSS – Sistema de Soporte de Decisiones

Aplicaciones que proveen análisis de los datos almacenados en la BD.

Beneficios:

- ▶ Ayuda a identificar tendencias clave.
- ▶ Ayuda en la administración y que los tomadores de decisiones transformen los datos en información.
- ▶ Ayuda a las empresas a prever acontecimientos y actuar con anticipación.
- ▶ Ayuda al administrador a comprender el entorno y así rediseñar las prácticas empresariales en reacción a los sucedido

OLAP – Online Analytical Processing

Son un tipo de aplicaciones para soporte de decisiones. Habilitan al usuario para consultar los datos y verlo de múltiples perspectivas.

OLAP Provee:

- De forma rápida, diferentes puntos de vista de los datos.
- Análisis de datos avanzado a través de tablas pivote.
- Flexibilidad, fácil reporteo mediante consultas a la medida.
- Pronósticos´.



Herramientas para la toma de decisiones.

Diferencia entre EIS y OLAP.

- ❑ Un EIS (*Executive Information System*) es un sistema de información y un conjunto de herramientas asociadas:
 - Proporciona a los directivos acceso a la información de estado y sus actividades de gestión.
 - Está especializado en analizar el estado diario de la organización (mediante indicadores clave) para informar rápidamente sobre *cambios* a los directivos.
 - La información solicitada suele ser, en gran medida, numérica (*ventas semanales, nivel de stocks, balances parciales, etc.*) y representada de forma gráfica al estilo de las hojas de cálculo.

- ❑ Las herramientas OLAP (*On-Line Analytical Processing*) son más genéricas:
 - Funcionan sobre un sistema de información (transaccional o almacén de datos)
 - Permiten realizar agregaciones y combinaciones de los datos de maneras mucho más complejas y ambiciosas, con objetivos de análisis más estratégicos.



Herramientas para la toma de decisiones.

Diferencia entre “informes avanzados” y OLAP

- ❑ Los sistemas de informes o consultas avanzadas:
 - están basados, generalmente, en sistemas *relacionales u objeto-relacionales*,
 - utilizan los operadores clásicos: concatenación, proyección, selección, agrupamiento, ... (en SQL y extensiones).
 - el resultado se presenta de una manera tabular.

- ❑ Las herramientas OLAP
 - Están basadas, generalmente, en sistemas o *interfaces multidimensionales*,
 - Utilizando operadores
 - El resultado se presenta de una manera matricial o híbrida.



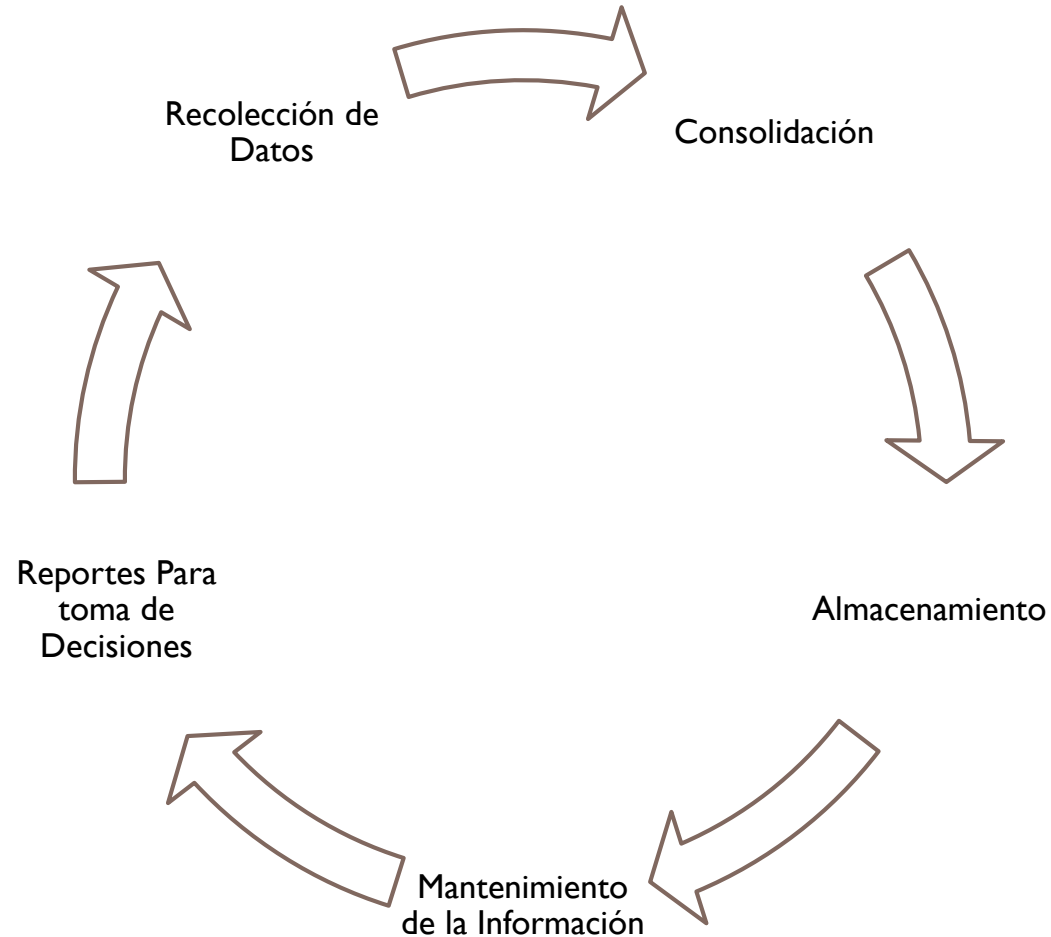
Almacenes de Datos (Data Warehouse)

- ▶ Lugar de almacenamiento centralizado para diferentes tipos de datos de la empresa
- ▶ Datos orientados por tema
- ▶ Sensible al tiempo
- ▶ Organizado de forma que ofrece un análisis simplificado.

Ejemplo: DW Podría contener información de los clientes y sus ventas de los últimos cinco años.



Proceso Data Warehouse



Características del Data Warehouse

- ▶ Ofrece una ventana única de consulta y provee un almacenamiento de información de toda la empresa.
- ▶ Soporta datos para toma de decisiones
- ▶ Resúmenes pre-calculados, incrementa la velocidad al contar con acumulados de información
- ▶ Ofrece instantáneas del tiempo para facilitar el análisis empresarial
- ▶ Carga de datos operacionales.

OLAP vs OLTP

OLTP	OLAP
Orientada a transacciones entre entidades y relaciones	Orientado a hechos y dimensiones
Muchas tablas, esquemas normalizados	Pocas tablas, esquema denormalizado
Usuarios insertan, borran y actualizan registros	Usuario consultan registros de solo lectura.
Muchos usuario accesan los registros, pocos usuarios utilizan reportes	Más reporteo
Utiliza Transacciones / Roll back del log	No utiliza transacciones, no requiere rollback del log
Muchos registros detalle	Registros consolidados, resúmenes de información. Totales
Menor cantidad de índices para velocidad de actualización	Gran cantidad de índices para optimización de consultas

Componentes de Data Warehousing

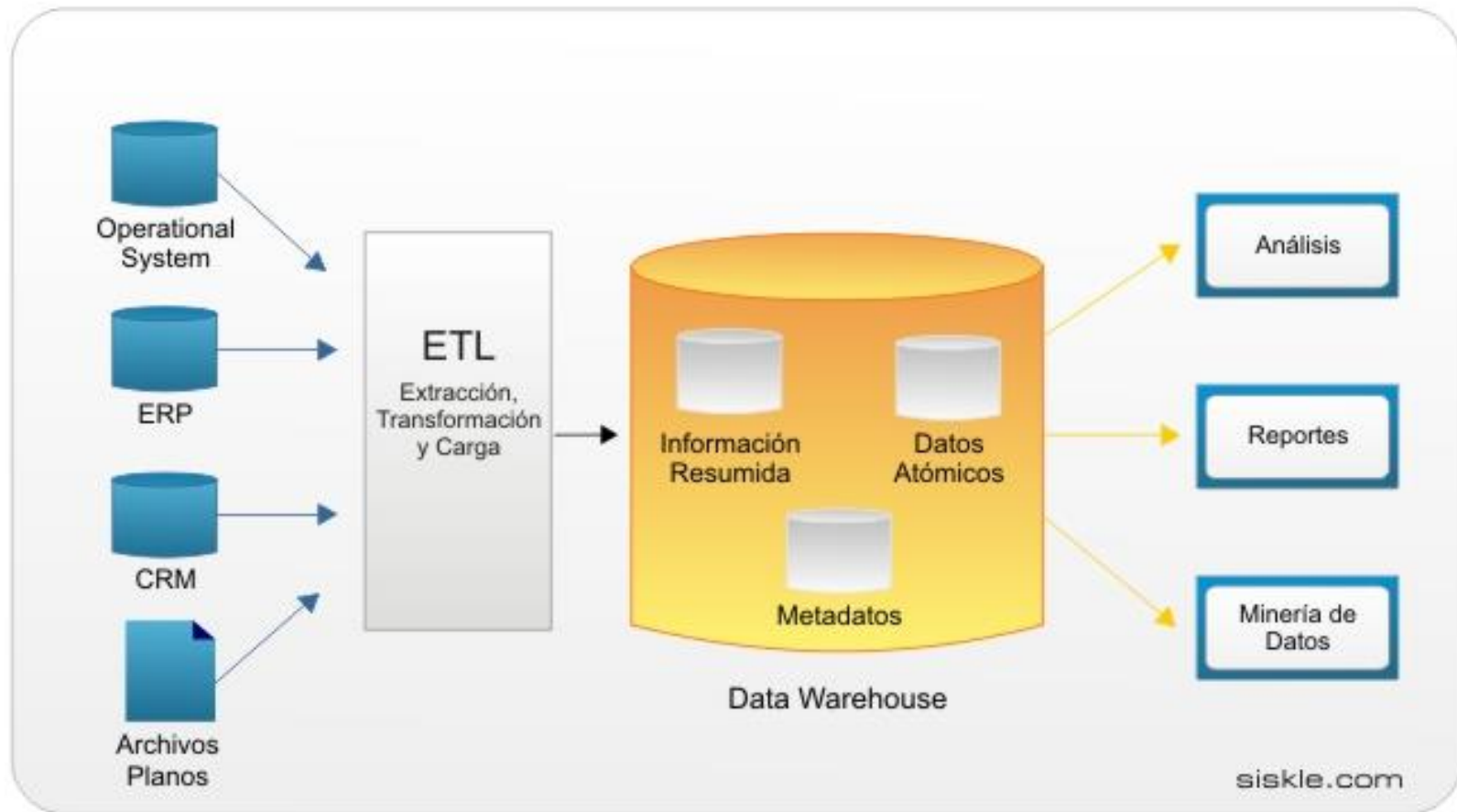


Figura1. Arquitectura del DW

Componentes de Data Warehousing

- ▶ **Almacenamiento de Datos Operacional:**

Sistemas OLTP actualmente en uso en la organización.

- ▶ **Data Transformations: (ETL)**

Proceso por el cual los datos se mueve del sistema OLTP, se modifican, totalizan, consolidan a un formato en el Data Warehouse.

- ▶ **Metadatos:**

Descripción de los datos a ser almacenados en el data warehouse.

- ▶ **Data Store:**

Servidor y Base de Datos que contiene la información. Puede consistir de un Data Warehouse o múltiples Data Marts especializados.



Componentes de Data Warehousing

► **Herramientas de Consulta / Acceso a Usuario Final:**

Herramientas de consulta de la información o aplicaciones hechas en casa que dan acceso a la información almacenada en el Data warehouse.



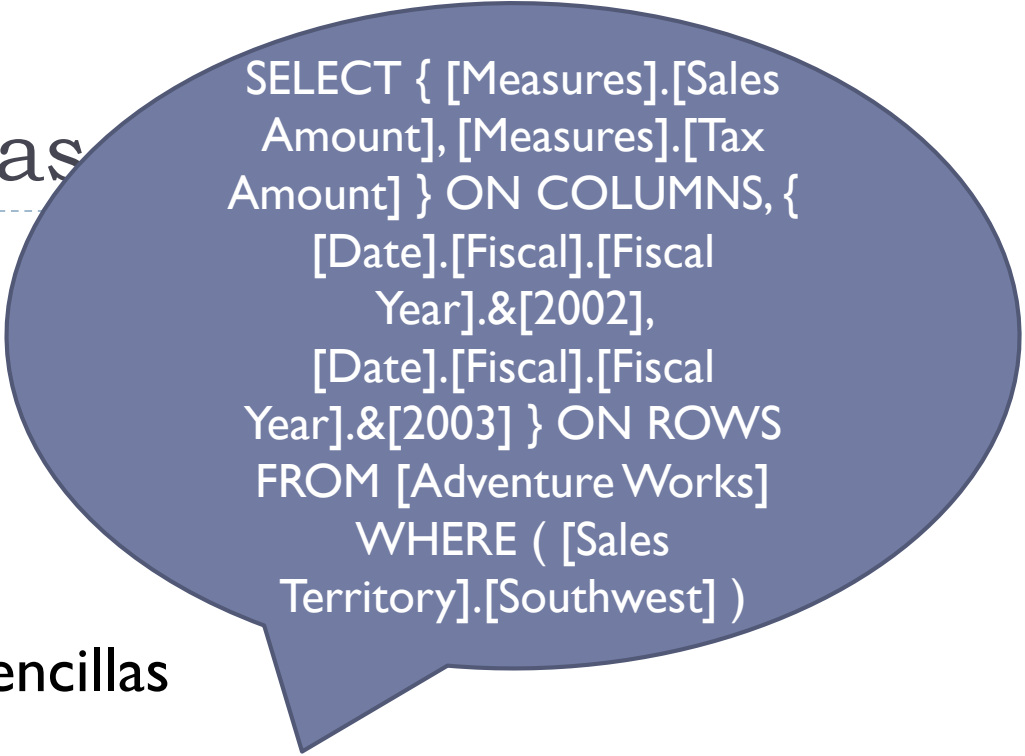
Data Mart

- ▶ Repositorio de datos obtenido de datos operacionales u otras fuentes y es diseñado para servir a un departamento particular o grupo funcional.
- ▶ Data Mart es menos genérico que un Data Warehouse
- ▶ La información se almacena en un formato familiar a el usuario.
- ▶ Consolida la información por cada grupo o departamento funcional.



Datamart - Ventajas

- ▶ Poco volumen de datos
- ▶ Mayor rapidez de consulta
- ▶ Consultas SQL y/o MDX sencillas
- ▶ Validación directa de la información
- ▶ Facilidad para la almacenar información histórica de los datos



```
SELECT { [Measures].[Sales  
Amount], [Measures].[Tax  
Amount] } ON COLUMNS, {  
    [Date].[Fiscal].[Fiscal  
    Year].&[2002],  
    [Date].[Fiscal].[Fiscal  
    Year].&[2003] } ON ROWS  
FROM [Adventure Works]  
WHERE ( [Sales  
Territory].[Southwest] )
```

Tipos de OLAP

ROLAP (Relational On-Line Analytical Process):

Herramientas OLAP que crean vistas multidimensionales extrayendo los datos de bases de datos relacionales. Simulan los datos multidimensionales usando técnicas de indexación, cachés, metadatos, etc.



MOLAP (Multidimensional On-Line Analytical Process):

Son herramientas que acceden a datos que no están almacenados en registros de tablas, sino que almacenan los datos en arreglos de varias dimensiones, llamados cubos. Estos cubos utilizan índices para optimizar el acceso a los datos

HOLAP (Hybrid On-Line Analytical Process):

Une los dos tipos anteriores. Ayuda a reducir costes de hardware ya que necesita menos disco que en las bases de datos relacionales. La respuesta de las consultas sobre las bases de datos multidimensionales son más rápidas que sobre las relacionales

Cuadrante gartner – plataformas bi



Cuadrante gartner – plataformas bi

Figure 1. Magic Quadrant for Business Intelligence and Analytics Platforms



Minería de Datos

- ▶ Proceso de extraer información válida, previamente desconocida, comprensible y útil partiendo de bases de datos de gran tamaño y utilizar dicha información para tomar decisiones de negocio cruciales.

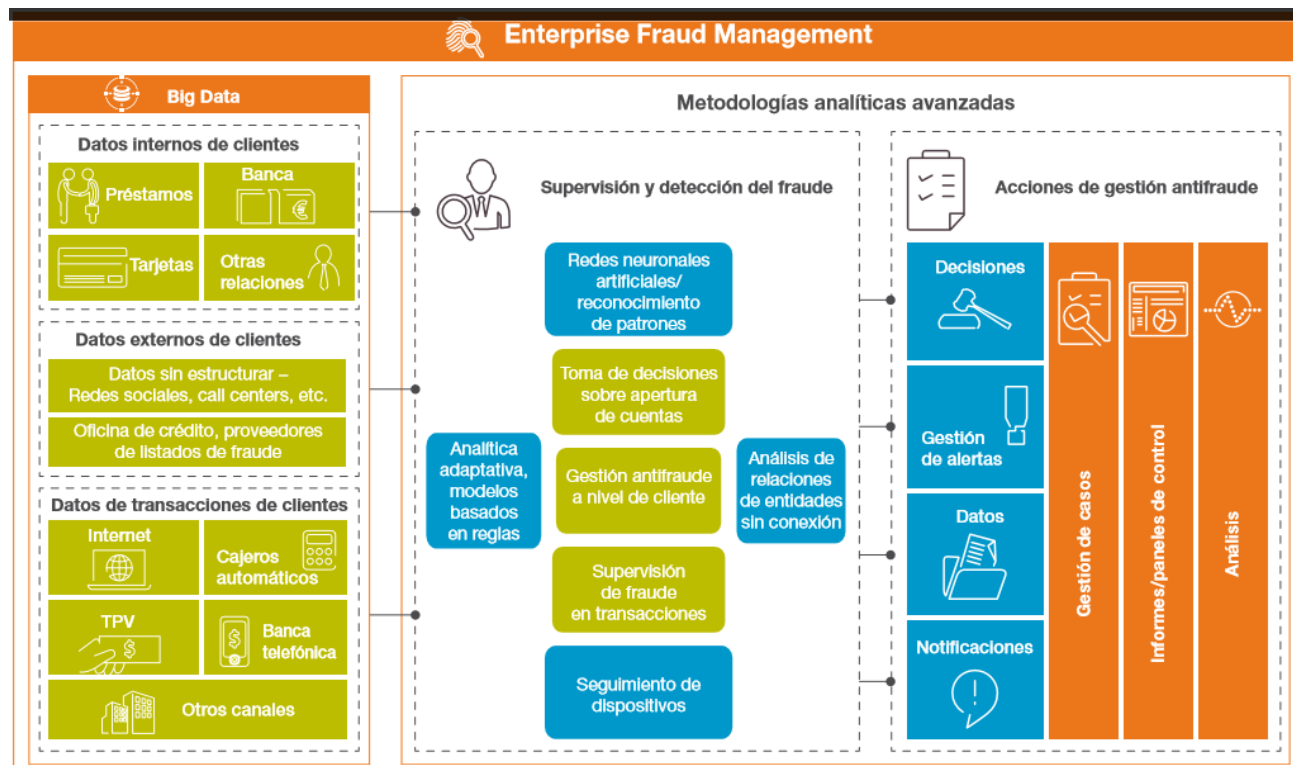


Ejemplos (Banca)

- ▶ Detección de patrones de uso fraudulento de tarjetas de crédito.
- ▶ Identificación de clientes leales
- ▶ Determinar gastos realizados por cierto grupo de clientes con la tarjeta de crédito.



MINERÍA DE DATOS - ANTIFRAUDES



Almacenes de Datos y Minería de Datos

- Los almacenes de datos no son *imprescindibles* para hacer extracción de conocimiento a partir de datos.
 - se puede hacer minería de datos sobre una BD OLTP.
- Las *ventajas* de organizar un almacén de datos para realizar minería de datos se amortizan sobradamente a medio y largo plazo cuando:
 - tenemos grandes volúmenes de datos, o
 - éstos aumentan con el tiempo, o
 - provienen de fuentes heterogéneas o
 - se van a combinar de maneras arbitrarias y no predefinidas.



Herramientas para la toma de decisiones.

Diferencia entre OLAP y Minería de Datos

☐ Las herramientas OLAP

- proporcionan facilidades para “manejar” y “transformar” los datos.
- **producen otros “datos”** (más agregados, combinados).
- ayudan a analizar los datos porque producen *diferentes vistas* de los mismos.

☐ Las herramientas de Minería de Datos:

- son muy variadas: permiten “extraer” patrones, modelos, descubrir relaciones, regularidades, tendencias, etc.
- **producen “reglas” o “patrones” (“conocimiento”).**



Business intelligence

- ▶ Habilidad para transformar los datos en información, y la información en conocimiento, de forma que se pueda optimizar el proceso de toma de decisiones en los negocios.

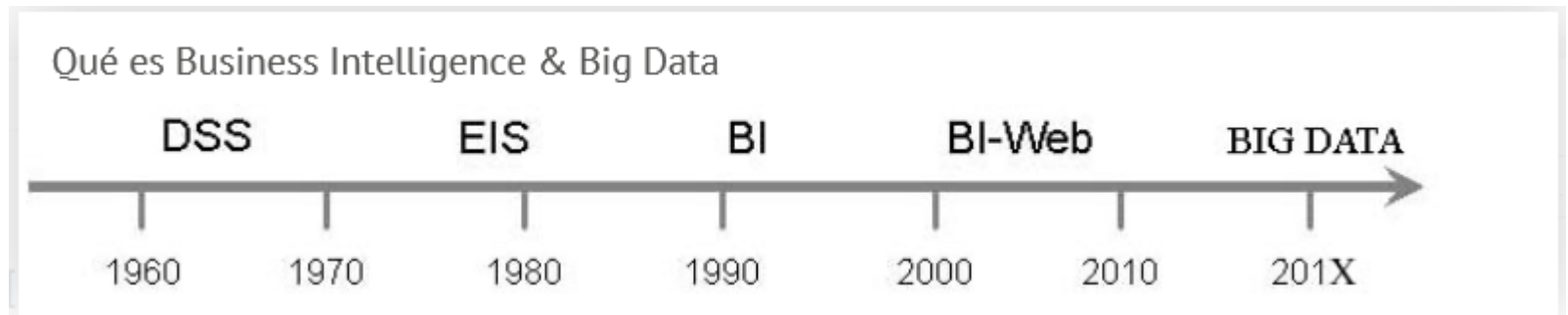


Business intelligence

- ▶ Conjunto de metodologías, aplicaciones y tecnologías que permiten reunir, depurar y transformar datos de los sistemas transaccionales e información desestructurada (interna y externa a la compañía) en información estructurada, para su explotación directa (reporteo, análisis OLTP / OLAP, alertas...) o para su análisis y conversión en conocimiento, dando así soporte a la toma de decisiones sobre el negocio



Evolución de Business intelligence





Big data

“big data en general es definido como alto volumen, velocidad y variedad de activos de información que demandan costo-efectividad, formas innovadores de procesamiento de información para mejorar la comprensión y toma de decisiones”

Origen: gartner



BIG DATA

- Gran Cantidad de Datos
- Datos Estructurados y No Estructurados
- La cantidad no es lo importante sino lo que las organizaciones hacen con esta información



BIG DATA

EL 90% DE LOS DATOS DEL MUNDO QUE EXISTEN HOY
HAN SIDO CREADOS SOLO EN LOS DOS ÚLTIMOS AÑOS.



Videos

Fotos

Big data – volúmenes de datos

A diagram illustrating the scale of data storage units. It consists of a vertical list of units on the left and their corresponding numerical values on the right. Each unit is followed by a horizontal green arrow pointing to the right, with the arrow's length increasing proportionally to the magnitude of the unit. The units are: Kilobyte (KB), Megabyte (MB), Gigabyte (GB), Terabyte (TB), Petabyte (PB), Exabyte (EB), Zettabyte (ZB), and Yottabyte (YB). The numerical values are: 1 thousand, 1 million, 1 billion, 1 trillion, 1 quadrillion, 1 quintillion, 1 sextillion, and 1 septillion.

Kilobyte (KB)	1 thousand
Megabyte (MB)	1 million
Gigabyte (GB)	1 billion
Terabyte (TB)	1 trillion
Petabyte (PB)	1 quadrillion
Exabyte (EB)	1 quintillion
Zettabyte (ZB)	1 sextillion
Yottabyte (YB)	1 septillion

Crecimiento de datos

▶ **Cada 60 segundos:**

- ▶ 98,000 tweets son creados en twitter
- ▶ 695,000 actualizaciones en Facebook
- ▶ 11 millones de mensajes instantáneos son enviados
- ▶ 698,445 búsquedas en google son generadas
- ▶ 1,820 tb de datos son creados
- ▶ 217 nuevos usuarios web mobile son creados.

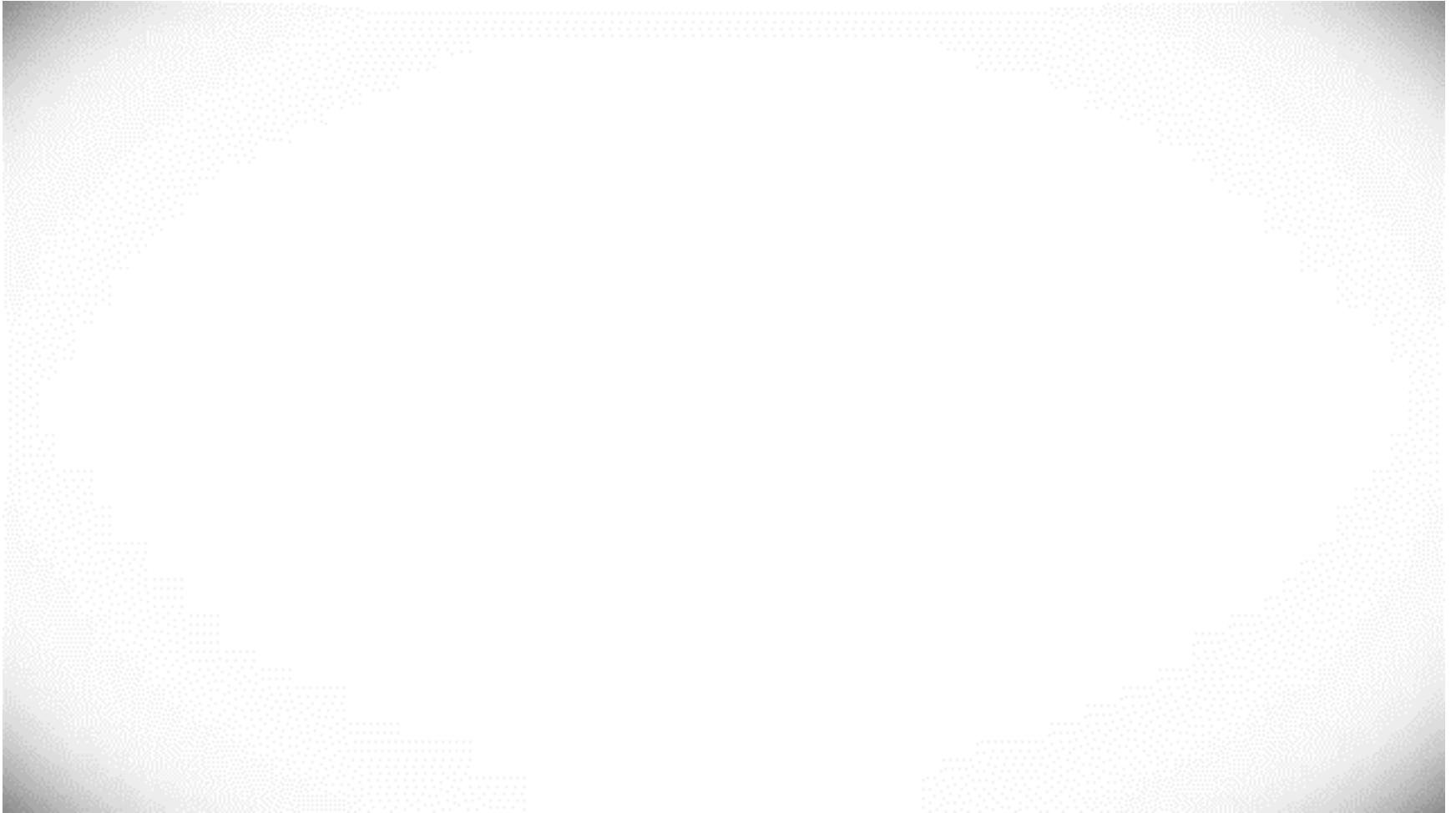


La nueva cara de analítica de negocios

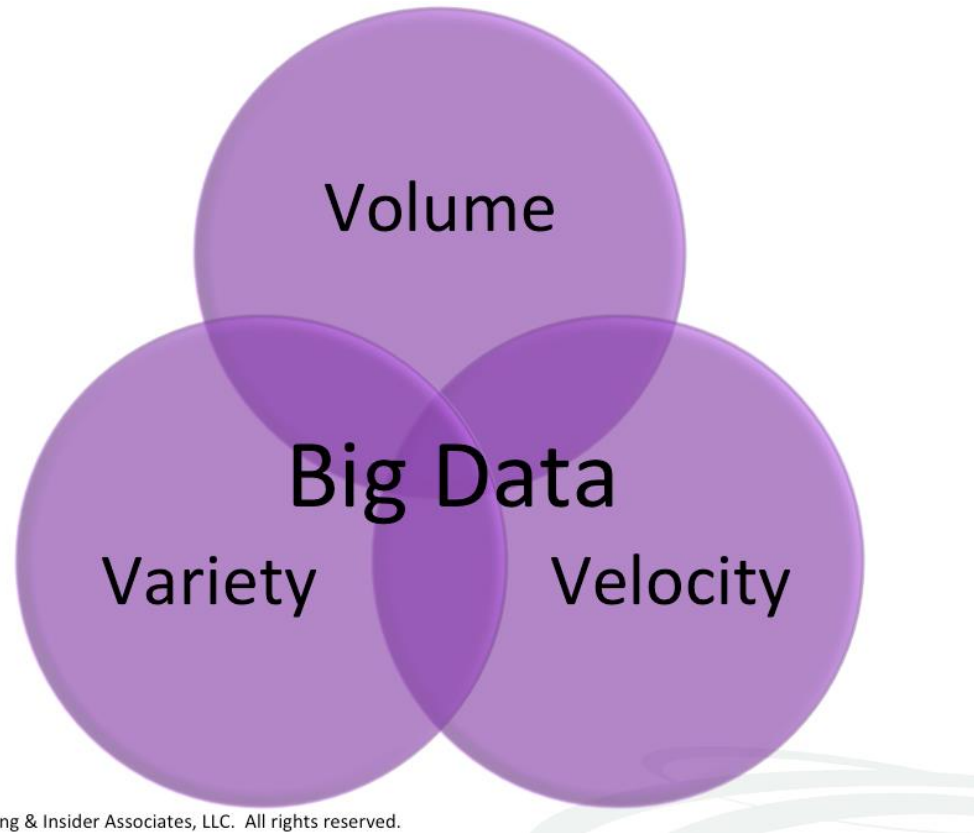
- ▶ **Nosotros tenemos el internet de las cosas**
 - ▶ Streams de datos de cada dispositivo
 - ▶ Múltiples dispositivos
 - ▶ Publicidad
 - ▶ Mercado en línea y detalle de tracking de clicks
 - ▶ Contenido impulsado por el consumidor (p.ej. Redes sociales y colaboración social)



Internet de las cosas



BIG DATA



© 2011 R Wang & Insider Associates, LLC. All rights reserved.



Soluciones BI

