



Tecnológico  
de Monterrey

# BASE DE DATOS AVANZADAS DATA WAREHOUSE

Enrique Alfonso Calderón Balderas



# ¿Qué es un Data Warehouse?

- William. H. Inmon lo define como:  
“Un Data Warehouse es un conjunto de datos integrados, orientados a un ámbito, no volátil y que varían con el tiempo, los cuales soportan al proceso de toma de decisiones la entidad que se utiliza.”
- Repositorio de información disponible para búsqueda y análisis de la misma.
- Considerado proceso continuo de adquisición de datos y análisis de los mismos.



# ¿Por qué un Data Warehouse?

- En la organización, se busca incrementar la productividad y ventajas competitivas
- Recurrir a los datos transaccionales presenta varios problemas:
  - Gran cantidad de información
  - Información genérica
  - Información irrelevante para el área.
- Unión de datos con negocio para:
  - Responder cuestionamientos de la gerencia o altos directivos.



# Objetivos de un Data Warehouse

1. Proporcionar *acceso* a datos organizacionales o corporativos: desde equipo personal, conexión inmediata, por demanda y con alto desempeño, con herramientas sencillas de manejar.
2. Los datos contenidos en un Data Warehouse son *consistentes*: mismas consultas dan mismos resultados
3. Los datos contenidos en un Data Warehouse pueden ser *separados y combinados* a través de cualquier posible medida/indicador del negocio (slice and dice)





# Objetivos de un Data Warehouse

4. El Data Warehouse provee herramientas de consulta, análisis y presentación de datos.
5. El Data Warehouse es el sitio donde se publican los datos utilizados por los usuarios (similitud a la función editorial de revista o de libros)
6. La calidad de los datos en el Data Warehouse hace las veces de motor de impulso de la reingeniería del negocio.



# Características de un Data Warehouse

- Típicamente implementado en un servidor independiente
- Dimensión del tiempo:
  - Representación de la historia de la organización a través de fotografías instantáneas (estados discretos) de los sistemas transaccionales
- Lo impráctico del modelo de datos entidad-relación para Data Warehouses
  - Complejidad de las estructuras de datos
  - Costo de relacionar múltiples tablas con miles o millones de registros
  - Complejidad para realizar consultas con atributos dimensionables (Ej. para tiempo: por semana, por mes, por año, por trimestre, por semestre).



# Características de un Data Warehouse

- Se guarda toda la información útil (proveniente de varias fuentes) en un único lugar
- Cargas masivas de datos en horarios “muertos” para los OLTP
- Bases de datos de gran tamaño.
- Integración de datos internos (80%) y datos externos (20%)



# Diferencias entre OLTP vs OLAP

- On-Line transaction processing (OLTP) vs Data Warehouse (OLAP On-Line Analytical processing):
  - Orientado al soporte de la operación diaria vs orientado a la toma de decisiones
  - Orientación a procesos vs orientación a sujetos/entidades
  - Consistencia a nivel micro vs consistencia a nivel global
  - Carga de transacciones (por segundo vs por unidad)
  - Nivel de granularidad en el manejo de datos de usuario (típicamente por registro/cuenta vs por conjunto respuesta)
  - Normalizado en su mayoría vs violación frecuente de as formas normales
  - Transacción vs “production data load”
  - Operación de rollback vs recarga de la imagen previa (snapshot) del sistema completo





Derechos Reservados 2020 Tecnológico de Monterrey  
Prohibida la reproducción total o parcial de esta obra sin expresa  
autorización del Tecnológico de Monterrey.

Vicerrectoría Académica y de Innovación Educativa