

CI2355 – Almacenes de datos y OLAP

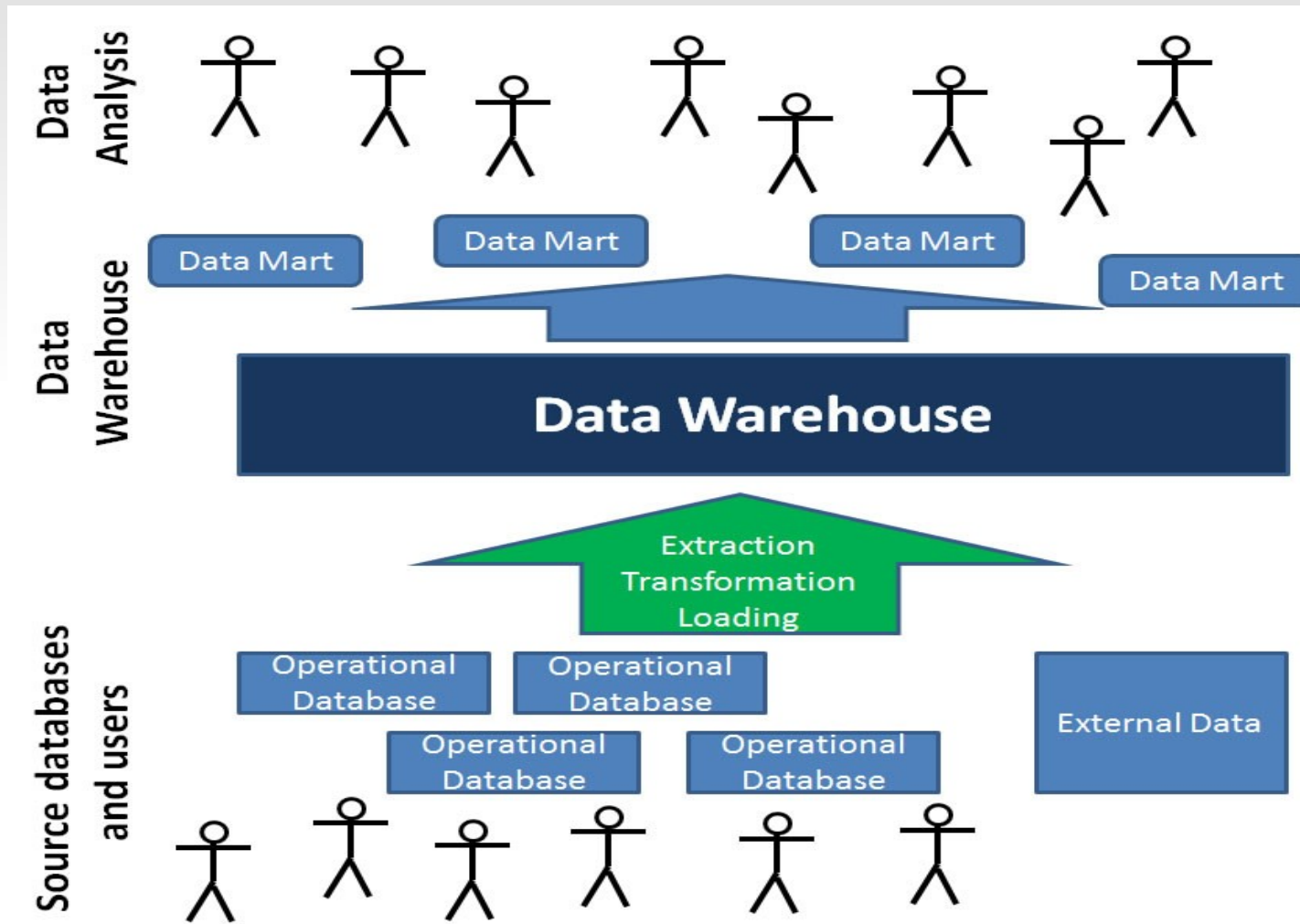


UNIVERSIDAD DE
COSTA RICA

CI2355 – Almacenes de datos y OLAP

- Profesor:
 - Diego Villalba
 - Casillero # 19
- Horario:
 - Lunes, 17:00 a 20:50

Introducción



Almacenes



Orígenes

- BI (1958)
- DSS
- Barry Devlin y Paul Murphy
 - IBM
 - 1988
 - Business Data Warehouse: '...el almacén lógico de toda la información necesaria para hacer reportes en un negocio'
- Bill Inmon y Ralph Kimball

Características de un AD - 1

- Orientado a temas
 - Todas las entidades y eventos relacionados con un tema (eg ventas)
- Variante en el tiempo
 - Todos los cambios son registrados para poder reportar cambios a través del tiempo
- No volátil
 - Cuando los datos son ingresados, no se sobrescriben ni borran

Características de un AD - 2

- Integrado
 - Contiene datos de múltiples fuentes
- A través de los años, algunas partes de esta descripción han sido criticadas (especialmente el aspecto de *no volátil*) y algunos autores han hecho diversas propuestas y variantes, por lo que existen diferentes arquitecturas y formas de modelar el AD. Sin embargo, en lo que la mayoría de los autores concuerdan es en la necesidad de tener repositorio de datos separado para las operaciones de análisis y reportes

Repositorio dedicado

- Justificación:
 - Asegurar que el rendimiento de los sistemas de producción no sea afectado por las consultas o análisis
 - Requerir que la información que los usuarios finales necesitan no cambie mientras la utilizan, es decir, datos en un punto en el tiempo

¿Para qué necesitamos AD?

- Toda la información en un solo lugar
- Información actualizada
- Rápido acceso
- No hay límites de espacio
- Datos históricos disponibles
- Fácil de comprender
- Definiciones claras y uniformes
- Datos estandarizados

Diferentes enfoques

- Almacenes de datos
- Mercados de datos (data marts) con dimensiones conformadas

- Enfoque centralizado
- Enfoque descentralizado / iterativo

- Modelo normalizado
- Modelo dimensional

De-normalización

country
country_id: SMALLINT(5) [PK]
country_name: VARCHAR(50) [AK]
ISO3166_country_code: CHAR(2) [AK]



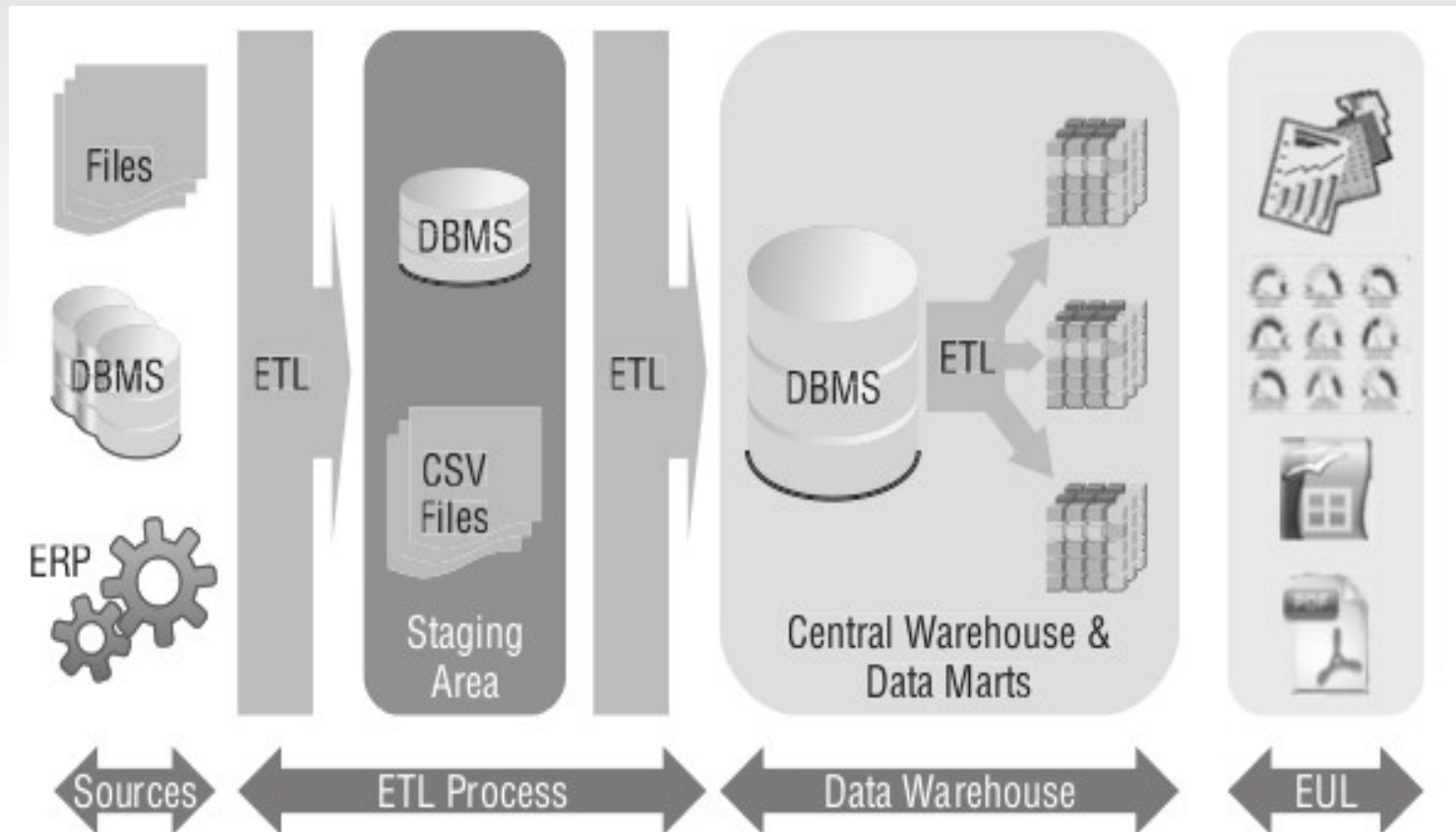
region
region_id: INT(8) [PK]
country_id: SMALLINT(5) [FAK]
region_name: VARCHAR(50) [AK]
ISO3166_region_code: CHAR(2) [AK]



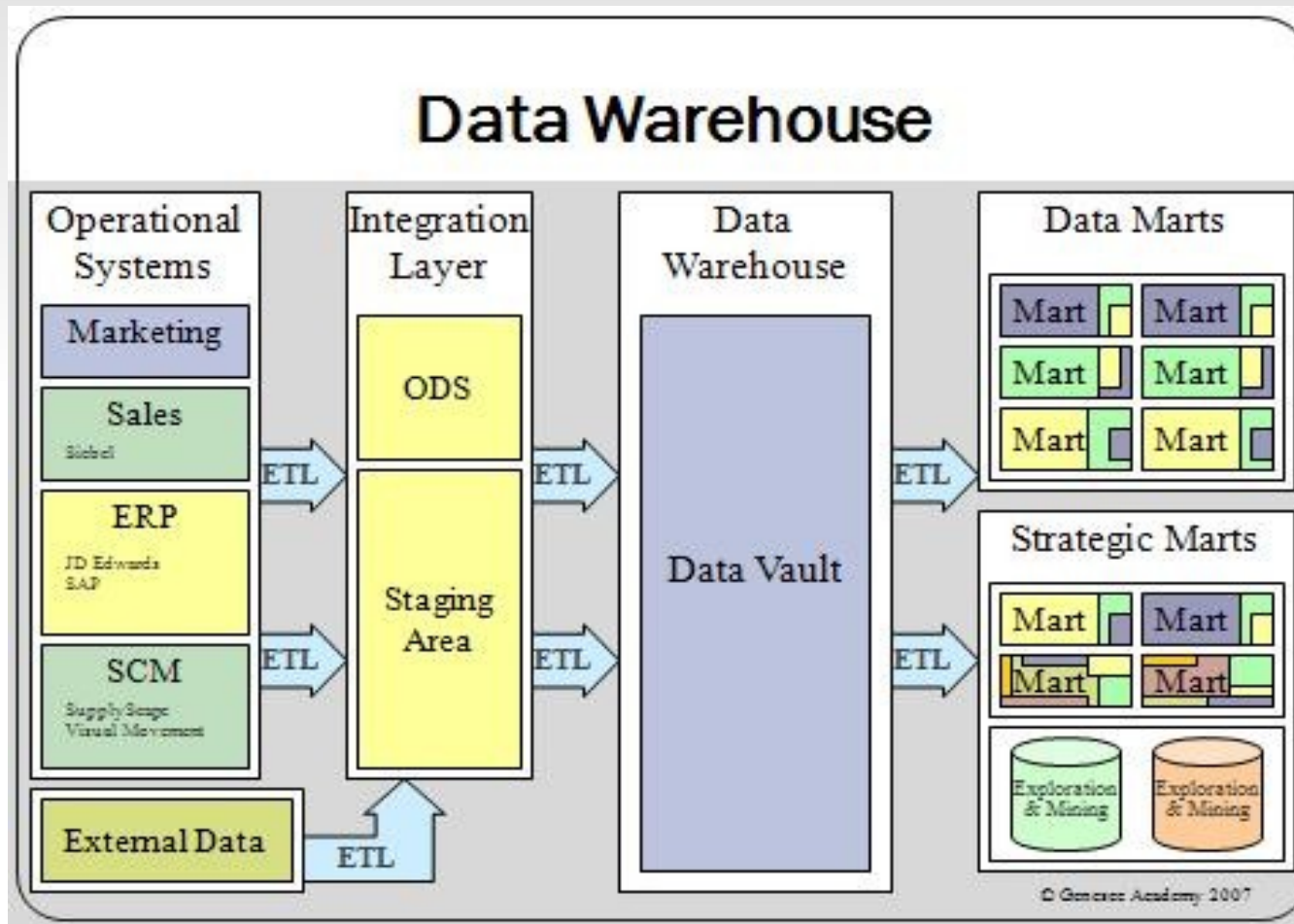
customer_normalized
customer_id: INT(8) [PK]
region_id: INT(8) [FK]

customer_denormalized
customer_id: INT(8) [PK]
ISO3166_region_code: CHAR(2)
region_name: VARCHAR(50)
ISO3166_country_code: CHAR(2)
country_name: VARCHAR(50)

Arquitecturas



Arquitecturas



Capítulo introductorio

- Matteo Golfarelli, Stefano Rizzi. *Data Warehouse Design: Modern Principles and Methodologies*. McGraw-Hill, 2009
- <http://www.mhprofessional.com/product.php?isbn=0071610391>