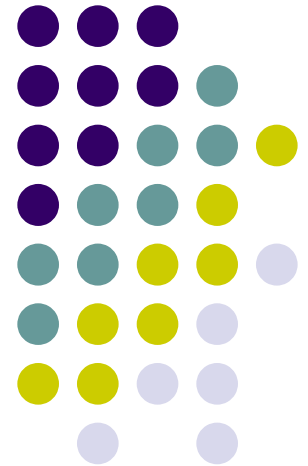


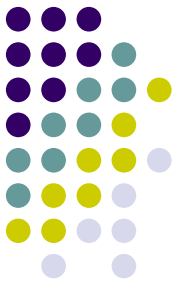
# Data Warehouse Generalidades

## Parte 2:

### Data Warehouse



# Data Warehouse



## Data Marts (DM)

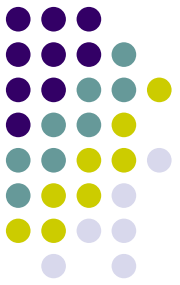
Es un DW restringido a:

- Un **área** funcional
- Un **problema** en particular
- Temas o grupos de **necesidades**

Suelen ser un paso previo al DW porque:

- Minimizar riesgos
- Producir una primera entrega en tiempos razonables
- Si la implementación es exitosa, su alcance se irá ampliando paulatinamente.

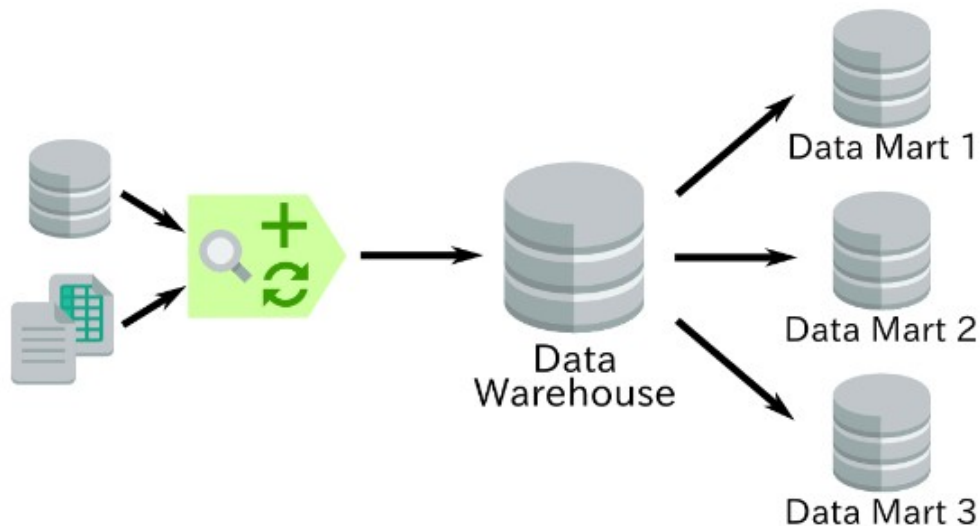
# Data Warehouse



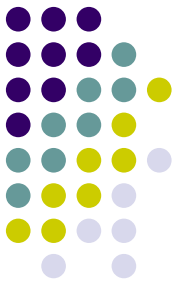
## DM – Arquitecturas de implementación

### Top-Down

Primero se define el DW y luego se desarrollan, construyen y cargan los DM



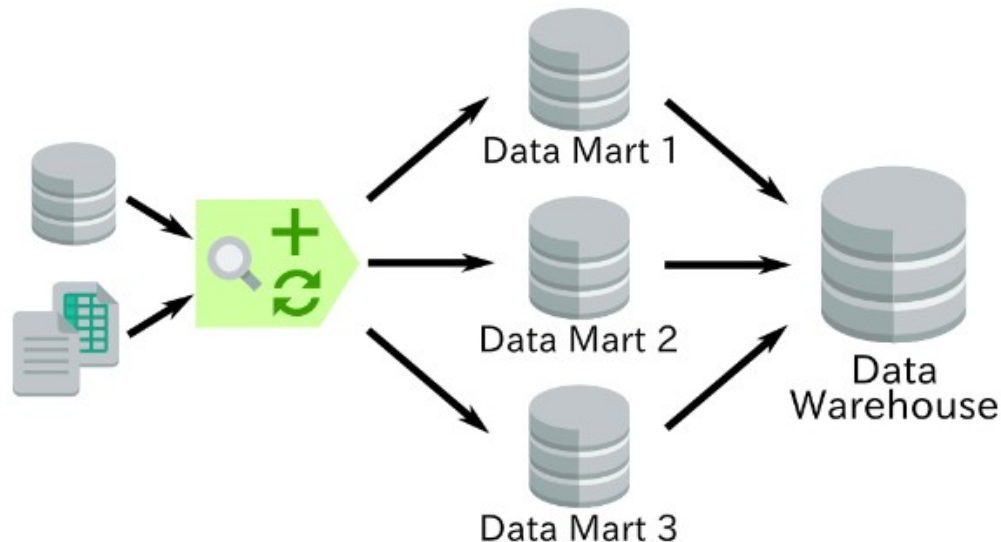
# Data Warehouse



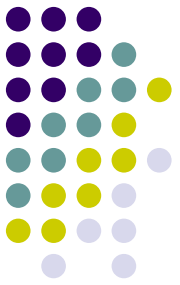
## DM – Arquitecturas de implementación

### Bottom-Up

Se definen inicialmente los DM y luego se integran en un DW centralizado



# Data Warehouse

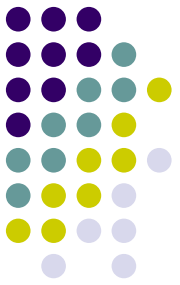


## Data Marts (DM)

### **Beneficios:**

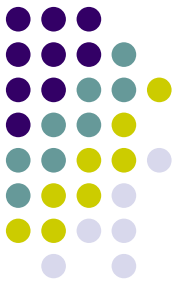
- Son simples de implementar.
- Conllevan poco tiempo de construcción y puesta en marcha.
- Reflejan rápidamente sus beneficios.
- Reducen la complejidad y demanda del DW.
- Pueden tratarse como sub-proyectos (considerando los elementos comunes con otras áreas para no repetir tareas)

# Data Warehouse



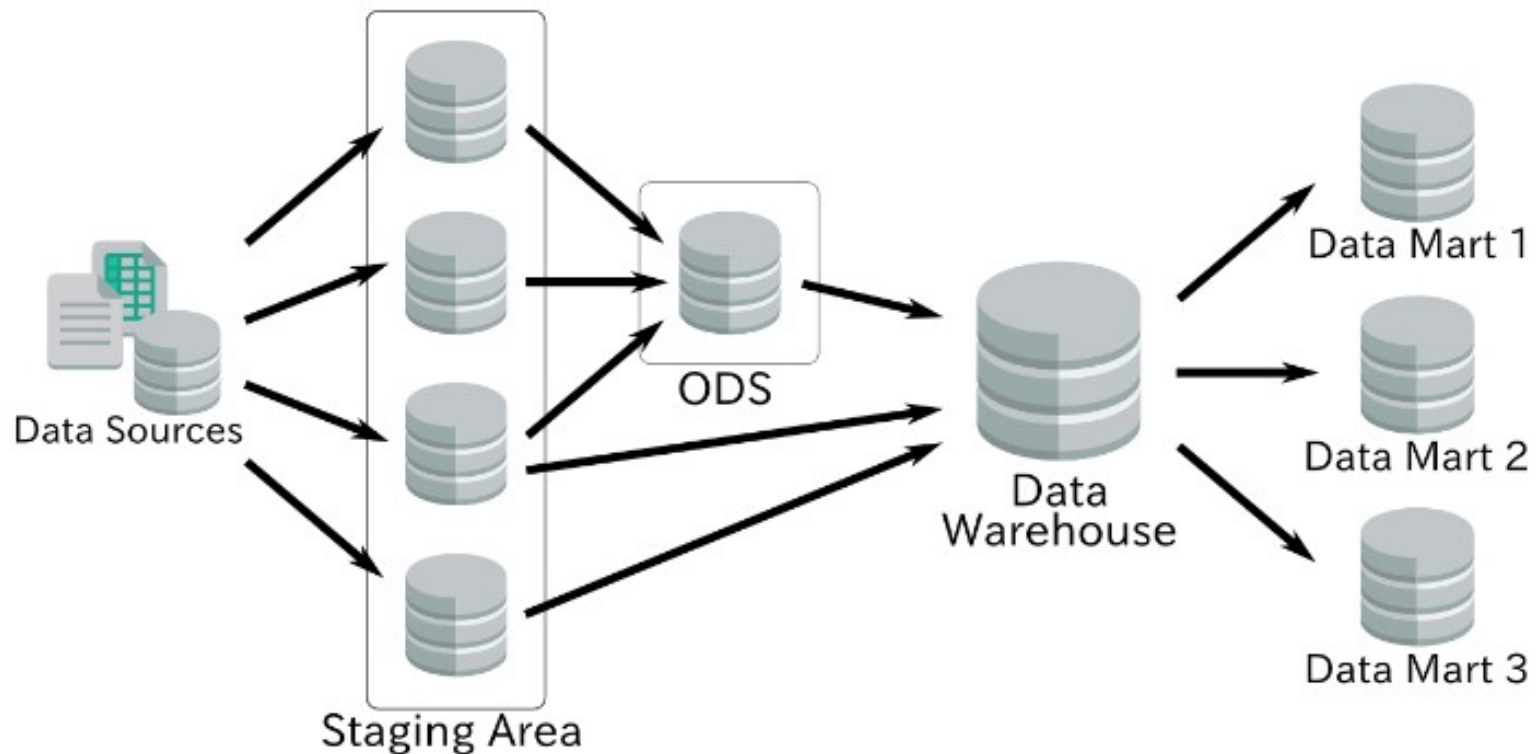
## Áreas de Datos

- Son los entornos por los que han de pasar los datos desde el Data Source hasta llegar al DW o DM.
- Deben considerarse todas las transformaciones que se han de realizar y la integración de datos para conformar el Almacén Corporativo (DW)
- Cada Área se distingue por las funciones que realiza, de qué manera se organizan los datos en la misma, y a qué tipo de necesidad pueden dar servicio.
- No hay una convención estándar sobre lo que abarca exactamente cada Área, y la obligatoriedad de utilizar todas las estudiadas.

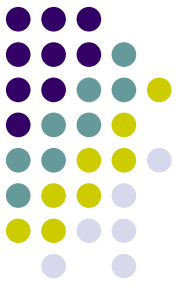


# Data Warehouse

## Áreas de Datos – Staging Area



# Data Warehouse

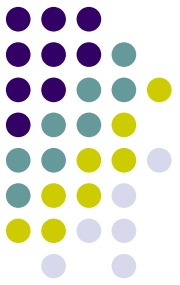


## Áreas de Datos – Staging Area

- **Es un área temporal meramente operativa donde se recogen los datos de los sistemas origen.**
- Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos.
- No se aplican restricciones de integridad ni se utilizan claves.
- Los datos se tratan como si las tablas fueran ficheros planos.
- Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.
- Algunos autores consideran que la Staging Area abarca más de lo comentado, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.



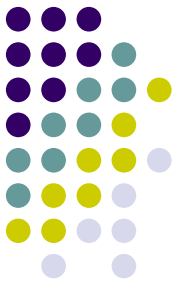
# Data Warehouse



## Áreas de Datos – Staging Area

- Es un área temporal meramente operativa donde se recogen los datos de los sistemas origen.
- **Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos.**
- No se aplican restricciones de integridad ni se utilizan claves.
- Los datos se tratan como si las tablas fueran ficheros planos.
- Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.
- Algunos autores consideran que la Staging Area abarca más de lo comentado, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.

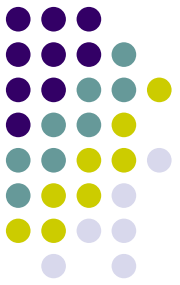
# Data Warehouse



## Áreas de Datos – Staging Area

- Es un área temporal meramente operativa donde se recogen los datos de los sistemas origen.
- Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos.
- **No se aplican restricciones de integridad ni se utilizan claves.**
- Los datos se tratan como si las tablas fueran ficheros planos.
- Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.
- Algunos autores consideran que la Staging Area abarca más de lo comentado, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.

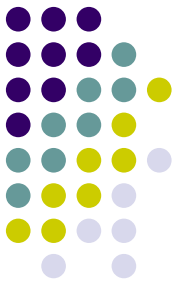
# Data Warehouse



## Áreas de Datos – Staging Area

- Es un área temporal meramente operativa donde se recogen los datos de los sistemas origen.
- Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos.
- No se aplican restricciones de integridad ni se utilizan claves.
- **Los datos se tratan como si las tablas fueran ficheros planos.**
- Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.
- Algunos autores consideran que la Staging Area abarca más de lo comentado, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.

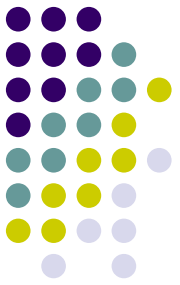
# Data Warehouse



## Áreas de Datos – Staging Area

- Es un área temporal meramente operativa donde se recogen los datos de los sistemas origen.
- Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos.
- No se aplican restricciones de integridad ni se utilizan claves.
- Los datos se tratan como si las tablas fueran ficheros planos.
- **Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.**
- Algunos autores consideran que la Staging Area abarca más de lo comentado, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.

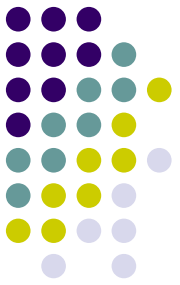
# Data Warehouse



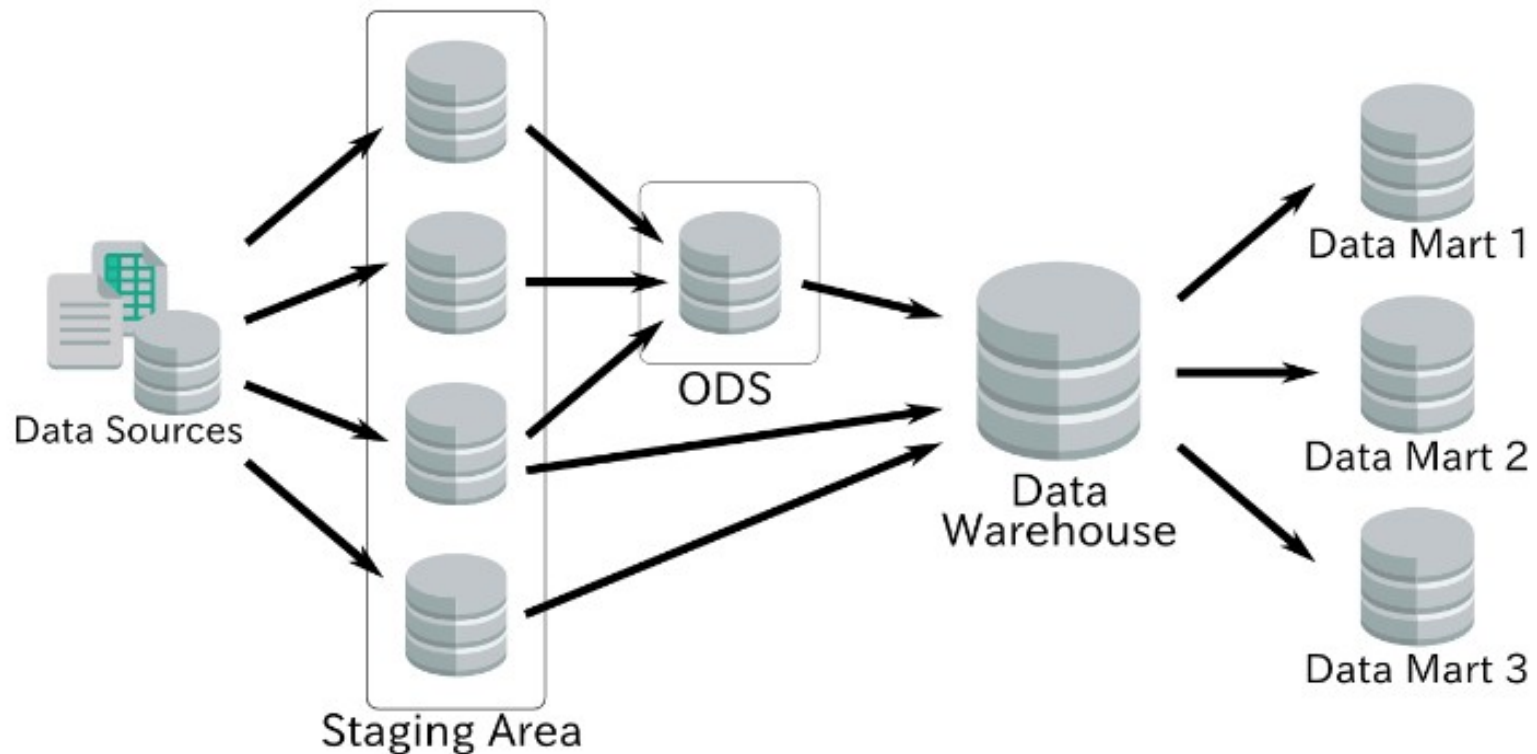
## Áreas de Datos – Staging Area

- Es un área temporal meramente operativa donde se recogen los datos de los sistemas origen.
- Se recogen los datos estrictamente necesarios para las cargas y se aplica el mínimo de transformaciones a los mismos.
- No se aplican restricciones de integridad ni se utilizan claves.
- Los datos se tratan como si las tablas fueran ficheros planos.
- Una vez que los datos han sido traspasados, el DW se independiza de los sistemas origen hasta la siguiente carga. Lo único que se suele añadir es algún campo que almacene la fecha de la carga.
- **Algunos autores consideran que la Staging Area abarca más de lo comentado, o incluso que engloba todo el entorno donde se realizan los procesos de ETL.**

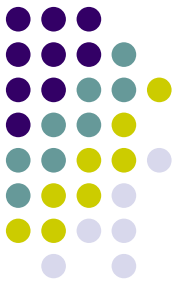
# Data Warehouse



## Áreas de Datos – ODS



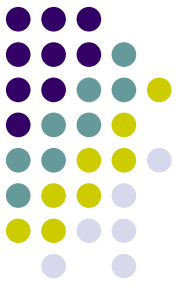
# Data Warehouse



## Áreas de Datos – ODS

- Da soporte a los sistemas operacionales.
- El modelo de datos del Almacén de Datos Operacional (ODS) sigue una estructura relacional normalizada, así cualquier herramienta de **reporting** o **sistema operacional** puede consultar sus datos.
- Está dentro del DW porque se aprovecha el esfuerzo de integración que supone la creación de DW para poder atender también a necesidades operacionales (no es obligatorio).
- No almacena datos históricos, muestra la imagen del momento actual.
- Los datos del ODS se recogen de la Staging Area, y en este proceso sí se realizan transformaciones, limpieza de datos y controles de integridad referencial.
- La actualización de los datos del ODS no es instantánea, los cambios no se ven reflejados hasta que finaliza la carga.

# Data Warehouse



## Áreas de Datos – DW y DM

- El Almacén de Datos Corporativo (DW) contiene datos históricos, y está orientado a la **explotación analítica** de la información.
- Las herramientas DSS o de reporting analítico consultan tanto los Data Marts como el Almacén de Datos Corporativo.
- El DW puede servir para consultas en las que se precisa mostrar a la vez información que se encuentre en diferentes Data Marts.
- En él se almacenan datos que pueden provenir tanto de la Staging Area como del ODS. Si ya se realizan procesos de transformación e integración en el ODS no se repiten para pasarlos al DW. Lo que no se pueda recoger desde el ODS, hay que ir a buscarlo a la Staging Area.
- la mayoría de las tablas han de incorporar columnas para controlar la fecha de carga, la fecha en que se produce un Hecho, o el periodo de validez del registro.