

DW y ETL

FUNDAMENTOS DE ETL





ÍNDICE

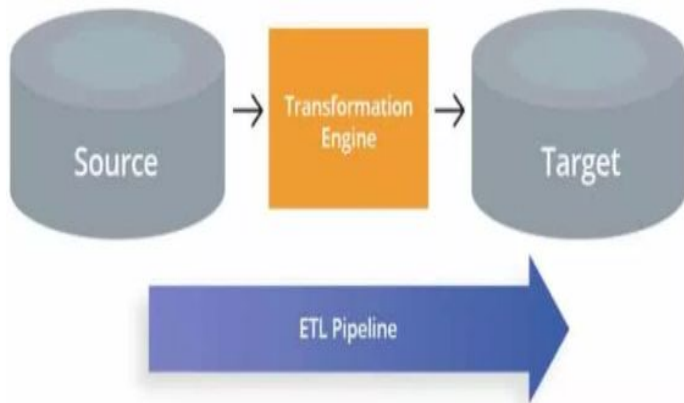
1. ¿QUÉ SIGNIFICA ETL?
2. PROCESOS ETL
3. ¿PARA QUÉ SE UTILIZA?
4. FASE EXTRACCIÓN
5. FASE TRANSFORMACIÓN
6. FASE CARGA
7. ESTRATEGIA

¿QUÉ SIGNIFICA ETL?

EL SIGNIFICADO DE ETL VIENE DE:

1. E: Extraer - **Extract.**
2. T: Transformar - **Transform**
3. L: Cargar - **Load**

PROCESOS ETL



DEFINIMOS **PROCESO ETL** COMO UN PROCESO DE GESTIÓN DE DATOS QUE PERMITE EXTRAER Y MOVER DATOS DESDE MÚLTIPLES FUENTES, TRANSFORMARLOS Y CARGARLOS EN EL DESTINO.

¿PARA QUÉ SE UTILIZA?

En prácticamente cualquier operación empresarial los datos desempeñan un papel importante; para que tengan valor, deben trasladarse y prepararse para su uso, y para ello se necesitan procesos de ETL. Enumeramos algunos casos de aplicación de ETL:

- **Migración de datos** de una aplicación a otra.
- **Replicación de datos** para copias de seguridad o análisis de redundancia.
- **Procesos operativos**, como la migración de datos desde un CRM a un ODS (almacén de datos operativos) para potenciar o enriquecer los datos y luego devolverlos al CRM.
- **Depositar los datos en un almacén de datos** para aprovisionar, clasificar y transformarlos en conceptos de negocio.
- **Migración de aplicaciones de infraestructuras locales a cloud**, cloud híbrida o multicloud.
- **Sincronización de sistemas** clave.

FASE EXTRACCIÓN

El objetivo de un proceso ETL es producir datos limpios y accesibles que puedan utilizarse para analíticas u operaciones comerciales.

- **Analizar el origen de los datos.** Los datos en bruto pueden extraerse de una variedad de fuentes.
- **Extraer los datos** desde los sistemas de origen.
- **Analizar los datos** extraídos.
- **Verificar que los datos extraídos** cumplen los requisitos establecidos en calidad y forma.
- Si fuese necesario, **convertir los datos a un formato preparado** para iniciar el proceso de transformación.

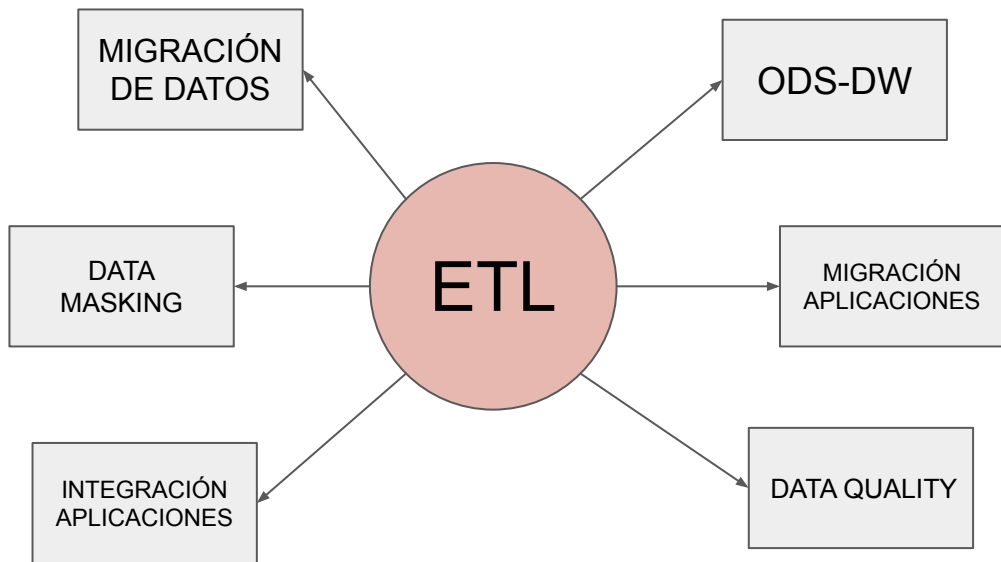
FASE TRANSFORMACIÓN

La fase de transformación de ETL es donde se produce la operación más crítica. La transformación modifica los datos en bruto para que presenten los formatos correctos.

- **Normalización:** Definir qué datos entrarán en juego, cómo se formatean y almacenan, y otras consideraciones básicas que definirán las etapas sucesivas.
- **Eliminación de duplicados:** Excluir y eliminar los datos redundantes.
- **Reglas de Negocio:** Aplicación de las reglas de negocio mediante la transformación de los datos.
- **Clasificación:** Maximizar la eficiencia de los almacenes de datos agrupando y clasificando elementos en colecciones de datos. Las normas de transformación determinarán cómo se clasifica cada dato y dónde se trasladará a continuación.

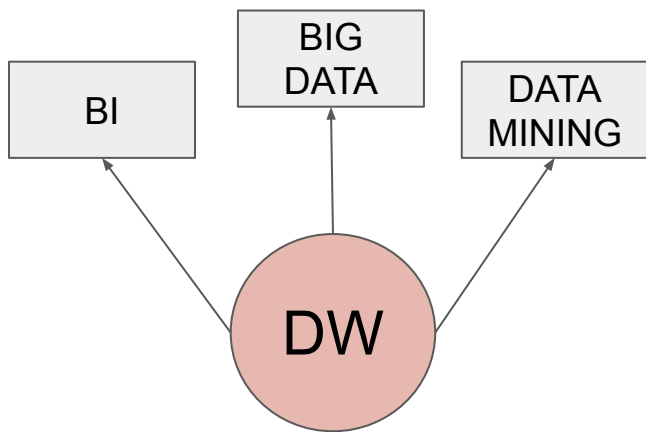
FASE CARGA

La fase de carga de un proceso ETL implica mover los datos procesados al sistema de destino.

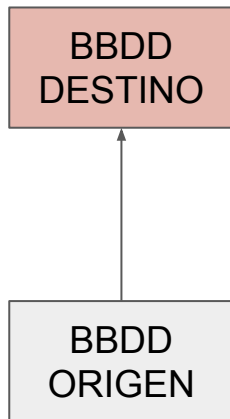


FASE CARGA

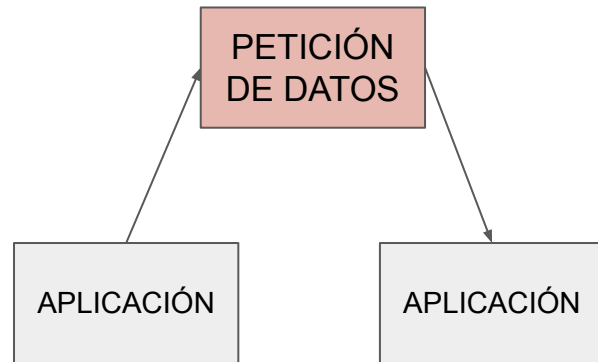
El destino y la estrategia que seguiremos en nuestra carga puede variar en función de nuestro objetivo.



DATAWAREHOUSE



DATA MASKING



**INTEGRACIÓN
REALTIME**

ESTRATEGIA PROYECTO ETL

ANÁLISIS

1. **Definir alcance del proyecto** de integración de datos para disponer de una visión clara de los requisitos que deben cumplir los procesos ETL.
2. **Evaluar y analizar los datos origen.** Ubicación de los datos en origen, análisis de volumetrías

DISEÑO

1. **Diseño de los Modelos Destino.** Diseño del Modelo Lógico y Físico de datos del sistema destino.
2. **Diseño Mapa Fuente-Destino** de los datos. Diseño detallado de los mapas lógicos y técnicos de carga de los campos origen y destino.
3. **Diseño del Plan de Pruebas.**

CONSTRUCCIÓN

1. **Construcción de los procesos ETL.**
2. **Ejecución y carga** de los datos.
3. **Ejecución del Plan de Pruebas** para certificar que los procesos está bien construidos y los datos corresponden a los requisitos establecidos.

IMPLANTACIÓN

1. **Implantación de los procesos ETL.** Traslado de los procesos al entorno de producción (real).
2. **Carga de datos históricos** (si procede)
3. **Planificación y ejecución** de los procesos.

COMENZAMOS

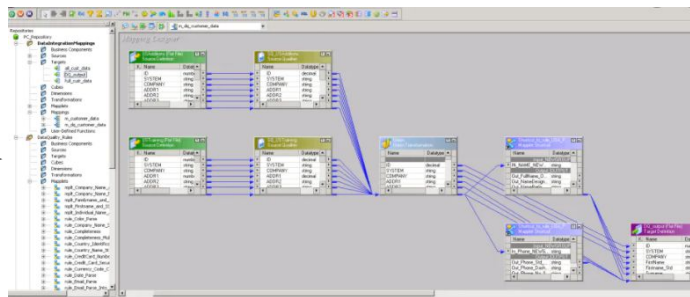
ORIGEN DE DATOS

ETL

ODS-DW

amazon

ORACLE®
DATABASE



Informatica™



PostgreSQL

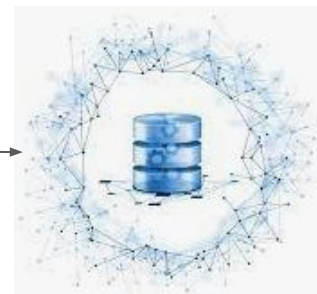
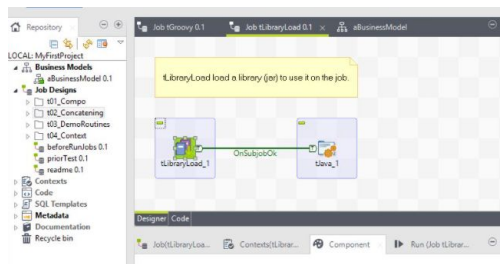
COMENZAMOS

ORIGEN DE DATOS

ETL

ODS-DW

DVD RENTAL DATABASE





KEEPCODING

Tech School

Madrid | Barcelona | Bogotá

Jon Pineda Galvez

jpinedagalvez@gmail.com

<https://www.linkedin.com/in/jonathanpg>