

Ingenierio de Datos (Rol)

Que hace un Data Engineer

Rol Escencial

- Primero se tienen **datos Crutos** datos no procesados o datos con informacion faltante
- Se limpian estos datos o se pasan por filtros
- Se almacenan en bases de datos ya como datos estructurados

Esto se hace con el fin de que **el demas equipo de data scientist** tenga datos con los cuales trabajar y realizar informes

Creacion de Pipelines ETL

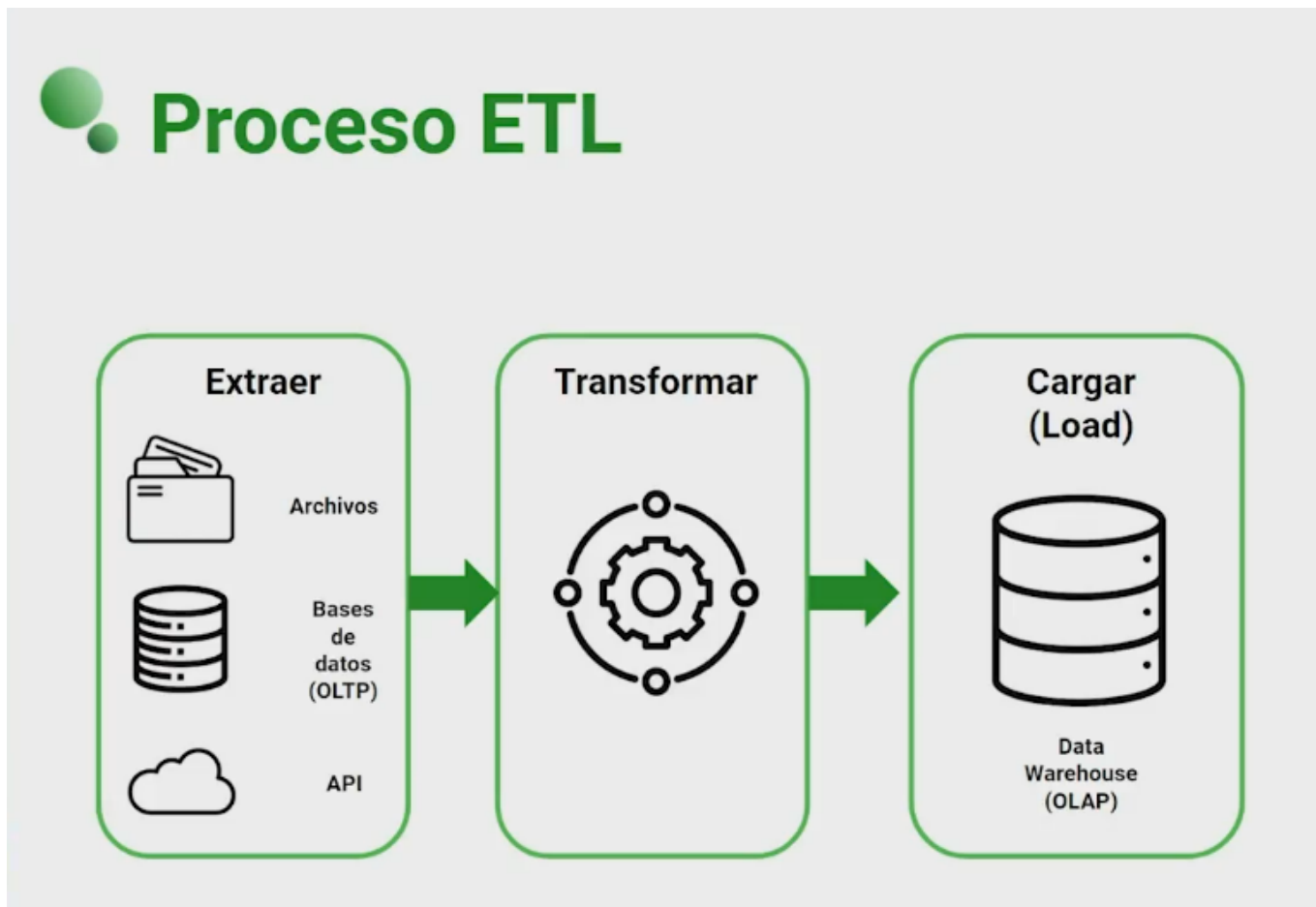
- E = Extraccion
- T = Transformacion
- L = Carga/Load

Trabajo de un Data Engineer

- Se tienen los datos en pipelines ETL e informacion en bases de datos
- Se realiza la informacion de estas fuentes o fuentes externas o internas (**lo importante esque estos datos suman valor a la empresa**)
- Se transforman los datos para su posterior analisis
- Se almacenan en **bases de datos especializadas para el analisis**

Estas bases de datos mencionadas anteriormente son especiales ya que pueden recibir multiples consultas, ya que las bases de datos comunes o (Transaccionales) en su mayoria, se rompen a momento de realizar muchas consultas

- Posterior al almacenado en estas bases de datos, se automatiza el proceso de ETL



Roles Relacionados

- Data Architect: Se encarga de plantear estrategias de datos en la organización
- Big data Architect: Trabaja con Big Data

Herramientas y Tecnologías para Data Engineer

- Lenguaje de Programación python (Programación Orientada a Objetos)
- Scala como lenguaje de Programación
- Sistemas linux

- Bash Scripting para automatizacion
- Jupyter Notabooks (editores deCodigo)
- Manejo de bases de datos SQL y NoSQL (Nivel Avanzado)
- Librerias de procesar datos Pandas, Spark y dask para el procesamiento de datos en mayor escala
- Apache Airflow para la automatizacion de su ejecucion
- Tecnologias Cloud (AWS, Azure, Google Cloud) ya que algunas bases de datos/sistemas no estaran en nuestra pc
- Contenedores Docker (Creacion de contenedores)
- Orquestadores Kubernetes
- Estadistica Descriptiva

Como empezar?

- Python y bases solidas de la ingeniera de Software
- Automatizacion y Scripting (bash)
- Librerias y conocimientos de ellas (python)
- Conocimientos (consultas y guardar datos) en SQL y noSQL