En este curso aprenderás a crear y automatizar un flujo ETL (Extract, Transform and Load) para extraer, transformar y guardar nuestros datos

Vamos a obtener datos y datasets de diferentes fuentes, de diferentes lugares.

Estos casi siempre viene en una estructura que no están listos para el análisis posterior.

Entonces una vez obtenidos los datos se van a procesar, limpiarlos y enriquecerlos para con posterioridad insertarlos dentro de un sistema especializado para el análisis de datos.



Los ingenieros de datos se preocupan principalmente por implementar los pipelines que permiten automatizar la obtención de datos y su posterior limpieza para que otros profesionales de los datos, puedan realizar su labor.

**PIPELINES:** La definición de pipelines es tuberías o caminos, cuando tu programas o haces análisis de datos tienes que transformar una entrada en una salida, esta transformación puede ser tan sencilla como la suma de dos valores o compleja como extraer valores, hacer algunos cálculos con estos cálculos hacer otra cosas hasta tener la salida de datos esperada, estas transformaciones se hacen usando funciones, este paso entre funciones de los resultados es conocido como pipelines.

Los ingenieros de datos son aquellos que saben realizar software de gran escala, llevarlo a producción.

**Archivos del curso:**