

Curso Data Engineer: Creando un pipeline de datos

Módulo E - Clase 11



Práctica Arquitectura

Ejercicio de arquitectura



Te acaban de contratar en una empresa de la industria minera como Data Engineer/Data Architect para delinear su arquitectura y sugerir qué herramientas deberían utilizar para ingestar la data, procesar la información, almacenarla en un datawarehouse, orquestar y realizar Dashboards que ayuden a la toma de decisiones basadas en datos.

Luego de realizar algunas reuniones con el team de analitica de la empresa pudimos relevar:

Sistema ERP: SAP con una base de datos Oracle

Sistema de Producción: App desarrollada "in house" con una base de datos Postgres.

Fuentes externas: un proveedor que realiza algunas mediciones de la calidad de las rocas le deja todos sus análisis en un bucket de AWS S3 con archivos Avro.

Mediciones en tiempo real: Utilizan +100 sensores de mediciones de vibración por toda la mina para detectar movimiento del suelo y se podrían utilizar para predecir posibles derrumbes.

Apps Mobile: La empresa cuenta con una app mobile donde trackean todos los issues pendientes con maquinaria de la mina.

Ejercicio de arquitectura



Desarrollar una arquitectura, que sea escalable, robusta, que sea orquestada automáticamente, que contemple seguridad, calidad, linaje del dato, que sea utilizada para procesar tanto información batch como información en tiempo real.

Responder las siguientes preguntas:

1. Utilizarían infraestructura on premise o en la nube?
2. ETL o ELT? Por qué?
3. Que herramienta/s utilizarían para ETL/ELT?
4. Que herramienta/s utilizarían para Ingestar estos datos?
5. Que herramienta/s utilizarían para almacenar estos datos?
6. Como guardarán la información, OLTP o OLAP?
7. Que herramienta/s utilizarían para Data Governance?
8. Data Warehouse, Data Lake o Lake House?
9. Qué tipo de información gestionarán, estructurada, semi estructurada, no estructurada?
10. Con que herramienta podrían desplegar toda la infraestructura de datos?