

Proyecto: Arquetipo de Procesamiento Avanzado

Descripción

Se requiere el desarrollo de un proceso ETL utilizando la plataforma Databricks, asegurando un alto rendimiento y escalabilidad en el procesamiento de datos. Este desarrollo deberá implementarse exclusivamente en Apache Spark con Scala, por lo que no se aceptarán implementaciones en PySpark ni Spark SQL.

Objetivo

El objetivo del proyecto es diseñar e implementar un flujo de procesamiento de datos eficiente, basado en las mejores prácticas de ETL (Extract, Transform, Load), optimizando el manejo de grandes volúmenes de información y garantizando la calidad e integridad de los datos.

Alcance

El participante deberá construir el proceso siguiendo la arquitectura establecida, asegurando la correcta integración y transformación de los datos dentro del entorno Databricks. Se evaluará el uso adecuado de Scala y Spark en cada etapa del desarrollo, así como el rendimiento y la optimización del procesamiento de datos.

Requisitos Técnicos

- Implementación en Apache Spark con Scala.
- Uso de la plataforma Databricks.
- Diseño y desarrollo del flujo ETL basado en la arquitectura proporcionada.
- Aplicación de buenas prácticas en procesamiento de datos y optimización de rendimiento.

Se espera que el proceso desarrollado cumpla con los estándares de calidad y eficiencia requeridos, garantizando su correcta ejecución dentro del ecosistema de Data Engineering.

