

Introducción a Databricks y el ecosistema Spark

Laboratorio: Pipeline de Datos en Databricks con Arquitectura Medallón



IMMUNE
TECHNOLOGY INSTITUTE



Laboratorio: Pipeline de Datos en Databricks con Arquitectura Medallón

- **Objetivo:**
 - Aprender sobre la arquitectura de datos en capas usando un pipeline de ingesta estructurado en tres niveles: **Bronze**, **Silver**, y **Gold**.
- **Datos:**
 - Base de datos de pilotos y carreras de **Fórmula 1 (1950 – 2023)**.
- **Estructura del Laboratorio:**
 1. **Notebook Bronze:** Ingesta de datos crudos desde la fuente.
 2. **Notebook Silver:** Limpieza y transformación inicial de los datos para mejorar la calidad.
 3. **Notebook Gold:** Preparación final de datos listos para consumo y análisis.
- **Paso Final:**
 - Construcción de visualizaciones para representar gráficamente la información

