

Introducción a Databricks y el ecosistema Spark

Jesús Arnau Villar



IMMUNE
TECHNOLOGY INSTITUTE



Agenda



Introducción (30 min)

- Breve historia del procesamiento de datos y Big Data.
- Evolución de Apache Spark y sus beneficios en el mundo del Big Data.
- Fundación y desarrollo de Databricks: historia, capitalización, número de clientes, valoración y ranking.



Databricks Onboarding(60 min)

- Qué es Databricks y cómo se diferencia de otros entornos de Spark.
- Arquitectura general y componentes clave.
- Interfaz de usuario y navegación básica.
- **Lab 1: Databricks Workspace UI**
- **Break (15 min)**

Agenda



Databricks - Data Engineering (90 min)

- Repaso conceptos básicos de Spark
- Introducción a Delta Lake: conceptos y beneficios.
- Gestión de metadatos y versionado de datos con Delta Lake.
- Orquestación de procesos
- Lab 2: Ingesta de datos en Delta Lake



Databricks – Data Analytics (60 min)

- Herramientas de análisis de datos en Databricks.
- Creación de dashboards y visualizaciones interactivas.
- Integración con herramientas externas de BI y visualización.
- Lab 3: Análisis de un conjunto de datos y creación de visualizaciones y dashboards.

Agenda



Conclusión (30 min)

- Repaso de los conceptos clave cubiertos en el workshop.
- Mejores prácticas para aprovechar Databricks en proyectos de Big Data.
- Recursos adicionales para seguir aprendiendo.
- Q&A

Recursos Adicionales



Slides y Código de los laboratorios

- Github Repo:
 - [jarnawer/databricks_workshop \(github.com\)](https://github.com/jarnawer/databricks_workshop)
- Databricks Community Edition:
- [Login - Databricks Community Edition](#)

Jesus Arnau Villar



Senior Cloud Data Engineer



Apache Spark 3 Certified



Databricks Solutions Architect Champion



<https://www.linkedin.com/in/jesus-arnau-84494441/>



IMMUNE
TECHNOLOGY INSTITUTE

