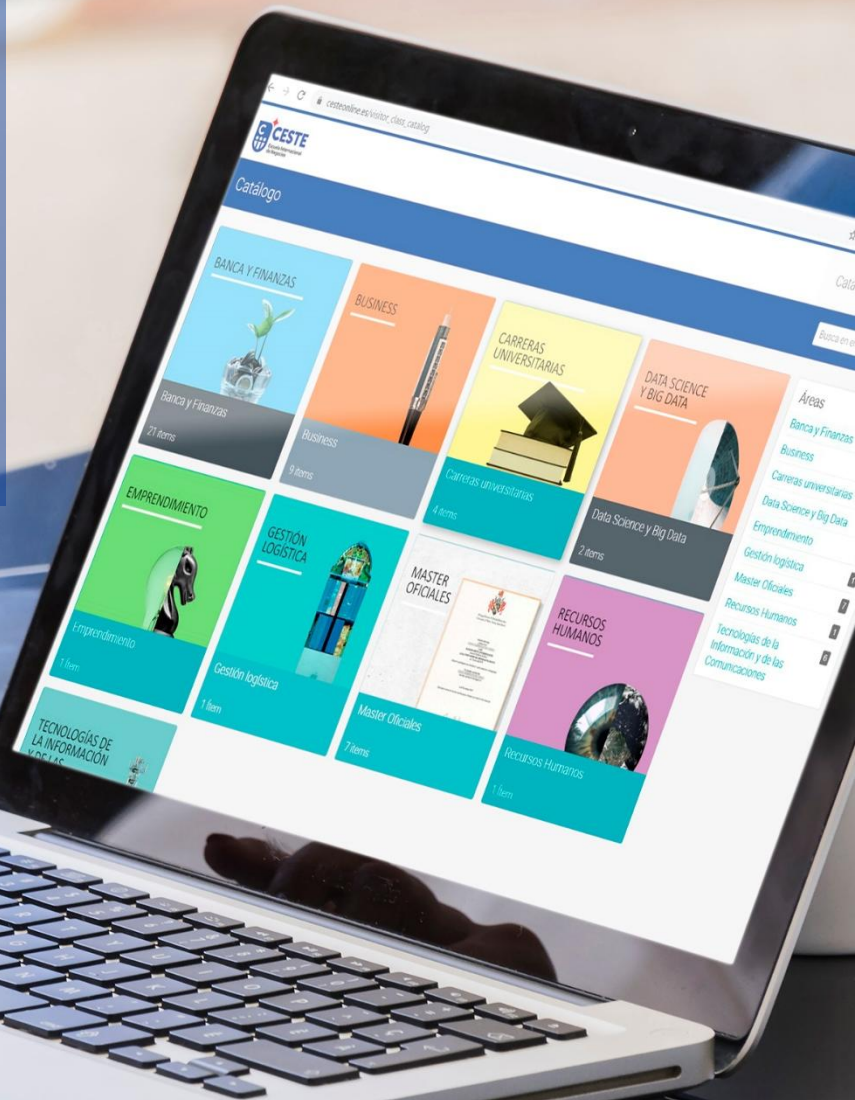


CESTE

Escuela Internacional de Negocios

Zaragoza (España)



Introducción a Apache Spark: RDDs, DataFrames, Datasets y Spark SQL



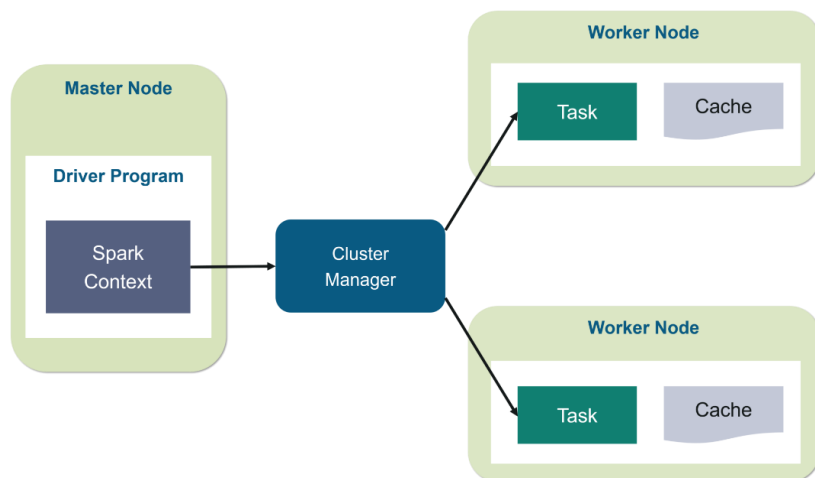
databricks

Objetivos de la sesión



- Entender Spark y sus principales componentes.
- Uso de Notebooks en Databricks.
- Uso de dbutils.
- Ejercicios sobre Spark.

¿Qué es Apache Spark?



- Motor unificado de analítica para procesar datos a gran escala
- Motor de procesamiento distribuido
- Permite desarrollo en Java, Scala, Python y R.
- Crea DAGs para programar y orquestar nodos de trabajo

Estructuras de datos en Spark



RDDs



DataFrames



Datasets

Práctica con Notebook



databricks

Próxima sesión



Introducción a Unity Catalog



Configuraciones básicas Unity Catalog



Análisis exploratorio en Databricks



Práctica: Creación de Dashboards y análisis exploratorio.

Preguntas / Sugerencias



www.ceste.es