

Databricks

- Plataforma de Análise de Dados
- Na nuvem
- Permite engenharia e ciência de dados



Baseado em Nuvem como Platform as a Service (PaaS)

- Não possui infraestrutura na Nuvem
- Você deve escolher qual infraestrutura usar
- Por exemplo, se usar AWS:
 - EC2
 - S3



Spark

- Plataforma de computação em cluster e tolerante a falhas
- Open Source
- Em memória
- Possui componentes como SQL, Mllib e GraphX
- Suporta variedade de linguagens: Java, Scala, Python, R, SQL



Qual a relação?

- Databricks foi fundado pelos criadores originais do Apache Spark
- Objetivo é simplificar e potencializar o uso do Spark
- Databricks é uma plataforma comercial
- Spark é um produto open source
- O uso de Spark não é exclusivo do Databricks

