

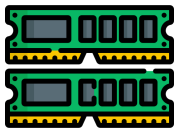
Visión general de Apache Spark

- **Marco de procesamiento distribuido para big data**



Está diseñado para manejar grandes volúmenes de datos

- **Caché en memoria y ejecución de consultas optimizadas**



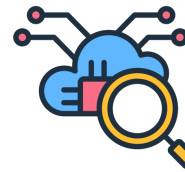
Permite realizar operaciones de datos a alta velocidad en comparación con MapReduce tradicional

- **Soporte multilinguaje**



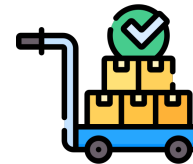
Soporta varios lenguajes de programación como Java, Scala, Python y R

- **Reutilización de código en múltiples tareas**



Procesamiento por lotes, consultas interactivas, análisis en tiempo real, aprendizaje automático y procesamiento de gráficos

- **No está destinado para OLTP**



Su enfoque principal es el procesamiento de datos en lotes y en tiempo real para análisis

- **Spark Streaming**



Procesamiento de datos en tiempo real. Se integra fácilmente con Amazon Kinesis y Apache Kafka