

Componentes de Apache Spark



Permite a los desarrolladores ejecutar consultas SQL en los datos almacenados en DataFrames y Datasets



Permite el procesamiento de flujos de datos en tiempo real



Proporciona varias implementaciones de algoritmos ML, incluyendo clasificación, regresión, clustering, filtrado colaborativo y minería de patrones



Permite a los usuarios de R aprovechar la capacidad de procesamiento distribuido de Spark para manipular y analizar grandes conjuntos de datos