



## DIFERENCIAS

### Modelo de Computación

**Hadoop** utiliza el modelo de programación **MapReduce**, mientras que **Spark** utiliza el modelo de programación Resilient Distributed Datasets (**RDD**). El modelo RDD es más rápido que MapReduce porque los datos se almacenan en caché en la memoria y se procesan en paralelo.

### Velocidad de procesamiento

Spark es más rápido que Hadoop debido a su capacidad para procesar datos en memoria.

### Soporte para Múltiples Lenguajes

Spark admite varios lenguajes, como Java, Python, Scala y R, mientras que Hadoop admite principalmente Java.

### Abstracción de Datos

Spark proporciona una abstracción de datos más alta que Hadoop. Por ejemplo, Spark tiene DataFrames y Datasets, mientras que Hadoop solo tiene MapReduce.

### Recuperación de Fallos

Ambos marcos tienen tolerancia a fallos incorporada. Sin embargo, Spark es más rápido en la recuperación de fallos debido a su modelo RDD.

## SIMILITUDES

### Escalabilidad.

Ambos marcos son altamente escalables y pueden manejar grandes conjuntos de datos.

### Procesamiento Distribuido.

Ambos marcos utilizan el procesamiento distribuido para procesar grandes conjuntos de datos.

### Ecosistema de Componentes.

Ambos marcos tienen un ecosistema rico y diverso de componentes para procesar datos.

### Integración con Almacenamiento.

Ambos marcos pueden integrarse con diferentes sistemas de almacenamiento como HDFS, S3, etc.

## REFERENCIAS

[Hadoop y Spark: diferencia entre los marcos de Apache. AWS \(amazon.com\)](#)