

Apache Spark is a lightning real-time processing Framework.

What are the main features of Apache Spark?

What are the main features of Apache Spark?

What is a Resilient Distribution Dataset in Apache Spark?

What is a Transformation in Apache Spark?

What are security options in Apache Spark?

How will you monitor Apache Spark?

What are the main libraries of Apache Spark?

What are the main functions of Spark Core in Apache Spark?

How will you do memory tuning in Spark?

What are the two ways to create RDD in Spark?

What are the main operations that can be done on a RDD in Apache Spark?

What are the common Transformations in Apache Spark?

What are the common Actions in Apache Spark?

What is a Shuffle operation in Spark?

What are the operations that can cause a shuffle in Spark?

What is purpose of Spark SQL?

What is a DataFrame in Spark SQL?

What is a Parquet file in Spark?

What is the difference between Apache Spark and Apache Hadoop MapReduce?

What are the main languages supported by Apache Spark?

What are the file systems supported by Spark?

What is a Spark Driver?

What is an RDD Lineage?

What are the two main types of Vector in Spark?

What are the different deployment modes of Apache Spark?

What is lazy evaluation in Apache Spark?

What are the core components of a distributed application in Apache Spark?

What is the difference in cache() and persist() methods in Apache Spark?

How will you remove data from cache in Apache Spark?

What is the use of SparkContext in Apache Spark?

Do we need HDFS for running Spark application?

What is Spark Streaming?

How does Spark Streaming work internally?

What is a Pipeline in Apache Spark?

How does Pipeline work in Apache Spark?

What is the difference between Transformer and Estimator in Apache Spark?