

What is shuffling in map reduce?

What is the difference between an HDFS Block and Input Split?

What are the mapfiles in Hadoop?

Why do we need a password-less SSH in Fully Distributed environment?

What is the use of .pagination class?

Why is Replication pursued in HDFS in spite of its data redundancy?

What are the core components of Hadoop?

What differentiates Hadoop from other parallel computing solutions?

What is Difference between Secondary namenode, Checkpoint namenode & backupnod Secondary Namennode, a poorly named component of hadoop.

What happens when a datanode fails ?

What are the Side Data Distribution Techniques.

Explain what is Sequencefileinputformat?

What is shuffleing in mapreduce?

What is partitioning?

Explain what happens in textinformat ?

Can we change the file cached by Distributed Cache

What if job tracker machine is down?

Explain what does the conf.setMapper Class do ?

Can we deploy job tracker other than name node?

What are the four modules that make up the Apache Hadoop framework?

How did you debug your Hadoop code?

Which modes can Hadoop be run in? List a few features for each mode.

What is Hadoop Streaming?

Where are Hadoop's configuration files located?

What is a combiner in Hadoop?

What is the functionality of JobTracker in Hadoop? How many instances of a JobTracker run on Hadoop cluster?

List Hadoop's three configuration files.

Is it necessary to write jobs for Hadoop in Java language?

What are “slaves” and “masters” in Hadoop?

What commands are used to see all jobs running in the Hadoop cluster and kill a job in LINUX?

How many datanodes can run on a single Hadoop cluster?

What is identity mapper?

What is job tracker in Hadoop?

How many job tracker processes can run on a single Hadoop cluster?

What is Identity reducer?

What sorts of actions does the job tracker process perform?

What is Commodity Hardware?

How does job tracker schedule a job for the task tracker?

What are the main components of Job flow in YARN architecture ?

What does the mapred.job.tracker command do?

What are the Main configuration parameters that user need to specify to run Mapreduce Job ?

What is “PID”?

What is “jps”?

Is there another way to check whether Namenode is working?

How would you restart Namenode?

What is Chain Reducer ?