21/05/2021

Quiz
Answers

1). __Hadoop__ is a Top level project, open-source implementation of frameworks for reliable, scalable, distributed computing and data storage.

2). __HDFS__ is the storage layer, __YARN__ is the resource management layers and __MapReduce__ is data processing layer of hadoop.

3). Name node is the __Master Node__ in the Apache Hadoop HDFS Architecture that maintains and manages the blocks present on the Data nodes __Slave node__

4). The __Secondary Namenode__ works concurrently with the primary Name node as a helper daemon.

5). Default size of each block in Apache Hadoop 2.x is __128MB__

6). I have a file "Example .txt" of size 514MB. If default configuration of block size, which is 128MB, Then, how many blocks will be created. __5__ 514/128 = 4.015, complete 4 & 1 just filled.

7). If you are storing a file of 100MB in HDFS using the default Configuration, and if it perform replication management policy then how much of data is residing in different data node __300MB__ 100 MB × 3 (replication factor)

8. _Hadoop Streaming_ is the utility that enables us to create or run Map reduce scripts in any language either java or non-Java, as mapper / reducer.

9. Hadoop framework is written in which language _Java_.

10. Hadoop YARN Research Manager (RM)

   → RM manages the global assignments of resources (cpu and memory) among all the applications. It at arbitrates system resources between competing applications.

11. Moving data in and out of Hadoop, is referred as data Ingress & egress.