21 02 2018

Big Data Hadoop Certification Training

Module 1: Understanding Big Data and Hadoop

Assignment

⇒ Data to store = 100 TB
RAM = 8 GB
HDD = 10 TB
Read-write speed = 100 MB/s
Replication factor = 3
Block size = 64 MB

Reading 100 TB data at a speed of 100 MB/s using only 1 node = 1048576 sec

So data to read in 5 minutes = 300 sec

Let x be the number of data nodes required to read 100 TB of data in 5 minutes (300 sec)

Since, we are reading 100TB of data

$$x \times 300 = 1 \times 1048576$$
$$x = 1048576 / 300$$
$$= 3495.25 \approx 3495$$

∴ To read 100TB data in 5 mins would require 3495 data nodes.