

# Module 1: Understanding Big Data and Hadoop

---

## Assignment Solution

edureka!

**edureka!**

© 2014 Brain4ce Education Solutions Pvt. Ltd.

# Introduction

This problem test your understanding of how distributed file system works and process data. Once the concept is clear, the solution requires only few calculations.

## 1. Problem Solution

### 1.1 Time required to read the data using Single DataNode

1 DataNode takes:

Total Data/Read –Write speed  
(100TB \* 1024\*1024) in MB / 100MB/s  
1048576 seconds or 291.77 hours to read 100TB Data

### 1.2 DataNodes required to read the data in FIVE minutes

Number of DataNodes required to read 100TB in 5 minutes

Time taken by 1 DataNode to read the 100TB data / Total time given to finish the read  
= ( 1048576 secs/60)/5 minutes  
= 3495.253333 DataNodes

**So, you would need ~ 3495 such DataNodes to read the 100TB data in 5 minutes**