



**LOYOLA COLLEGE (AUTONOMOUS), CHENNAI – 600 034**

**M.Sc. DEGREE EXAMINATION – COMPUTER SCIENCE**

**THIRD SEMESTER – NOVEMBER 2018**

**17PCS3MC01– BIG DATA ANALYTICS**

Date: 25-10-2018

Dept. No.

Max. : 100 Marks

Time: 09:00-12:00

**PART-A**

**Answer ALL questions**

**(10×2 = 20)**

1. What is the difference between analytics and reporting?
2. What is Prediction Error?
3. What is Archival Storage?
4. When do we use the Working Storage?
5. What do you mean by Job Tracker and Task Tracker?
6. What is the normal capacity of Hard Disk and RAM in a Slave Node?
7. What do you mean by striping of data blocks?
8. What do you mean by Hadoop Benchmarks?
9. What is HiveQL?
10. What do you mean by IBM Big Insights?

**PART-B**

**Answer ALL questions**

**(5×8 = 40)**

11. a. Briefly explain the need for Big Data Analytics.  
(or)  
b. Briefly explain the Variety and Value aspects of big data.
12. a. Briefly explain the Stream Data Management system.  
(or)  
b. What is an ad-hoc query? Give an example.
13. a. Briefly explain the Master/Slave architecture of MapReduce and HDFS.  
(or)  
b. Explain the main purposes of Sqoop with examples.
14. a. Briefly explain how Hadoop is installed in "roll your own" Hadoop approach.  
(or)  
b. Briefly explain the Hadoop components included in BigInsights 1.2 version.
15. a. How LOAD and TRANSFORM commands are used in Pig?  
(or)  
b. Explain the three steps involved in using BigSheets to perform big data analytics.

**PART-C**

**Answer any TWO questions**

**(2×20 = 40)**

16. a. Briefly explain the process Of sampling and filtering of stream data.  
b. Briefly explain how input splits are used in managing the distributed data.
17. a. Briefly explain the various tools available in the Hadoop ecosystem.  
b. Explain the Shuffle and Sort phases with the help of a suitable example.
18. a. Explain how the execution of a MapReduce program is monitored.  
b. Briefly explain the salient features of ZooKeeper and HBase.

\*\*\*\*\*