

M.C.A. - II (Two Year Programme) (New CBCS Pattern) Semester-III
PMCAT303 - Paper-III : Big Data & Hadoop

P. Pages : 2

Time : Three Hours



GUG/W/24/13649

Max. Marks : 80

-
- Notes :
1. All questions are compulsory and carry equal marks.
 2. Draw neat and labeled diagram and use supporting data wherever necessary.
 3. Avoid vague answers and write specific answers related to questions.

Either :

1. a) What is Big Data? Explain 4 V's in detail. 8
- b) Explain the drivers for Big Data and its application in detail. 8

OR

- c) Explain the Hadoop ecosystem and how data is moved in and out of Hadoop. 8
- d) Describe the concept of Mapreduce and the importance of data serialization in detail. 8

Either :

2. a) Explain the Hadoop architecture and the function of HDFS. 8
- b) Explain the role of NameNode, Secondary NameNode and DataNode in Hadoop. 8

OR

- c) Describe in detail about Mapreduce paradigm and the task of Job and Task Tracker. 8
- d) Explain how Hadoop cluster is set up using SSH and configuration. 8

Either :

3. a) Explain various components of the Hadoop ecosystem and their functions. 8
- b) Explain the new features in Hadoop 2.0 including Namenode high availability and HDFS Federation. 8

OR

- c) Describe in detail about YARN architecture and its benefits over MRV₁. 8
- d) Explain the role of schedulers like fair and capacity in Hadoop. 8

Either :

4. a) Give the detailed comparison between Hive architecture with traditional databases. 8
- b) Explain how HiveQL is used for querying, sorting and aggregating data. 8

OR

- c) What are the key features of HBase, including a cluster? Explain how it is used in HBase. 8
- d) How does zookeeper help in monitoring a cluster? Explain its use in HBase. 8

5. Solve all the questions.

- a) Explain the role of distributed file system in Big Data. 4
- b) List common Hadoop shell commands. 4
- c) Discuss how is MRV₁ run in YARN. 4
- d) Describe the purpose of MapReduce scripts in HiveQL. 4
