

- Q.29 Discuss statistical model in "R"? (CO5)  
 Q.30 Explain how Hadoop Clusters works? (CO2)  
 Q.31 Explain Hadoop Map Reduce Word-Count problem? (CO3)  
 Q.32 Explain Apache Hadoop and Hadoop Ecosystem.(CO2)  
 Q.33 Write and Explain Daemon of Hadoop with their functionality? (CO2)  
 Q.34 Explain reading and getting data into "R". (CO5)  
 Q.35 Explain Shuffling and Sorting in Map Reduce? (CO3)

#### Section-D

**Note: Long answer questions. Attempt any two question out of three Questions. (2x10=20)**

- Q.36 a) Explain in detail about HIVE? (CO4)  
 b) Explain Hadoop Architecture? (CO1)  
 Q.37 a) Explain ordered & unordered factor in "R"? (CO1)  
 b) Explain Map Reduce Architecture? (CO3)  
 Q.38 a) Explain the utility of popular Big Data tools with examples? (CO1)  
 b) Explain the structured life cycle approach to analyze Big Data and create statistical model? (CO5)

No. of Printed Pages : 4

180854B/170854B

Roll No. ....

#### 5th Sem., Branch : Comp Subject : Big Data

Time : 3 Hrs.

M.M. : 100

#### SECTION-A

**Note: Multiple type Questions. All Questions are compulsory. (10x1=10)**

- Q.1 Hadoop is distributed under which License? (CO1)  
 a) Apache License 2.0 b) Mozilla Public License  
 c) Sharware d) Commercial  
 Q.2 What is Hadoop? (CO1)  
 a) A programming language  
 b) A distributed storage and processing framework  
 c) A relational database management system  
 d) An operating system  
 Q.3 Which programming language is commonly used to write Map Reduce programs in Hadoop? (CO2)  
 a) Java b) Python  
 c) C++ d) Ruby  
 Q.4 Which of the following platforms does Hadoop run on? (CO2)  
 a) Bare metal b) Debian  
 c) Cross-platfrom d) Unix-Like

- Q.5 Hadoop is a framework that works with a variety of related tools. Common cohorts / groups include \_\_\_\_\_. (CO3)
- MapReduce, Hive and HBase
  - MapReduce, MySQL and Google Apps
  - MapReduce, Hadoop and HBase
  - MapReduce, Hadoop and HBase
- Q.6 \_\_\_\_\_ has the world's largest Hadoop cluster. (CO2)
- Apple
  - Datamatics
  - Facebook
  - None of the above
- Q.7 The Pig Latin scripting language is not only a higher-level data flow language but also has operators similar to \_\_\_\_\_. (CO4)
- SQL
  - JSON
  - XML
  - All of the mentioned
- Q.8 Which of the following phases occur simultaneously? (CO5)
- Shuffle and Sort
  - Reduce and Sort
  - Shuffle and Map
  - All of the mentioned
- Q.9 Input of the \_\_\_\_\_ is the sorted output of the mappers. (CO5)
- Reducer
  - Mapper
  - Shuffle
  - All of the mentioned
- Q.10 The daemons associated with the Map Reduce phase are \_\_\_\_\_ and task-trackers. (CO5)
- Job tracker
  - Map tracker
  - Reduce tracker
  - All of the above

## Section-B

**Note: Objective type questions. All questions are compulsory. (10x1=10)**

- Q.11 What is HDFS? (CO2)
- Q.12 Map is Map Reduce? (CO3)
- Q.13 What is YARN? (CO2)
- Q.14 What is Big Data? (CO1)
- Q.15 List best practices of Big Data Analytics? (CO1)
- Q.16 List any two Big Data Platforms? (CO4)
- Q.17 Write challenges for processing Big Data? (CO1)
- Q.18 Write any two tools of Big Data? (CO2)
- Q.19 Write any two HIVE components? (CO4)
- Q.20 Write four computing resources of Big Data Storage? (CO5)

## Section-C

**Note: Short answer type Question. Attempt any twelve questions out of fifteen Questions. (12x5=60)**

- Q.21 Explain the Hadoop Distributed File System? (CO1)
- Q.22 Explain the characteristics of Big Data? (CO1)
- Q.23 Discuss the Evolution of Big Data? (CO1)
- Q.24 Explain in detail about HDFS? (CO2)
- Q.25 Write a note on Pig Data Flow Engine? (CO4)
- Q.26 Explain Hive with traditional RDBMS? (CO4)
- Q.27 Discuss five Vs in Big Data? (CO1)
- Q.28 Explain lists and data frames in "R"? (CO5)