

**PARUL UNIVERSITY**  
**FACULTY OF ENGINEERING & TECHNOLOGY**  
**B.Tech., Summer 2018 - 19 Examination**

**Semester: 8**  
**Subject Code: 03105480**  
**Subject Name: Big Data Analytics**

**Date: 01/05/2019**  
**Time: 10:30 am To 1:00 pm**  
**Total Marks: 60**

---

**Instructions:**

1. All questions are compulsory.
2. Figures to the right indicate full marks.
3. Make suitable assumptions wherever necessary.
4. Start new question on new page.

---

**Q.1 Objective Type Questions (All are compulsory) (Each of one mark) (15)**

1. Point out the correct statement :
  - a) Hadoop do need specialized hardware to process the data
  - b) Hadoop 2.0 allows live stream processing of real time data
  - c) In Hadoop programming framework output files are divided in to lines or records
  - d) None of the mentioned
2. Hadoop is a framework that works with a variety of related tools. Common cohorts include:
  - a) MapReduce, Hive and HBase
  - b) MapReduce, MySQL and Google Apps
  - c) MapReduce, Hummer and Iguana
  - d) MapReduce, Heron and Trumpet
3. Point out the wrong statement :
  - a) Hadoop processing capabilities are huge and it's real advantage lies in the ability to process terabytes & petabytes of data
  - b) Hadoop uses a programming model called "MapReduce", all the programs should confirms to this model in order to work on Hadoop platform
  - c) The programming model, MapReduce, used by Hadoop is difficult to write and test
  - d) All of the mentioned
4. \_\_\_\_\_ systems are scale-out file-based (HDD) systems moving to more uses of memory in the nodes.
  - a) NoSQL
  - b) NewSQL
  - c) SQL
  - d) All of the mentioned
5. All of the following accurately describe Hadoop, EXCEPT:
  - a) Open source
  - b) Real-time
  - c) Java-based
  - d) Distributed computing approach
6. HDFS provides a command line interface called \_\_\_\_\_ used to interact with HDFS.
  - a) "HDFS Shell"
  - b) "FS Shell"
  - c) "DFS Shell"
  - d) None of the mentioned
7. The daemons associated with the MapReduce phase are \_\_\_\_\_ and task-trackers.
  - a) job-tracker
  - b) map-tracker
  - c) reduce-tracker
  - d) all of the mentioned
8. The \_\_\_\_\_ machine is a single point of failure for an HDFS cluster.
9. Pig operates in mainly how many nodes ?
  - a) Two
  - b) Three
  - c) Four
  - d) Five
10. Hive uses \_\_\_\_\_ for logging.
  - a) logj4
  - b) log4l
  - c) log4i
  - d) log4j

11. Which of the following is not a NoSQL database?
  - a) SQL Server
  - b) MongoDB
  - c) Cassandra
  - d) None of the mentioned
12. Point out the correct statement :
  - a) Hive is not a relational database, but a query engine that supports the parts of SQL specific to querying data
  - b) Hive is a relational database with SQL support
  - c) Pig is a relational database with SQL support
  - d) All of the mentioned
13. Which hdfs command is used to check for various inconsistencies?
  - a) fsk
  - b) fsck
  - c) fetchdt
  - d) none of the mentioned
14. For YARN, the \_\_\_\_\_ Manager UI provides host and port information.
  - a) Data Node
  - b) NameNode
  - c) Resource
  - d) Replication
15. \_\_\_\_\_ command display the list of databases in MongoDB.

**Q.2 Answer the following questions. (Attempt any three)**

**(15)**

- A) Explain horizontal scaling and vertical scaling. Which is used for Hadoop? Justify your answer.
- B) What is the difference between Hadoop and Traditional RDBMS?
- C) Discuss the following terms
  - a. NodeManager
  - b. Job History Server
  - c. Datanode
  - d. Namenode
- D) Discuss fair scheduling in YARN.

**Q.3 A) Introduce Hadoop ecosystem and all of its components.**

**(07)**

- B) Explain Sharding and Replication in NoSQL.

**(08)**

**OR**

- B) Explain anatomy of File Read in Hadoop.

**(08)**

**Q.4 A) Explain five V's of Big data with proper examples.**

**(07)**

**OR**

- A) Define HDFS and YARN, and talk about their respective components.

**(07)**

- B) Explain HIVE architecture

**(08)**