

**GUJARAT TECHNOLOGICAL UNIVERSITY****MCA Integrated- SEMESTER– VIII EXAMINATION – WINTER 2019****Subject Code: 4480601****Date:15/11/2019****Subject Name: Big Data****Time: 02:30 PM TO 05:00 PM****Total Marks: 70****Instructions:**

1. Attempt all questions.
2. Make suitable assumptions wherever necessary.
3. Figures to the right indicate full marks.

- Q.1** (a) i) Define Terms: DataNode, NameNode, Cluster, Big Table **04**  
 ii) Differentiate between File Systems and distributed file system. **03**
- (b) **Say TRUE or FALSE with justification.** **07**  
 i) MapReduce works on the principle of divide-and-conquer.  
 ii) MongoDB scales horizontally using Replication.  
 iii) MapReduce cannot be used for batch processing of data and aggregation operations.  
 iv) There is no similarity between SQL and NoSQL databses.  
 v) Hadoop can be used as an ETL engine.  
 vi) A fundamental principle related to Big Data processing is the CAP principle.  
 vii) Big data characteristic Value refers to the quality or fidelity of data.
- Q.2** (a) i) Differentiate between RDBMS and NoSQL Databases. **04**  
 ii) Why Document databases are schema-less in form? **03**
- (b) Justify statement: Developments in ICT have accelerated the pace of Big Data adoption in businesses. **07**
- OR**
- (b) State the stages of Big Data analytics lifecycle. **07**
- Q.3** (a) What is Big Data? List and explain characteristics of big data with example **07**  
 (b) Describe the Speed, Consistency and Volume (SCV). why only two out of the three SCV properties are simultaneously supportable. **07**
- OR**
- Q.3** (a) How Sharding and Replication can be combined benefit fault tolerance, availability and scalability? **07**  
 (b) Describe the Consistency, Availability, and Partition tolerance (CAP) theorem. why only two out of the three CAP theorem properties are simultaneously supportable. **07**
- Q.4** (a) Discuss A/B Testing, Correlation and Regression using suitable example. **07**  
 (b) What is MapReduce? Write a Mapper and Reducer function for word count. **07**
- OR**
- Q.4** (a) Discuss Classification, Clustering and Filtering Machine learning techniques. **07**  
 (b) What is the objective of MapReduce? Explain the key considerations when developing a MapReduce algorithm. **07**
- Q.5** (a) Explain how to access data from MongoDB, Hbase and Redis databases. **07**  
 (b) Explain mongoimport, mongoexport and mongodump utility. **07**
- OR**
- Q.5** (a) Explain the concepts “task parallelism” and “data parallelism” using suitable example. Which of these concepts is adopted by MapReduce? How? **07**

- (b) MongoDB : Create collection Election with fields : SeatID, SeatName, PartyName, CandidateName, Gender, Age
1. List all documents containing Patel in their CandidateName.
  2. Take backup of collection.
  3. Count no. of Party wise candidates..
  4. Create index on PartyName.
  5. Arrange documents in ascending order of names.
  6. Display 4th document to 5th document.
  7. Remove documents whose Age is more than 65

\*\*\*\*\*