

GUJARAT TECHNOLOGICAL UNIVERSITY
BE – SEMESTER 7 (NEW SYLLABUS) EXAMINATION- SUMMER 2018

Subject Code: 2171607**Date: 28-04-2018****Subject Name: BIG DATA ANALYTICS (Department Elective-II)****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) What is Big Data? Explain how big data processing differs from distributed processing. **03**
- (b) List various application of big data. How it can be used to improve business for a superstore. **04**
- (c) Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail. **07**

- Q.2** (a) Explain Avro data serialization technique in MapReduce. **03**
- (b) Explain characteristics of Big Data. **04**
- (c) What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem. **07**

OR

- (c) What is data serialization? With proper examples discuss and differentiate structured, unstructured and semi-structured data. Make a note on how type of data affects data serialization. **07**
- Q.3** (a) Explain following commands with syntax and at least one example of each. (1) copyFromLocal (2) showing the content of outputfile. **03**
- (b) Explain “Map Phase” and “Combiner Phase” in MapReduce. **04**
- (c) Write Map Reduce steps for counting occurrences of specific numbers in the input text file(s). Also write the commands to compile and run the code. **07**

OR

- Q.3** (a) List various configuration files used in Hadoop Installation. What is use of mapred-site.xml? **03**
- (b) Explain “Shuffle & Sort” phase and “Reducer Phase” in MapReduce. **04**
- (c) Write Map Reduce steps for counting sum of numbers in the input text file(s). Also write the commands to compile and run the code. **07**
- Q.4** (a) What is Zookeeper? What are the benefits of Zookeeper? **03**
- (b) Draw architecture of APACHE PIG and explain in short. **04**
- (c) Define HDFS. Discuss the HDFS Architecture and HDFS Commands in brief. **07**

OR

- Q.4** (a) What is HBase? Write a query to create a table in HBase. **03**
- (b) Discuss role of Data node and Name node in HDFS. **04**

- (c) Draw and explain Architecture of APACHE HIVE. Explain various data insertion techniques in HIVE with example. **07**

[P.T.O]

- Q.5** (a) Explain following in brief with respect to Mongo DB : **03**
1) Collections and documents
2) Indexing and retrieval
- (b) Write difference between MangoDB and Hadoop. **04**
- (c) What is NoSQL database? List the differences between NoSQL and relational databases. Explain in brief various types of NoSQL databases in practice. **07**

OR

- Q.5** (a) Explain scaling in MangoDB. **03**
- (b) Explain CRUD operations in MongoDB. **04**
- (c) What is Resilient Distributed Dataset in Apache Spark? Explain in detail. **07**
Make a note on why RDD is better than Map Reduce data storage?
