

Seat No.: _____

Enrolment No. _____

GUJARAT TECHNOLOGICAL UNIVERSITY

MCA – SEMESTER – V • EXAMINATION – SUMMER 2018

Subject Code: 3650010

Date: 09-May-2018

Subject Name: Big Data Tools

Time: 10.30 am to 1.00 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 data? List and explain types of Data with e.g. 07
(b) What is big data? List and explain characteristics of big data in detail. 07
- Q.2** (a) Explain Hadoop Ecosystem in detail with diagram. 07
(b) What is NoSQL? Why we need NoSQL? List and explain types of NoSQL databases in detail with e.g. 07
- OR**
- (b) What is NewSQL? Why we need NewSQL. Give difference between SQL, NoSQL and NewSQL. 07
- Q.3** (a) Give differences : 07
1. Hadoop 1.x and Hadoop 2.x
2. Traditional BI and Big Data
(b) Explain HDFS read and write process in detail with diagram. 07
- OR**
- Q.3** (a) What is Map – Reduce? Explain Map Reduce – Mapper task and Reducer task in detail with e.g. 07
(b) Explain NameNode, Secondary NameNode, DataNode , Job Tracker and Task Tracker in detail. 07
- Q.4** (a) Explain count, limit, skip and sort in MongoDB with syntax and e.g. 07
(b) What is Pig? How pig jobs are running on cluster? Give some overview of Pig Latin. 07
- OR**
- Q.4** (a) What is HQL? Explain following with syntax and e.g. of HQL : 07
1. Static partition
2. Dynamic partition
3. Bucketing
(b) What is Hive? Explain Hive architecture in detail. Give difference between Pig and Hive: 07
- Q.5** (a) What is spark? Explain RDD in detail by giving difference of interactive operation on Map Reduce and interactive operation on RDD. 07
(b) Write Pig Script for word count. 07
- OR**
- Q.5** (a) Write short note on MLlib, Tensor Flow and Theone. 07
(b) Write Hive queries for word count. 07
