Seat No.: _____ Enrolment No._____

GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VII(NEW) • EXAMINATION – WINTER 2016

Subject Code:2171607 Date:25/11/2016

Subject Name:Big Data Analytics(Departmental Elective - II)

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 Big Data? Explain characteristics of Big Data.	**07**
	(b)	What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.	**07**
Q.2	**(a)**	Write a short note on NoSQL databases. List the differences between NoSQL and relational databases?	**07**
	(b)	What are the benefits of Big Data? Discuss challenges under Big Data. How Big Data Analytics can be useful in the development of smart cities.(Discuss one application)	**07**
		OR	
	(b)	Explain working of following phases of Map Reduce with one common example. (i) Map Phase (ii) Combiner Phase (iii) Shuffle and Sort Phase (iv) Reducer Phase	**07**
Q.3	**(a)**	Explain working of Hive with proper steps and diagram.	**07**
	(b)	(i) What is Zookeeper? List the benefits of it. (ii) Differentiate: Apache pig Vs Map Reduce.	**03** **04**
		OR	
Q.3	**(a)**	What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example.	**07**
	(b)	(i) Explain Metastore in Hive. (ii) Explain Storage mechanism in HBase. Compare Raw oriented and Column Oriented database structures.	**03** **04**
Q.4	**(a)**	Write Map Reduce code for counting occurrences of specific words in the input text file(s). Also write the commands to compile and run the code.	**07**
	(b)	Explain Spark components in detail. Also list the features of spark.	**07**
		OR	
Q.4	**(a)**	Explain Replication and scaling feature of MongoDB.	**07**
	(b)	Draw HDFS Architecture. Explain any two commands of HDFS from following commands with syntax and al least one example of each. (i) copyFromLocal (ii) setrep (iii) checksum	**07**
Q.5	**(a)**	What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset? Explain RDDs in detail.	**07**
	(b)	Explain CRUD operations in MongoDB.	**07**
		OR	

Q.5 **(a)** Explain following for MongoDB **07**

(i) Indexing

(ii) Aggregation

(b) Explain Job Scheduling in Map Reduce. How it is done in case of **07**

(i) The Fair Scheduler

(ii) The Capacity Scheduler

****