## **GUJARAT TECHNOLOGICAL UNIVERSITY** BE - SEMESTER-VII(NEW) • EXAMINATION – WINTER 2016

Subject Code:2171607Date:25/11/2016Subject Name:Big Data Analytics(Departmental Elective - II)Time:10.30 AM to 1.00 PMInstructions:1. Attempt all questions.					
	<b>1.</b> <b>2.</b> 3.	Attempt all questions. Make suitable assumptions wherever necessary. Figures to the right indicate full marks.			
Q.1	(a) (b)	What is Big Data? Explain characteristics of Big Data. What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.	07 07		
Q.2	<b>(a)</b>	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		
	( <b>b</b> )	OR Explain working of following phases of Man Badwas with one common	07		
	(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) (b)	<ul><li>Explain working of Hive with proper steps and diagram.</li><li>(i) What is Zookeeper? List the benefits of it.</li><li>(ii) Differentiate: Apache pig Vs Map Reduce.</li></ul>	07 03 04		
Q.3	(a)	<b>OR</b> What do you mean by HiveQL Data Definition Language? Explain any three	07		
	( <b>b</b> )	<ul><li>HiveQL DDL command with its syntax and example.</li><li>(i) Explain Metastore in Hive.</li></ul>	03		
		(ii) Explain Storage mechanism in HBase. Compare Raw oriented and Column Oriented database structures.	04		
Q.4	<b>(a)</b>	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>(b)</b>	Explain Spark components in detail. Also list the features of spark. OR	07		
Q.4	(a) (b)	Explain Replication and scaling feature of MongoDB. Draw HDFS Architecture. Explain any two commands of HDFS from following commands with syntax and al least one example of each. (i) copyFromLocal	07 07		
		(ii) setrep (iii) checksum			
		(III) Checksum			
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		

Q.5	(a)	Explain following for MongoDB	07
-		(i) Indexing	
		(ii) Aggregation	
	<b>(b)</b>	Explain Job Scheduling in Map Reduce. How it is done in case of	07
		(i) The Fair Scheduler	
		(ii) The Capacity Scheduler	

\*\*\*\*\*