

**GUJARAT TECHNOLOGICAL UNIVERSITY****BE - SEMESTER-VII (NEW) EXAMINATION – WINTER 2017****Subject Code: 2171607****Date:02/11/2017****Subject Name: Big Data Analytics(Departmental Elective - II)****Time: 10:30 AM TO 01:00 PM****Total Marks: 70****Instructions:**

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

		MARKS
<b>Q.1</b>	(a) How Big Data Analytics can be useful in the development of smart cities.	<b>03</b>
	(b) What is Big data? Discuss it in terms of volume and velocity.	<b>04</b>
	(c) What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram.	<b>07</b>
<b>Q.2</b>	(a) Explain Storage mechanism in HBase.	<b>03</b>
	(b) In Map Reduce how Job Scheduling is done in case of the Fair Scheduler?	<b>04</b>
	(c) Draw and explain HDFS Architecture. Explain the function of NameNode and DataNode. What is a Secondary Namenode? Is it a substitute to the Namenode?	<b>07</b>
<b>OR</b>		
	(c) Explain following commands of HDFS with syntax and at least one example of each. (i) get (ii) cp (iii) chown	<b>07</b>
<b>Q.3</b>	(a) Explain Metastore in Hive.	<b>03</b>
	(b) Explain Map-reduce framework in brief.	<b>04</b>
	(c) Explain the concept of Blocks and Heartbeat Message in HDFS Architecture. What are the benefits of block transfer?	<b>07</b>
<b>OR</b>		
<b>Q.3</b>	(a) What is Zookeeper? List the benefits of it.	<b>03</b>
	(b) Explain the 5 P's of Data science in brief.	<b>04</b>
	(c) Explain Spark components in detail. Also list the features of spark.	<b>07</b>
<b>Q.4</b>	(a) Differentiate NoSQL and relational databases?	<b>03</b>
	(b) What do you mean by HiveQL Data Definition Language?	<b>04</b>
	(c) Explain following for MongoDB. (i) Indexing (ii) Aggregation	<b>07</b>
<b>OR</b>		
<b>Q.4</b>	(a) Explain any three HiveQL DDL command with its syntax and example.	<b>03</b>
	(b) Explain scaling feature of MongoDB.	<b>04</b>
	(c) What are the problems related to Map Reduce data storage? How Apache Spark solves it using Resilient Distributed Dataset? Explain RDDs in detail.	<b>07</b>
<b>Q.5</b>	(a) Compare Row oriented and Column Oriented database structures.	<b>03</b>
	(b) Discuss how Pig data model will help in effective data flow.	<b>04</b>
	(c) Explain how HBase uses Zookeeper to Build Applications with Zookeeper.	<b>07</b>

**OR**

- Q.5** (a) What is Spark? **03**  
(b) Differentiate: Apache pig Vs Map Reduce. **04**  
(c) Explain CRUD operations in MongoDB. **07**

\*\*\*\*\*