GUJARAT TECHNOLOGICAL UNIVERSITY

BE - SEMESTER–VIII (NEW) - EXAMINATION – SUMMER 2017

Subject Code: 2180710   **Date: 02/05/2017**

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

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.**

Q.1	**(a)**	What is big data analytics? Explain four 'V's of Big data. Briefly discuss applications of big data.	**07**
	(b)	What is Map Reduce? Explain working of various phases of Map Reduce with appropriate example and diagram.	**07**
Q.2	**(a)**	Explain core architecture of Hadoop with suitable block diagram. Discuss role of each component in detail.	**07**
	(b)	What is Hadoop Ecosystem? Discuss various components of Hadoop Ecosystem.	**07**

OR

	(b)	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)**	With suitable block diagram explain architecture of HDFS. Discuss role of Data node and Name node in HDFS. Give commands with appropriate arguments to perform data transfer between local file system and HDFS.	**07**
	(b)	Explain working of Hive with proper steps and diagram.	**07**

OR

Q.3	**(a)**	(i) What is Zookeeper? What are the benefits of Zookeeper? (ii) Write a short note on Apache Pig. Enlist applications of Apache Pig.	**07**
	(b)	What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example.	**07**
Q.4	**(a)**	What is Apache Spark? What are the advantages of using Apache Spark over Hadoop? Explain in brief four major libraries of Apache Spark.	**07**
	(b)	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.4	**(a)**	What is transformation and actions in Apache Spark? Discuss various commands available for this activities in Apache Spark?	**07**
	(b)	What is Resilient Distributed Dataset in Apache Spark? Explain in detail. Make a note on why RDD is better than Map Reduce data storage?	**07**
Q.5	**(a)**	Explain following in brief with respect to Mongo DB : • Collections and documents • Indexing and retrieval • Data aggregation	**07**
	(b)	Requirement specification of blog application in social networking is as follows. • Every post has a unique title, description and url. • Every post can have one or more tags.	**07**

1

- Every post has the name of its publisher and total number of likes.
- Every post has comments given by users along with their name, message, data-time and likes.
- On each post, there can be zero or more comments.

For this set of requirements design a Mongo DB schema.

OR

Q.5 **(a)** What is Mongo DB? Explain in brief key features of Mongo DB. Show basic CRUD operations in Mongo DB with proper example. **07**

(b) Requirement specification for a meeting dashboard application in an organization is as follows: **07**

- Any member in an organization can host a meeting and send an invitations to other members within an organization.
- Invitees can accept or reject the meeting with proper reason.
- Every meeting has the title, timestamp and place/location associated.
- Every meeting has predefined agendas and documents associated.
- Meeting discussion concludes with identifying tasks to accomplish.
- Every task has title, priority, deadline and note associated with it. Task can be assigned to any attendee of meeting.

For this set of requirements design a Mongo DB schema.

****