C.U.SHAH UNIVERSITY
Winter Examination-2018

Subject Name : Business Intelligence and Big Data Analytics

Subject Code : 4TE07BIB1   **Branch: B.Tech (CE)**

Semester : 7   **Date : 01/12/2018**   **Time : 10:30 To 01:30**   **Marks : 70**

Instructions:
(1) Use of Programmable calculator & any other electronic instrument is prohibited.
(2) Instructions written on main answer book are strictly to be obeyed.
(3) Draw neat diagrams and figures (if necessary) at right places.
(4) Assume suitable data if needed.

Q-1		**Attempt the following questions:**	
	a)	What are the most common Input Formats in Hadoop?	**01**
	b)	Compare HDFS & HBase.	**01**
	c)	What is the use of Big Data Analysis for an enterprise?	**01**
	d)	How is Hadoop different from traditional RDBMS?	**01**
	e)	List any two statistical methods that are useful for data-analyst?	**01**
	f)	What is meant by Data node?	**01**
	g)	What are the main components of Hadoop?	**01**
	h)	What are the main features of HDFS?	**01**
	i)	What is the function of job tracker?	**01**
	j)	Define clustering.	**01**
	k)	List some tools used for Big Data.	**01**
	l)	What is default HDFS Block size?	**01**
	m)	What is the name of the framework that Apache developed for processing massive dataset for an application in a computing environment that is distributed?	**01**
	n)	What does Visual Analytics mean?	**01**

Attempt any four questions from Q-2 to Q-8.

Q-2		**Attempt all questions**	
	a)	Define NoSQL and where is it used?	**03**
	b)	Explain: i)Document Oriented Database ii) Graph based Database	**04**
	c)	Write & Explain Apriori algorithm for discovering frequent item sets for mining Boolean Association Rules.	**07**

Q-3 **Attempt all questions**

   a) Explain structure and semi structure.   **03**

   b) How Map Reduce work and explain terminology used in Map reduces.   **04**

   c) Write the steps of the K-means Clustering Algorithm. Also state its limitation.   **07**

Q-4 **Attempt all questions**

   a) Difference between Hive and RDBMS.   **03**

   b) Explain Hive Data types.   **04**

   c) 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-5 **Attempt all questions**

   a) What is Big Data?   **03**

   b) Explain characteristics of Big Data.   **04**

   c) 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.   **07**

Q-6 **Attempt all questions**

   a) What are the advantages of Hadoop?   **03**

   b) Explain Hadoop Architecture.   **04**

   c) Draw HDFS Architecture. Explain any two commands of HDFS from following commands with syntax and at least one example of each. (i) copyFromLocal (ii) setrep (iii) checksum   **07**

Q-7 **Attempt all questions**

   a) Explain Storage mechanism in HBase.   **03**

   b) Compare Raw oriented and Column Oriented database structures.   **04**

   c) What do you mean by HiveQL Data Definition Language? Explain any three HiveQL DDL command with its syntax and example.   **07**

Q-8 **Attempt all questions**

   a) What is Zookeeper? List the benefits of it   **03**

   b) Explain HBase Architecture.   **04**

   c) Explain working of Hive with proper steps and diagram.   **07**