SRN | | | | | | | | | |

REVA UNIVERSITY
Bengaluru, India

Seventh Semester B.Tech CSE Semester End Examination December 2018
Course Title (Subject): Big Data and Hadoop
Course Code:BTCS15F7310 (Qp code – 32722)

Time: 3 Hours **Max. Marks: 100**

Note: *Answer ONE FULL question from each section.* **Marks**

UNIT – I

1. a) Digital data can be classified in to structured, unstructured and semi structured data. Point out the differences between structured, unstructured and semi-structured data. 10
   b) Big data (Hadoop) will replace the traditional RDBMS and datawarehouse.With the help of a suitable diagram, Outline the key considerations of Hadoop for being it as the most wanted technologies to solve the problems of traditional RDBMS. 10
   c) Data continues to be a precious and irreplaceable asset. Data is present in homogeneous sources as well as heterogeneous sources. List and explain the key characteristics of data. 5

OR

2. a) Define big data analytics and its importance. If we go by the industry buzz, every place there seems to be talk about big data and big data analytics. Outline the key aspects of Big data analytics. 10
   b) Differentiate between Analytics 1.0, Analytics 2.0 and Analytics 3.0. 10
   c) Unstructured data is growing by the day and growing by leaps and bounds. List the sources of un-structured data. 5

UNIT – II

3. a) HDFS is a file system designed for storing very large files with streaming data access patterns, running on clusters of commodity hardware. Outline the functionalities of the five important Daemons in Hadoop. 10
   b) Point out the differences between SQL and NOSQL. 10
   c) NOSQL Databases are non-relational. Explain the types of NOSQL Databases. 5

OR

4. a) Hadoop is a distributed Master-Slave Archiecture.With the help of a neat diagram, enumerate the functionalities of Master HDFS and Master MapReduce components in a high-level architecture of Hadoop. 10
   b) Hadoop is best known for Map Reduce and its distributed file system (HDFS, renamed from NDFS), the term is also used for a family of related projects that fall under the umbrella of infrastructure for distributed computing and large-scale data processing. With the help of a neat diagram, outline the key aspects of Hadoop Ecosystem. 10
   c) Enumerate the advantages of NOSQL Databases. 5

UNIT – III

5. a) When a MapReduce job is fed to Hadoop for execution, the Hadoop system performs the job by splitting it into a number of map and reduce tasks. With an aid of a neat data flow diagram, Illustrate the data flow when a single reduce task and multiple reduce tasks are present. 10

   b) Shuffle and sort sits between a mapper and a reducer. With the help of a neat diagram, illustrate the process of Shuffle and sort in MapReduce. 10

   c) Compare and contrast Hadoop with RDBMS to show how Hadoop is similar/dissimilar to RDBMS. 5

OR

6. a) Enumerate the steps involved by Hadoop to run a MapReduce job using a Classic Framework with suitable diagram. 10

   b) MapReduce Paradigm was introduced based on the concept of Functional Programming. With an aid of a neat block diagram, illustrate the steps involved in the working model of MapReduce Programming. 10

   c) Reducer is to reduce a set of intermediate values to a smaller set of values. Reducer has three primary phases. List and explain. 5

UNIT – IV

7. a) Outline the different types of metastores and their relevant functionalities that are used by Hive. 10

   b) Define Pig. Outline the components (anatomy) of Pig that makes it suitable to work with Hadoop for large scale data analysis. 10

   c) Apache pig is a platform for data analysis. Outline the key features of Pig. 5

OR

8 a) Compare and contrast Pig with Hive. 10

   b) Enumerate the DDL and DML statements that are used in Hive. 10

   c) Identify the situations that you would recommend and discourage the use of Pig during data analysis. 5

***