39

SRN | | | | | | | | | |

REVA UNIVERSITY
Bengaluru, India

Seventh Semester B.Tech-CSE Semester End Examination December 2017
Course (Subject): Big Data and Hadoop
Course Code: BTCS15F7310

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

Note: Answer **ONE FULL** question from each unit.

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) Define Big Data .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) Outline the major challenges posed by Big Data in general. 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) Point out the differences between SQL and NoSQL. 10
   b) 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) Outline the important features of NoSQL that make it to be suitable for any big data application. Also, enumerate the major big data applications that use NoSQL. 5

OR

4. a) 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. 15
   b) 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

UNIT-III

5. a) With the help of a neat diagram, illustrate the steps involved by Hadoop to run a MapReduce job using a Classic Framework. 10
   b) In the real world, User code is buggy, Processes crash and machine fails.one of the major benefits of using Hadoop is its ability to handle failures. Outline the different types of failures. 10
   c) Define "Data Locality Optimization" in Hadoop. 5

OR

6.	a)	With the aid of a neat data flow diagram, Illustrate the data flow of general case of single and multiple reduce tasks.		10
	b)	With the help of a neat diagram, illustrate the process of Shuffle and sort in MapReduce.		10
	c)	Compare and contrast RDBMS with MapReduce.		5

UNIT-IV

7.	a)	With the help of a neat diagram, illustrate the architecture of 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

***