647510 **0304212C2020** **Seat No :_____**

M.Sc.(IT & CA) Semester - 2 (*CBCS*) Examination
August/September -2020 [OLD COURSE]
INTRODUCTION TO BIG DATA AND HADOOP(CORE)

Time: 2:00 Hours **Marks: 56**

Instructions:

1. Figure to the right indicate marks.
2. There are five questions in the question paper.
3. Answer any four of the following questions.

Q.1(A) Answer following questions. (04)

1. Default block size in HDFS is (Hadoop 2.0) _____
2. HDFS stands for _____
3. What is data node?
4. What is secondary name node?

Q.1(B) Answer any one. (02)

1. Explain CRC checksum policy.
2. Explain function of data node.

Q.1(C) Answer any one. (03)

1. Explain function of name node.
2. Explain rack awareness in HDFS.

Q.1(D) Answer any one. (05)

1. Explain advanced HDFS features.
2. Explain 3V characteristics of Big Data.

Q.2(A) Answer following questions. (04)

1. YARN stands for _____
2. In pig _____ is used to display the output to the screen.
3. Pig was developed by _____
4. What is Tuple in pig?

Q.2(B) Answer any one. (02)

1. Explain Hadoop data Types.
2. Explain input formats in MapReduce.

Q.2(C) Answer any one. (03)

1. Explain schedulers in MR.
2. Explain different mode of executing pig constructs.

Q.2(D) Answer any one. (05)

1. Explain MapReduce architecture.
2. Explain user defined functions in pig using suitable example.

Q.3(A) Answer following questions. (04)

1. Hive is based suited for OLAP. (True/False)
2. HQL stands for _____
3. In Pseudo distributed mode Name node and Data Node are not available. (True/False)
4. In Standalone mode of Hadoop, there are no Hadoop daemons running in background. (True/False)

Q.3(B) Answer any one. (02)

1. Explain partitions in Hive.
2. Explain external table in Hive.

Q.3(C) Answer any one. (03)

1. Explain Standalone mode of MR.
2. Explain different mode of executing hive queries.

Q.3(D) Answer any one. (05)

1. Explain Hive Architecture.
2. Explain views in Hive using suitable example.

Q.4(A) Answer following questions. (04)

1. HBAse is NoSQL, distributed database. (True/False)
2. HBase stores data in _____ form.
3. What is HBase coprocessor?
4. DDL stands for _____

Q.4(B) Answer any one. (02)

1. Explain Master and Region Server.
2. Explain bulk loading in HBase.

Q.4(C) Answer any one. (03)

1. Explain HBase Vs Cassandra.
2. List tables in HBase using Java API.

Q.4(D) Answer any one. (05)

1. Explain catalog tables in Hive.
2. Explain HBase architecture.

Q.5(A) Answer following questions. (04)

1. EMR stands for _____
2. RDD stands for _____
3. EC2 stands for _____
4. Write down a name of Graph Database Management system on Hadoop.

Q.5(B) Answer any one. (02)

1. Explain Graph cessing in Neo4j.
2. Explain Oozie work flow.

Q.5(C) Answer any one. (03)

1. Explain RDD.
2. Explain how Sqoop works.

Q.5(D) Answer any one. (05)

1. Explain import export process across RDBMS in HDFS.
2. Explain different mode of Hadoop cluster.

****