



PAR-003-1082002

Seat No. _____

M. Sc. (IT & CA) (Sem. II) Examination

August / September - 2020

CS - 08 : Introduction To Big Data And HADOOP

Faculty Code : 003

Subject Code : 1082002

Time : $2\frac{1}{2}$ Hours]

[Total Marks : 70

1 (A) Answer the following : 4

- (1) _____ Command list the files and folders for the given path.
- (2) _____ task takes the output from a map task as input and produces output.
- (3) HDFS stands for _____.
- (4) HDFS can contain only one data node.
True or False?

(B) Answer any **one** in brief : 2

- (1) What is Virtual Box ?
- (2) Explain 3V characteristics of Big Data.

(C) Answer any **one** in detail : 3

- (1) Explain Brewers CAP Theorem.
- (2) Explain Rack awareness and Block placement policy.

(D) Answer any **one** : 5

- (1) List and Explain HDFS features.
- (2) Explain HDFS architecture.

- 2 (A) Answer the following : 4
- (1) Give full form of BSP.
 - (2) _____ node is responsible for storing the individual file blocks on the slave nodes in Hadoop cluster.
 - (3) _____ receives the job information from Job Tracker daemon and initiates a task on slave node.
 - (4) Filter is used to remove unwanted rows from a relation in pig. True or False?
- (B) Answer any **one** in brief : 2
- (1) What is slot and container?
 - (2) What is Name Node?
- (C) Answer any **one** in detail : 3
- (1) Explain load, dump and store pig construct.
 - (2) How mapreduce works?
- (D) Answer any **one** : 5
- (1) Describe Pig Architecture.
 - (2) Explain MapReduce Architecture.
- 3 (A) Answer the following : 4
- (1) Sequence files stores data in serialized _____ pair.
 - (2) _____ is a data serialization tool which serializes data in Hadoop.
 - (3) _____ is a data warehouse system for querying and analysis of data that is stored in HDFS.
 - (4) Hive runs SQL like queries called HQL. True or False.
- (B) Answer any **one** in brief : 2
- (1) What is view?
 - (2) Explain select statement with suitable example.

- (C) Answer any **one** in detail : 3
- (1) List and Explain any one Optimization Technique.
 - (2) Differentiate External vs. Managed Tables.
- (D) Answer any **one** : 5
- (1) Explain Term Frequency and Inverse Document Frequency.
 - (2) Explain create table and insert data in table with suitable example.
- 4 (A) Answer the following : 4
- (1) _____ is a distributed column-oriented database built on top of HDFS.
 - (2) The _____ table holds the list of META table regions.
 - (3) _____ command is used to list all the tables in HBase.
 - (4) HBase is based on Google BigTable. True or False ?
- (B) Answer any **one** in brief : 2
- (1) What is Meta table?
 - (2) What is Minor Compaction?
- (C) Answer any **one** in detail : 3
- (1) Differentiate HBase vs. Cassandra.
 - (2) Explain EndPoints and Observers in HBase.
- (D) Answer any **one** : 5
- (1) Explain HBase Architecture.
 - (2) What is Filter? List and Explain any two filters.

- 5 (A) Answer the following : 4
- (1) _____ is a module for structured data processing in spark.
 - (2) _____ is an open source engine which can process data in real-time using its distributed architecture.
 - (3) _____ is a modern, open source, distributed SQL query engine for Apache Hadoop.
 - (4) Neo4j is open source graph database.
True or False?
- (B) Answer any **one** in brief : 2
- (1) What is RDD?
 - (2) What is Amazon EC2?
- (C) Answer any **one** in detail : 3
- (1) What is Sqoop? How Sqoop works?
 - (2) Explain workflow in Oozie.
- (D) Answer any **one** : 5
- (1) What is Flume? Explain Flume Architecture.
 - (2) Write a note on Spark Architecture.
-