Q.28 Discuss the strengths and weaknesses of map reduce.

Q.29 What is Pig and when Pig should be used?

Q.30 How Hive differs from traditional RDBMS?

Q.31 How to read and get data into R?

Q.32 What are the statistical models in R?

Q.33 What could be the various sources of big data? Explain

Q.34 What are the ecosystem projects used for processing Big data?

Q.35 Explain any two popular tools of big data.

SECTION-D

Note: Long answer type questions. Attempt any two questions out of three questions. (2x10=20)

Q.36 Explain different Big data applications.

Q.37 What is map reduce? Explain the map reduce architecture with neat and clean diagram.

Q.38 Illustrate main features and Architecture of Hive with neat diagram.

No. of Printed Pages : 4 180854B/170854B
Roll No. .....

5th Sem / Comp
Subject:- Big Data

Time : 3Hrs. M.M. : 100

SECTION-A

Note: Multiple choice questions. All questions are compulsory (10x1=10)

Q.1 Which of the following is NOT a characteristics of big data?

a) Volume b) Velocity
c) Variety d) Simplicity

Q.2 Which of the following tools is commonly used for big data processing?

a) Microsoft Excel b) Hadoop
c) Notepad d) Adobe Photoshop

Q.3 What does HDFS stand for?

a) High Definition File System
b) Hadoop Distributed File System
c) Hyper Data Framework System
d) High Density File storage

Q.4 What is the role of the NameNode in Hadoop?

a) Store data blocks
b) Handle user authentication
c) Manage metadata and file system namespace
d) Process data

Q.5 Which of the following is NOT a phase in Mapreduce?

a) Map phase b) Shuffle phase
c) Load phase d) Reduce phase

Q.6 What data structure is primarily used to represent data in MapReduce?

a) DataFrame b) Key-value Pair
c) Table d) JSON

Q.7 What is the default file format for Hive tables?

a) CSV b) JSON
c) Text d) Parquet

Q.8 How do you create a table in Hive?

a) CREATE DATABASE
b) CREATE TABLE
c) MAKE TABLE
d) NEW TABLE

Q.9 How can R connect to a Hadoop cluster?

a) Using the RODBC package
b) Using the RJDBC package
c) Using the rjava package
d) All of the above

Q.10 What does the term "data lake" refer to?

a) A structured database
b) A storage repository for new data
c) A type of relational database
d) A programming language

SECTION-B

Note: Objective type questions. All questions are compulsory. (10x1=10)

Q.11 What is data frame in R.

Q.12 Define big Data.

Q.13 Name one common use case for big data analytics.

Q.14 What does ETL stand for?

Q.15 Name one component of the Hadoop ecosystem.

Q.16 What is the purpose of the HDFS?

Q.17 What does YARN stand for in the context of Hadoop?

Q.18 What type of database does Hive use?

Q.19 What are the two main components of Pig?

Q.20 Who created the R programming language?

SECTION-C

Note: Short answer type questions. Attempt any twelve questions out of fifteen questions. (12x5=60)

Q.21 What is Hadoop, List various Hadoop vendors.

Q.22 What are the challenges for processing Big data?

Q.23 What is Distributed File System? Explain in brief.

Q.24 Explain how Hadoop cluster works?

Q.25 Define any 3 daemons of Hadoop and their functionalities .

Q.26 Explain how map reduce works.

Q.27 What is word count problem? Epxlain in brief.