

SECTION-C

Note: Short answer type questions. Attempt any twelve questions out of fifteen questions. $(12 \times 5 = 60)$

- Q.21 Write challenge for processing big data? (CO1)
Q.22 Write characteristics of big data. (CO1)
Q.23 Explain map reduce and R programming tool? (CO5)
Q.24 Explain pig data flow engine? (CO4)
Q.25 Explain architecture of hive? (CO4)
Q.26 Explain apache hadoop & hadoop eco system? (CO2)
Q.27 Explain map reduce architecture? (CO3)
Q.28 Explain how to get data in "R"? (CO5)
Q.29 Explain any two big data tool? (CO2)
Q.30 Write various benefits of big data? (CO1)
Q.31 Explain ordered and unordered factors in "R"? (CO5)
Q.32 Write statistical model in "R"? (CO5)
Q.33 Write working of hadoop cluster? (CO2)
Q.34 Explain loading data in to HDFS? (CO3)
Q.35 Explain big data with example? (CO1)

SECTION-D

Note: Long answer type questions. Attempt any two questions out of three questions. $(2 \times 10 = 20)$

- Q.36 Write the utility of big data tools in detail? (CO2)
Q.37 Explain big data in detail? Write various application of big data? (CO1)
Q.38 Write.
a) Array and matrices in "R"? (CO5)
b) List and data frames in "R"? (CO5)

No. of Printed Pages : 4

Roll No.

180854B/170854B

5th Sem./Computer

Subject:- Big Data

Time : 3Hrs.

M.M. : 100

SECTION-A

Note: Multiple choice questions. All questions are compulsory $(10 \times 1 = 10)$

- Q.1 What is hadoop ? (CO1)
a) A programming language
b) A distributed storage and processing framework
c) A relational database management system
d) An operating system
- Q.2 Which programming language is commonly used to write Map Reduce programs in Hadoop? (CO3)
a) Java b) Python
c) C++ d) Ruby
- Q.3 What is the primary purpose of Hadoop's HDFS (Hadoop Distributed File System)? (CO2)
a) To store and manage large volumes of structured data
b) To provide real-time processing capabilities
c) To partition data for parallel processing across a distributed cluster
d) To optimize SQL queries for faster performance

- Q.4 Which component of Hadoop is responsible for managing resources and scheduling tasks on the cluster? (CO3)
- a) HDFS
 - b) Map Reduce
 - c) YARN (Yet Another Resource Negotiator)
 - d) H Base
- Q.5 Which of the following is NOT a common use case for Hadoop? (CO2)
- a) Analyzing social media data
 - b) Real-time stream processing
 - c) Running transactional database
 - d) Log processing and analysis
- Q.6 Which programming language is commonly used for processing Big Data in Hadoop? (CO3)
- a) Java
 - b) Python
 - c) C++
 - d) Ruby
- Q.7 What is the purpose of Hadoop Map Reduce? (CO2)
- a) To store and manage data in Hadoop
 - b) To process and analyze data in parallel across a distributed cluster
 - c) To provide a query and analysis interface for Hadoop
 - d) To schedule and manage resource allocation in Hadoop

(2)

180854B/170854B

- Q.8 Hadoop MapReduce divides the data processing tasks into _____ and _____ phases. (CO4)
- a) Read, write
 - b) Map, reduce
 - c) Input, output
 - d) Start, End
- Q.9 _____ is a platform for constructing data flows for extract, transform, and load (ETL) processing and analysis of large data sets. (CO4)
- a) Pig Latin
 - b) Oozie
 - c) Pig
 - d) Hive
- Q.10 The default input format is _____ which treats each value of input a new value and the associated key is byte off set. (CO5)
- a) Text format
 - b) Text input Format
 - c) input format
 - d) All of the above

SECTION-B

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

- Q.11 HDFC stands for _____ (CO2)
- Q.12 Pig operates in which made? (CO3)
- Q.13 What is data node? (CO2)
- Q.14 What is array and matrices in "R" (CO5)
- Q.15 What is hive? (CO3)
- Q.16 Write any challenge for processing big data. (CO1)
- Q.17 Write benefits of big data (CO1)
- Q.18 Write any two tools of big data (CO2)
- Q.19 Write when pig should be used? (CO4)
- Q.20 Hadoop is an virtual framework (T/F)? (CO1)

(3)

180854B/170854B