

Modified

# CBCS SCHEME

USN 

--	--	--	--	--	--	--	--	--	--

20MCA352

## Third Semester MCA Degree Examination, Feb./Mar. 2022 Big Data Analytics

Time: 3 hrs.

Max. Marks: 100

Note: Answer any FIVE full questions, choosing ONE full question from each module.

### Module-1

- 1 a. Define Big Data. (02 Marks)
- b. Explain different characteristics of Big data Analytics. (10 Marks)
- c. Write a note on missing values. (08 Marks)

OR

- 2 a. Discuss different types of data sources of big data analytics. (10 Marks)
- b. Discuss the applications of big data analytics. (10 Marks)

### Module-2

- 3 a. Briefly explain mobile intelligence and big data. (10 Marks)
- b. Write a note on cloud and big data. (10 Marks)

OR

- 4 a. Explain working together of HDFS and map reduce. (10 Marks)
- b. Define crowd sourcing. Explain different types of crowd sourcing. (10 Marks)

### Module-3

- 5 a. Compare between Big data, RDBMS and grid computing. (10 Marks)
- b. Write a note on: i) History of Hadoop ii) Volunteer computing. (10 Marks)

OR

- 6 a. Illustrate the disabilities of implementing storage and analysis support for big data. (10 Marks)
- b. Explain different Hadoop versions and releases. (10 Marks)

### Module-4

- 7 a. Briefly explain the anatomy of file write. (10 Marks)
- b. Explain different HDFS concepts in detail. (10 Marks)

OR

- 8 a. Discuss different file system operations. (10 Marks)
- b. Define file system interface. Explain different file system interfaces. (10 Marks)

### Module-5

- 9 a. What is MapReduce? Sketch the neat diagram and explain the logical data flow in MapReduce. (10 Marks)
- b. Write a short note on: i) MapReduce UI ii) Hadoop streaming. (10 Marks)

OR

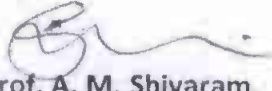
- 10 a. Write a java MapReduce code to find maximum temperature from the weather data set. (10 Marks)
- b. Write a note on: i) Hadoop logs ii) Remote Debugging. (10 Marks)

\*\*\*\*\*

Important Note : 1. On completing your answers, compulsorily draw diagonal cross lines on the remaining blank pages.  
2. Any revealing of identification, appeal to evaluator and /or equations written eg, 42+8= 50, will be treated as malpractice.

Dear Sir,

Sub Code	Subject Name	Comments/Corrections/Solutions
20MCA31	Data Analytics using Python	No Modification. It is Approved
20MCA32	Internet of Things	It is Approved. But Marks distribution may be as shown below 1b) 5 points: 5m +5 m explanation 2a) Diagram 5marks+ Explanation 5 marks 2b ) Diagram 5marks +Description 5marks, 3a) Six points : 6m +Explanation 4m 3b)diagram 7m + Explanation 3m 4a) Diagr 6m + expla 4m 4b)Smart Objs 5m + Wireless Sensor N/Ws 5m 5a) 5 Key advantg of IPs 5marks +Expla 5m 5b) Diagr 6m +Expln 4m 6a)COSE Definition 3m +Diagr 5m +Explan 2m 6b)Diagr 6m +Explan 4m 7a)Diagram 6marks +Explanation 4marks 7b)Definition with exampl for Supervised & Unsupervised Learning 5m each 8a)Diagram & Explanation 5 each 8b)OCTAVE 5 marks + FAIR 5marks 9a) 5x2=10 marks 9b)diagram 6 + explanation 4marks 10a)Diagram 6 + Explan 4marks 10b) Python code 8marks + Text File 2marks
20MCA33	Advnces in JAVA	Q.No 1c in Q. Paper may be read as POST instead of PSOT. It is Approved
20MCA342	Cloud Computing	3a. Question could have been 5th or 6th question (Module 3). The concepts of virtual machine exists in "3rd module" doesn't belong to this 2nd module. Students lost a choice by this mistake. So 10 marks may be liberally given. 6.b Theorem stated may be considered for Full 10 marks. 8.b. Types of clouds is huge topic, but asked for 10 marks. Any relevant points written under each cloud(Pvt, Public, Community, hybrid) can be considered and award full marks.
20MCA352	Big Data Analytics	10 a) : Write a map reduce program to find maximum temperature from Weather dataset. 10Marks MapReduce is based on set of key value pairs. So first we have to decide on the types for the key/value pairs for the input. Map Phase: The input for Map phase is set of weather data files as shown in snap shot. The types of input key value pairs are LongWritable and Text and the types of output key value pairs are Text and IntWritable. Each Map task extracts the temperature data from the given year file. The output of the map phase is set of key value pairs. Set of keys are the years. Values are the temperature of each year. Reduce Phase: Reduce phase takes all the values associated with a particular key. That is all the temperature values belong to a particular year is fed to a same reducer. Then each reducer finds the highest recorded temperature for each year. The types of output key value pairs in Map phase is same for the types of input key value pairs in reduce phase (Text and IntWritable). The types of output key value pairs in reduce phase is too Text and IntWritable. So, in this example we write three java classes: <input checked="" type="checkbox"/> HighestMapper.java – 3 Marks <input checked="" type="checkbox"/> HighestReducer.java -- 3 Marks <input checked="" type="checkbox"/> HighestDriver.java – 4 Marks

  
 Prof. A. M. Shivaram,  
 Chairman BOE, MCA, VTU,  
 Associate Professor, Dept. of MCA,  
 BIT, Bengaluru-560004.

**"APPROVED"**  
  
 Registrar (Evaluation)  
 Visvesvaraya Technological University  
 BELAGAVI - 590018



# Visvesvaraya Technological University

Belagavi, Karnataka - 590 018



211120MCA35241838

## Scheme & Solutions

  
Signature of Scrutinizer

Subject Title : Big Data Analytics

Subject Code : 20MCA352

Question Number	Solution	Marks Allocated
1.a)	Definition - 2m	02
b)	Any 5 V's - Volume, velocity, variety, veracity, validity	10 5x2 =
c)	missing values Any 4 points. OR	08
2.a)	Type of data sources - Transactional data - Unstructured data - Qualitative data - publicly available data - Data pools	10
2.b)	Application : i) Big data in Banking      ii) Retail ii) — in Government      iii) Finance sector iii) — in Healthcare      iv) Telecom	10

Question Number	Solution	Marks Allocated
3.a)	Applications and usage of mobile device in the field of big data	10
b)	How cloud is related with the Big data OR	10
4.a)	Diagram — Explanation — OR	5 05
b)	Definition Types - crowd source Design - crowd funding - micro tasks - open Innovation	02 08
5.a)	comparison between Big data, RDBMS & Grid computing - Any 4 points	10
b)	(i) History of Hadoop should discuss initial development of Hadoop (ii) Definition & Example	05 05

Question Number	Solution	Marks Allocated
6-a)	Diagram	4
	Explanation	6
b)	should discuss any 3 version, releases and comparisons. OR	10
7-a)	Explanation	05
	Diagram	05
b)	HDFS concepts - Blocks - Namenode and Datanode - HDFS federation - HDFS High Availability OR	10
8.a)	Definition Discuss any 4 file system operation command	2 8
b)	Definition Types - HTTP - C programming - FUSE	2 10

Question Number	Solution	Marks Allocated
9.a)	Definition - Diagram - Explanation	02 04 04 <hr/>
b)	(i) map Reduce UI short Notes (ii) Hadoop Streaming short Notes	05 05 <hr/>
10.a)	write down the program	10
b)	(i) Hadoop logs - short paragraph	05
	(ii) Remote Debugging - Paragraph writing	05 <hr/>

"APPROVED"



Registrar (Evaluation)

 Visvesvaraya Technological University  
 BELAGAVI - 590018

