PVP 19

Code: 19CS4702A

IV B.Tech - I Semester – Regular Examinations - DECEMBER 2022

BIG DATA

(COMPUTER SCIENCE & ENGINEERING)

Duration: 3 hours Max. Marks: 70

Note: 1. This question paper contains two Parts A and B.
2. Part-A contains 5 short answer questions. Each Question carries 2 Marks.
3. Part-B contains 5 essay questions with an internal choice from each unit. Each question carries 12 marks.
4. All parts of Question paper must be answered in one place.

BL – Blooms Level CO – Course Outcome

PART – A

		BL	CO
1. a)	Define Grid Computing.	L1	CO1
1. b)	Describe why block in HDFS is so large.	L1	CO2
1. c)	Define Hive Shell.	L1	CO3
1. d)	Describe how to create a Spark application in Scala.	L1	CO2
1. e)	List which model is used for recommendation system.	L1	CO4

PART – B

			BL	CO	Max. Marks
UNIT-I					
2	a)	Explain how Hadoop is used in data storage and analysis.	L2	CO1	6 M
	b)	Define Volunteer Computing. Differentiate between traditional RDBMS and MapReduce.	L2	CO1	6 M

OR					
3	a)	Explain how to write a MapReduce program in Java.	L2	CO1	6 M
	b)	Discuss how job runs on distributed MapReduce.	L2	CO1	6 M
UNIT-II					
4	a)	Interpret how does Namenode High availability work.	L3	CO2	6 M
	b)	Describe how does HDFS provide data transfer from hdfs to local file system and vice versa.	L2	CO2	6 M
OR					
5	a)	Explain how do you read a file from HDFS URL and explain with an example.	L2	CO2	6 M
	b)	Discuss why cluster balance is important in HDFS. Explain with an example.	L2	CO2	6 M
UNIT-III					
6	a)	Explain how can you create and manage the databases in Hive.	L2	CO3	6 M
	b)	Explain main features and Architecture of Hive with neat diagram.	L2	CO3	6 M
OR					
7	a)	Describe storage formats in HiveQL.	L2	CO3	6 M
	b)	Illustrate what happens when a table is dropped in HiveQL. Explain importing data in HiveQL.	L3	CO3	6 M

UNIT-IV					
8	a)	Define RDD in Spark. How are they computed in Spark? Explain various ways in which it can create.	L2	CO2	6 M
	b)	Explain how to create a Spark application in Scala. Write a Spark code to find the maximum temperature.	L2	CO2	6 M
OR					
9	a)	Explain the use of accumulator and broadcast variable in Spark discuss them with an example.	L2	CO2	6 M
	b)	Explain different types of tasks in DAG construction and explain with an example.	L2	CO2	6 M
UNIT-V					
10	a)	Illustrate how to calculate term frequency and inverse document frequency and explain with an example.	L3	CO4	6 M
	b)	Illustrate how would you get the features of document in content based systems.	L3	CO4	6 M
OR					
11	Explain implementation strategies of collaborative filtering model.		L2	CO4	12 M