21AI62

USN | | | | | | | | | | |

RV COLLEGE OF ENGINEERING®

(An Autonomous Institution Affiliated to VTU)

VI Semester B. E. Examinations Sept/Oct– 2024

Artificial Intelligence & Machine Learning

BIG DATA TECHNOLOGIES

Time: 03 Hours ***Maximum Marks: 100***

Instructions to candidates:

1. Answer all questions from Part A. Part A questions should be answered in first three pages of the answer book only.
2. Answer FIVE full questions from Part B. In Part B question number 2 is compulsory. Answer any one full question from 3 and 4, 5 and 6, 7 and 8, 9 and 10.

PART-A

			M	BT	CO
1	1.1	What is volunteer computing?	02	2	1
	1.2	Give any 4 differences between traditional RDBMS and map reduce.	02	1	1
	1.3	How does JVM handle task failure?	02	2	1
	1.4	Give reason to choose an appropriate block size for a Hadoop development?	02	1	1
	1.5	Give two reasons, why buckets are needed in hive?	02	2	1
	1.6	Give reason: what is need of external table.	02	1	1
	1.7	Define an accumulator	02	1	1
	1.8	What is a view in HiveQL	02	4	1
	1.9	The two styles of application that benefit greatly from Spark's processing model are _____ and _____	02	2	2
	1.10	How do we scale a set of Flume agents?	02	1	1

PART-B

2	a	Give the structure of HDFS. Brief the conditions when it is suitable for loading the unstructured data.	06	2	2
	b	Elaborate the anatomy of file read and file write with neat diagram.	10	2	1
3	a	Discuss the job submission and job initialization steps in anatomy of map reduce job run with a neat diagram.	06	3	3
	b	Write a map reduce program using java to find minimum integer value from a large file of integers	10	1	1
		OR			
4	a	Elaborate the shuffle and sort process in map phase?	06	3	3
	b	Write a map reduce program using java to find the maximum temperature from the NCDC data set.	10	1	1

<table>
<tr><td>5</td><td>a</td><td>Discuss the role of partition in a table? Discuss the reason to organize the tables (or partitions) into buckets. Give example</td><td>06</td><td>2</td><td>2</td></tr>
<tr><td></td><td>b</td><td>Consider following Tables and demonstrate following queries
<table>
<tr><td>ID</td><td>Name</td><td>Salary</td><td>Designation</td><td>Dept.</td></tr>
<tr><td>1201</td><td>Gopal</td><td>45000</td><td>Technical manager</td><td>TP</td></tr>
<tr><td>1202</td><td>Manisha</td><td>45000</td><td>Proofreader</td><td>PR</td></tr>
<tr><td>1203</td><td>Masthanvali</td><td>40000</td><td>Technical writer</td><td>TP</td></tr>
<tr><td>1204</td><td>Kiran</td><td>40000</td><td>Hr Admin</td><td>HR</td></tr>
<tr><td>1205</td><td>Kranthi</td><td>30000</td><td>Op Admin</td><td>Admin</td></tr>
</table>
a) Creating and loading the data is mandatory
b) Display employee id, name and Department order by Department
c) Display total count of the department using group by clause</td><td>10</td><td>3</td><td>2</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>6</td><td>a</td><td>Compare:
1. Managed Table and External Table
2. Partitions and Buckets</td><td>06</td><td>2</td><td>2</td></tr>
<tr><td></td><td>b</td><td>Justify the purpose VIEW in the Hive with simple example.</td><td>10</td><td>3</td><td>2</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td>7</td><td>a</td><td>With a neat diagram discuss the role of a flume agent? How do you start a flume agent?</td><td>08</td><td>4</td><td>3</td></tr>
<tr><td></td><td>b</td><td>“Aggregating Flume events is achieved by having tiers of Flume agents.” Justify the statement by connecting two flume agents by Avro sink-source pair and elaborate the process in detail</td><td>08</td><td>2</td><td>2</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>8</td><td>a</td><td>With a neat diagram describe fanning out to an HDFS and logger sink</td><td>08</td><td>4</td><td>3</td></tr>
<tr><td></td><td>b</td><td>“Flume event data, is partitioned by time.” Justify the statement by elaborating the working of Partitioning and Interceptors in flume.</td><td>08</td><td>4</td><td>2</td></tr>
<tr><td colspan="6"></td></tr>
<tr><td>9</td><td>a</td><td>Elaborate the working of spark task scheduling and execution.</td><td>08</td><td>2</td><td>3</td></tr>
<tr><td></td><td>b</td><td>Explain the process of RDD creation.</td><td>08</td><td>2</td><td>3</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>10</td><td>a</td><td>How do we implement spark in a stand alone system. Give an example using scala?</td><td>08</td><td>2</td><td>3</td></tr>
<tr><td></td><td>b</td><td>Discuss the two categories of operation on RDDs.</td><td>08</td><td>2</td><td>3</td></tr>
</table>