Department of Artificial Intelligence and Machine Leaning

Date: 06.06.2025	Test - 2	Max. Marks: 50+10
Semester: VI	UG	Duration: $1\frac{1}{2}$ Hrs + ½ Hr
Course Title: Big Data Technologies		Course Code: AI362IA

Sl No	Questions	M	BT	CO
1.	Tasks are considered failed if they exceed a timeout period of _____ minutes and it can set by _____ property	2	2	1
2.	The maximum number of attempts given to run a Map Reduce job on application master is _____ and it can be controlled by _____ property	2	2	1
3.	Blacklisting of a Node Manager is performed by the_____, and tasks are rescheduled on a different Node Manager if more than _____ tasks fail	2	2	1
4.	List two differences between Traditional Databases and Hive?	2	2	1
5.	State the two index types used in HiveQL?	2	2	1
	PART- B			
1	Summarize the Job Submission and Job Initialization stages in the anatomy of a Map Reduce job run with a diagram.	10	2	1
2(a)	Explain how progress and status updates are handled during a MapReduce job run and identify their main components.	3+3	2	1
2(b)	Explain how a Node Manager failure occurs in Hadoop and how it is handled by the system.	4	2	2
3(a)	Outline the Shuffle and Sort phase from the perspective of the Reduce side in a Map Reduce job.	6	2	2
3(b)	State the reasons for disabling speculative execution in a MapReduce job.	4	2	2
4(a)	Discuss how hive organises tables into partitions?	5	2	2
4(b)	Explain the sorting and aggregating concepts in hive with an example?	5	2	2
5	A music shop records information about its customers, including their customer ID, first name, and last name. The shop also keeps track of all the movies available, identified by a movie ID and movie name. Orders placed by the shop are tracked using an order ID, along with the supplier ID, movie ID, and the number of copies ordered. Details of all suppliers who provide the CDs are maintained, including their supplier ID and supplier name. Finally, the renting information is recorded with the customer ID, tape ID, checkout date, and rental duration. For the above scenario, perform the following a. Create relevant tables b. Load the data into the table c. Write following queries using HiveQL i. List the movies that have more than 4 copies been ordered? ii. Which customers rented movies for 5 days or more? iii. Which movies aren't in the inventory? iv. Which movie was rented for the longest duration (by any customer)?	10	3	4