



## Department of Artificial Intelligence and Machine Learning

Course Code:	: 21AI62	Date	: 26 .07.2024
Semester	: VI	Time	: 02:00 – 04:00
Max Marks	: 50 Marks	Duration	: 120 mins

### Big Data Technologies

#### Improvement Test

Note: Answer all the Questions

SL. No	Questions	M	BT
<b>PART - A</b>			
1	Define flume agent?	01	2
2	State the use of transaction in flume agent?	01	2
3	The two different types of channel used in flume are _____ and _____	01	1
4	The file format used by flume for storing the data is _____	01	2
5	State the two fields in Avro Record?	01	1
6	State two types of algorithm processing that best suits for spark processing model?	01	2
7	State two categories of operation on RDD's in Spark	01	1
8	Define Accumulator in Spark?	01	1
9	_____ sends the update message when a task has finished or failed in Spark Job	01	2
10	The number of core allocated for task execution in spark is _____	01	1
<b>PART - B</b>			
1	a Discuss with the diagram and code snippet the working of flume?	06	2
	b "The point of Flume is to deliver large amounts of data into a Hadoop data store". Show how to configure a Flume agent to deliver events to an HDFS sink?	04	2
2	a Discuss scaling of agent in flume?	06	2
	b "Flume uses transactions to ensure that each batch of events is reliably delivered from a source to a channel, and from a channel to a sink". How this is achieved?	04	4
3	a List out the features of apache spark?	04	4
	b Write and explain Scala application to find the maximum temperature, using Spark?	06	2
4	a Compare and Contrast reduceByKey() and foldByKey() transformations for aggregating RDDs in Spark?	06	4
	b How is data serialization handled in Apache Spark?	04	
5	Discuss with a neat diagram Job Submission in Anatomy of Spark Job Run	10	