UPES
End Semester Examination, May 2023

Course: Big Data Analysis
Program: B.Tech CSE (Big Data)
Course Code: CSBD3009

Semester : VI
Time : 03 hrs.
Max. Marks: 100

Instructions: Calculator is not allowed

SECTION A
(5Qx4M=20Marks)

S. No.	Answer all the questions:	Marks	CO
Q 1	State the various aspects of traditional closed-source Data warehouses.	4	CO1
Q 2	Explain two major differences in Joins between Hive and SQL.	4	CO2
Q 3	Illustrate data munging in brief.	4	CO3
Q 4	Explain Multi-Dimensional View of Data Mining in brief.	4	CO4
Q 5	Explain the advantages of Impala over Hive and Pig.	4	CO5

SECTION B
(4Qx10M= 40 Marks)

Q 6	Write the commands/query corresponding to the following statements: a) Copy files from one directory to another directory in HDFS. b) Create tables and insert data into them in Hive. c) Formatting Hadoop Name and Data Nodes. d) Starting and Stopping Impala servers. e) Change the permission (600) and Ownership of Linux files.	10	CO1
Q 7	Describe the internal architecture of the hive and its different components with a suitable diagram.	10	CO2
Q 8	Illustrate the workings of Apache Pig. Demonstrate the differences between Apache Pig and Hadoop Map-Reduce program. OR Illustrate clustering, Data Cube Aggregation, and Principal Component Analysis in brief.	5+5 10	CO3

Q 9	Explain Impala Statestore, Meta-Data, Metastore, and Impala demons precisely.	10	CO4
	SECTION-C **(2Qx20M=40 Marks)**		
Q 10	Explain Exploratory Data Analysis (EDA) in brief. Illustrate the various approaches and goals of EDA? Explain the core architecture of Impala and explain each of its components precisely.	4+6+10	CO4

Q 11	Explain various goals of R in web project development? Interpret the correlation analysis and X2 (chi-square) test elaborately. Explain various data reduction strategies in brief.
	
	OR
	
	Explain the following in brief:
	  a) Hive Metastore,
	  b) Execution engine,
	  c) Hive data model partitions.
	  d) OLTP Vs. OLAP
		6+6+8
		
		20
			CO5