Register No : _____

THE KAVERY ENGINEERING COLLEGE

(An Autonomous Institution)

B.E/ B.TECH DEGREE END SEMESTER THEORY EXAMINATIONS, APRIL-MAY 2025

Fifth Semester

Information Technology

CCS334 - Big Data Analytics
(Common to Computer Science and Engineering)

Regulations - 2021

Duration : 3 Hrs ***Maximum Marks : 100***

PART – A (ANSWER ALL QUESTIONS) (Marks 10*2=20)

Q.No	*Questions*	*BLT*	*CO*	*Marks*
1.	What are the common challenges associated with big data? Why is traditional data processing insufficient for handling big data sets?	L1	1	2
2.	Infer the role of Trans Firewall Analytics in monitoring and securing data transmissions across multiple firewalls within a small organization.	L2	1	2
3.	Define the term “schema-less” in the context of NoSQL databases.	L1	2	2
4.	Interpret the circumstance in which NoSQL database are chosen over a relational database.	L3	2	2
5.	Describe a scenario where HDFS is used in a data processing pipeline.	L1	3	2
6.	Recall the primary purpose of HDFS in the Hadoop ecosystem.	L1	3	2
7.	Give a real-world application where Map Reduce is used for log analysis.	L2	4	2
8.	How does YARN help with resource allocation and job scheduling for map reduce tasks in a multi-tenant Hadoop cluster?	L2	4	2
9.	Specify how time-series data are handled in HBase.	L1	5	2
10.	Enumerate how pig simplify data processing tasks when working with large datasets in Hadoop.	L1	5	2

PART – B (ANSWER ALL QUESTIONS) (Marks 5*16 = 80)

Q.No			*Questions*	*BLT*	*CO*	*Marks*
11.	a	i	Define data and big data. Also explain structured, semi structured and unstructured data.	L2	1	8
		ii	Explain the characteristics of big data and illustrate, how data is big with an example of E- Commerce?	L3	1	8
			Or			
	b	i	Discuss in brief about the basic building blocks of Hadoop.	L2	1	8
		ii	Explain the concept of crowd sourcing analytics and its applications in business environments provide specific examples.	L3	1	8

12.	a	An E-commerce company that wants to personalize product recommendations for its customer in real-time. Describe how to use a NoSQL database to store and manage customer profile and purchase history. Discuss the data model and indexing techniques to ensure data consistency and accuracy.	L3	2	16
		Or			
	b	Discuss how the Cassandra data model can be applied to real-world scenarios like product catalog management, order tracking, and user activity logs. Analyze how the Cassandra data model is used to design scalable and high-performance applications.	L3	2	16
13.	a i	Discuss the fundamentals architectural components and design principle of Hadoop distributed file system.	L2	3	8
	ii	Enumerate the HDFS write and read pipelines in detail. Explain how data is integrated and stored in HDFS including block replication and fault tolerances mechanisms.	L2	3	8
		Or			
	b	Discuss how Hadoop can be integrated with other technologies to enhance data processing and analytics capabilities.	L2	3	16
14.	a	Analyze how the MapReduce framework is utilized to process large-scale datasets in distributed computing environments. Using a case study of a social media platform, demonstrate how MapReduce can be applied to real-world scenarios.	L4	4	16
		Or			
	b	Examine, how YARN is used to manage resources in a Hadoop ecosystem for running distributed applications. Explain the key components of YARN and their roles in resource allocation and task scheduling.	L4	4	16
15.	a	Examine the HBase data model and its implementation in managing large-scale, real-time datasets. Using a case study of a telecommunications company, illustrate how HBase can be implemented to handle real-time customer data, call records, and service usage.	L4	5	16
		Or			
	b	Using a case study of an e-commerce platform analyzing customer purchase behavior, demonstrate how HiveQL can be used to execute various queries, such as data aggregation, filtering, and joining multiple tables.	L4	5	16