

THE KAVERY ENGINEERING COLLEGE*(An Autonomous Institution)***B.E/B.TECH DEGREE END SEMESTER THEORY EXAMINATIONS, NOV/DEC 2025***Fifth Semester**Artificial Intelligence and Data Science***CCS334 - Big Data Analytics****(Common to Computer Science and Engineering & Information Technology)***Regulations - 2021**Duration : 3 Hrs**Maximum : 100 Marks***PART - A (ANSWER ALL QUESTIONS) (Marks 10*2=20)**

<i>Q.No</i>	<i>Questions</i>	<i>BTL</i>	<i>CO</i>	<i>Marks</i>
1.	Define big data and list its key characteristics.	L1	1	2
2.	List any two industry examples where big data analytics is used.	L2	1	2
3.	What is a NoSQL database? Mention two features.	L1	2	2
4.	Differentiate between Key-Value and document data models.	L2	2	2
5.	State the role of HDFS in Hadoop ecosystem.	L1	3	2
6.	What is serialization in Hadoop? Give one example.	L1	3	2
7.	List any two failures that occur in classic MapReduce.	L1	4	2
8.	Explain shuffle and sort phase in MapReduce.	L2	4	2
9.	Mention any two features of HBase.	L1	5	2
10.	Define Pig Latin.	L1	5	2

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

<i>Q.No</i>	<i>Questions</i>	<i>BLT</i>	<i>CO</i>	<i>Marks</i>
11.	a i Explain convergence of key trends leading to Big Data growth.	L2	1	8
	ii Classify the role of Cloud and Mobile BI in Big Data applications.	L2	1	8
	Or			
	b i Explain web analytics and its importance with examples.	L2	1	8
	ii Illustrate crowdsourcing analytics with case examples.	L2	1	8
12.	a i Summarize the aggregate data models in NoSQL with neat diagram.	L2	2	8
	ii Clarify the Cassandra data model with an example.	L2	2	8
	Or			
	b i Compare graph databases with relational databases.	L2	2	8
	ii Explain consistency models in NoSQL.	L2	2	8

13.	a	i	Enlighten the architecture and concepts of HDFS.	L2	3	8
		ii	Illustrate Hadoop I/O, data integrity and compression techniques.	L2	3	8
Or						
	b	i	Enlighten the process of using Apache Spark with Cassandra in a Hadoop ecosystem. What are the performance considerations?	L3	3	8
		ii	Deliberate different serialization frameworks used in Hadoop (Writable, Avro, Protocol Buffers, Thrift). Compare them based on performance, ease of use, and compatibility.	L3	3	8
14.	a	i	Evaluate the efficiency and limitations of each phase in a MapReduce job run. Suggest improvements or alternatives based on workload characteristics.	L5	4	8
		ii	Evaluate the effectiveness of YARN's job scheduling and task execution strategies across different workload types.	L5	4	8
	Or					
	b	i	Analyze the architectural differences between Classic MapReduce and YARN, and explain how YARN overcomes the limitations of the classic model.	L4	4	8
				ii	Explicate input and output formats in MapReduce with examples.	L4 4 8
15.	a	i	Describe the HBase data model with suitable example.	L3	5	8
		ii	Discuss the features and applications of HiveQL queries.	L3	5	8
	Or					
	b	i	Explain Pig Latin scripts development and testing with examples.	L2	5	8
				ii	Compare HBase, Pig and Hive in terms of data handling capabilities.	L2 5 8