Semester: 7th	Department of AIML	AY: 2026-27
Course Code: BCS741D	Course Name: Big Data Analytics	

Question Bank

Q.	Module-01	M	L	C
1.	Define Big Data Analytics. Explain classification of analytics	10	L2	CO1
2.	Define Big Data. Explain its characteristics and challenges	10	L2	CO1
3.	List and explain the terminologies used in Big Data Environment	10	L2	CO1
4.	Explain Hadoop ecosystem components for data processing and data analysis.	10	L2	CO1
5.	What is NoSQL? Explain the different types of NoSQL databases with example.	10	L2	CO1
6.	Explain shared nothing architecture, CAP theorem and list its advantages.	10	L2	CO1
7.	Explain classification of BigData with diagram.	10	L2	CO1
8.	Explain CAP theorem	10	L2	CO1
9.	List and Explain terminologies used in BigData environment.	10	L2	CO1
10.	What is NoSql?. Explain the characteristics of NoSql.	10	L2	CO1
	Module-02			
11.	With a neat diagram explain Hadoop distributed file system architecture.	9	L2	CO2
12.	Differentiate between. i. RDBMS and Hadoop ii. Mapper and Reducer iii. Traditional BI and Big Data Architecture	10	L3	CO2
13.	Identify and explain key aspects and components of Hadoop	10	L3	CO2
14.	Give the HDFS commands to perform the following operations: i) To get the list of directories and files at the root of HDFS ii) To create a directory (Say sample) in HDFS iii) To copy a file from HDFS to local file system iv) To display the contents of an IIDFS file on console.	10	L3	CO2
15.	Identify and explain different phases of map task and reduce task for the word count problem.	10	L3	CO2
16.	Explain anotomy of HDFS file read and write.	10	L2	CO2
17.	With a neat diagram, demonstrate YARN architecture	10	L3	CO2
18.	Explain any two HDFS commands with Example	10	L2	CO2
19.	Implement a word count program in Hadoop	10	L2	CO2

Aditya College of Engineering and Technology

20.	Explain ease of working with structured data and limitations.	10	L2	CO2
	Module-03			
21.	What is CRUD operation in Mongo DB? Explain the following commands. a) Insert method b) Update on array c) Save method	10	L1	CO3
22.	Determine and Explain the characteristics of MongoDB	10	L3	CO3
23.	Write short notes on. a) Aggregate function b) Map reduce function in Mongo DB	10	L2	CO2
24.	Determine and explain creation of database, dropping of database and datatypes of Mongo DB	10	L3	CO2
25.	What is MongoDB? Explain the key differences between RDBMS and MongoDB with respect to terminology such as Table, Row, Column, Index, and Join. Also discuss the data modeling philosophy of MongoDB.	10	L3	CO3
26.	List and explain the various data types supported by MongoDB with suitable examples for each, including ObjectId, String, Array, Embedded Document, Date, Boolean, and Null.	10	L3	CO3
27.	Illustrate the use of query operators such as $gt, $lt, $in, $and, and $or with appropriate examples.	10	L3	CO3
28.	Why is MongoDB preferred over traditional RDBMS systems for big data and web-scale applications? Discuss any five advantages of MongoDB, supporting your answer with relevant real-world use cases.	10	L3	CO3
29.	Explain MongoDB Query Language (MQL) with examples demonstrating the following operations: insert, find, update, and delete.	10	L3	CO3
30.	What is MongoDB? Explain the key differences between RDBMS and MongoDB.	10	L2	CO3
	Module-04			
31.	Define Hive. Explain its main task and features.	10	L1	CO4
32.	Analyze Hive architecture with its diagram.	10	L4	CO4
33.	Define Pig. Explain its feature, anatomy and philosophy.	10	L1	CO4
34.	Examine identifiers, keywords, data types and operators of Pig with an example.	10	L4	CO4
35.	Explain the primitive and complex data types supported by Hive. Also describe the Hive file formats.	10	L4	CO4
36.	What is a User Defined Function (UDF) in Apache Hive? With a step-by-step Java example, explain how to create, compile, package, register, and invoke a custom UDF in HiveQL.	10	L4	CO4

Aditya College of Engineering and Technology

37.	With a neat diagram, explain the architecture of Apache Hive. Describe the role of each component: Metastore, Driver, Compiler, Optimizer, Executor, and HiveServer2 in the query processing lifecycle.	10	L4	CO4
38.	Explain the anatomy of Apache Pig and how it runs on Hadoop. Describe the Pig philosophy, its two execution modes (Local and MapReduce), and the significance of the HDFS commands used in a Pig workflow.	10	L4	CO4
39.	With appropriate Pig Latin scripts, demonstrate the use of the following relational operators: LOAD, FILTER, GROUP, JOIN, ORDER BY, and STORE. Also explain the complex data types in Pig (Tuple, Bag, Map) with examples, and describe at least two built-in eval functions.	10	L4	CO4
40.	Compare Apache Pig and Apache Hive across the following dimensions: programming model, ease of use, execution engine, schema enforcement, ideal use cases, and support for User Defined Functions. Present your comparison in a tabular format and conclude with when to use Pig over Hive and vice versa.	10	L4	CO4
	Module-05			
41.	Analyze spark SQL and Pandas for data analysis.	10	L4	CO4
42.	Determine and explain component, features of spark architecture.	10	L3	CO5
43.	Analyze different phase of text mining process with a diagram.	10	L4	CO5
44.	Identify and explain preprocessing steps, mining task for web content.	10	L3	CO5
45.	Discuss the different components of the Apache Spark ecosystem and explain their applications in Big Data processing.	10	L3	CO5
46.	Explain the process of data analysis using Apache Spark with a suitable example.	10	L3	CO5
47.	Compare Apache Spark and Hadoop MapReduce with respect to architecture, processing speed, memory utilization, and applications.	10	L3	CO5
48.	Explain the PageRank algorithm with its working principle, mathematical formula, and a suitable example.	10	L3	CO5
49.	Explain the structure of the World Wide Web and describe the process of analyzing a web graph, including nodes, edges, in-degree, and out-degree.	10	L3	CO5
50.	Explain Apache Spark architecture with a neat diagram. Discuss the functions of the Driver Program, Cluster Manager, and Executors.	10	L3	CO5

Aditya College of Engineering and Technology