Course Code: 23CS1104 **R23**

SIDDHARTH INSTITUTE OF ENGINEERING &TECHNOLOGY:: PUTTUR (AUTONOMOUS)
Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code: BIG DATA ANALYTICS (23CS1104) **Course & Branch**: B.Tech - CAD
Year &Sem: III-B.Tech& II-Sem **Regulation:** R23

UNIT –I

Data Structure in Java

1	**a**	Define a data structure.	**[L1][CO1]**	**[2M]**
	b	Define Linked List?	**[L1][CO1]**	**[2M]**
	c	Define Generics Class in Java.	**[L1][CO1]**	**[2M]**
	d	What is a Wrapper class in java?	**[L1][CO1]**	**[2M]**
	e	Define serialization in java	**[L1][CO1]**	**[2M]**
2	**a**	Explain the structure and working of a Linked List.	**[L2][CO1]**	**[5M]**
	b	Illustrate Stack operations with an example.	**[L2][CO1]**	**[5M]**
3		Analyze Linked List in detail. Explain its operations, advantages, and disadvantages with examples.	**[L4][CO1]**	**[10M]**
4	**a**	Apply the Queue data structure to demonstrate its operations with a suitable example.	**[L3][CO1]**	**[5M]**
	b	Compare Stack and Queue.	**[L4][CO1]**	**[5M]**
5		Explain Stack data structure. Evaluate its applications in real-time systems.	**[L4][CO1]**	**[10M]**
6	**a**	Explain Set interface and its Methods.	**[L2][CO1]**	**[5M]**
	b	Describe Map interface with suitable examples.	**[L2][CO1]**	**[5M]**
7	**a**	Demonstrate the use of Generics with a simple Java example.	**[L2][CO1]**	**[5M]**
	b	Apply Wrapper classes in Java to demonstrate their need in handling primitive data types with suitable examples.	**[L3][CO1]**	**[5M]**
8	**a**	Analyze the advantages of using Generics in Java programs.	**[L4][CO1]**	**[5M]**
	b	**Justify** the use of Serialization in Java for object persistence with suitable examples.	**[L5][CO1]**	**[5M]**
9		Analyze Sets and Maps in Java Collections and justify their use with examples.	**[L4][CO1]**	**[10M]**
10		Describe Serialization in Java. Create a program to serialize and deserialize an object.	**[L6][CO1]**	**[10M]**

Course Code: 23CS1104 **R23**

UNIT –II

Working with Big Data

1	**a**	Define Big Data.	**[L1][CO2]**	**[2M]**
	b	State the role of NameNode in Hadoop.	**[L1][CO2]**	**[2M]**
	c	Define Secondary NameNode.	**[L1][CO2]**	**[2M]**
	d	What is Job Tracker?	**[L1][CO2]**	**[2M]**
	e	What is meant by pseudo-distributed mode in Hadoop?	**[L1][CO2]**	**[2M]**
	f	**Discuss the Usage of Big Data Analytics**	**[L2][CO2]**	**[2M]**
	g	what are the 5 V's in big data	**[L1][CO2]**	**[2M]**
2		Apply the Google File System (GFS) architecture to illustrate how large files are stored, accessed, and managed in a distributed environment..	**[L3][CO2]**	**[10M]**
3		Apply the HDFS architecture to illustrate how data is stored and accessed in a Hadoop cluster with a neat diagram.	**[L3][CO2]**	**[10M]**
4		Analyze the importance of Hadoop ECO System in detail.	**[L4][CO2]**	**[10M]**
5		Illustrate the procedure for installing and configuring Hadoop in pseudo-distributed mode with suitable commands.	**[L3][CO2]**	**[10M]**
6	**a**	Explain the **HDFS Read** INTERFACES in detail	**[L2][CO2]**	**[5M]**
	b	Explain the **HDFS write** INTERFACES in detail	**[L2][CO3]**	**[5M]**
7	**a**	Explain the functions of NameNode and DataNode.	[L2][CO2]	[5M]
	b	Explain Job Tracker and Task Tracker with their responsibilities.	**[L2][CO2]**	**[5M]**
8		Apply the different Hadoop cluster modes to demonstrate their use in suitable application scenarios.	**[L3][CO2]**	**[10M]**
9		Analyze the importance of Secondary NameNode in HDFS.	**[L4][CO2]**	**[10M]**
10		Design a Hadoop configuration setup using XML files and explain how they control cluster behavior.	**[L6][CO2]**	**[10M]**

UNIT –III

Writing Map Reduce Programs

1	**a**	Define Mapper in Hadoop MapReduce.	**[L1][CO3]**	**[2M]**
	b	What is the function of Reducer?	**[L1][CO3]**	**[2M]**
	c	What is a Driver program in MapReduce?	**[L1][CO3]**	**[2M]**
	d	What is meant by Hadoop MapReduce API?	**[L1][CO3]**	**[2M]**
	e	What is the role of a Partitioner in MapReduce? ….	**[L1][CO4]**	**[2M]**
2	**a**	Explain the MapReduce programming model with a neat diagram.	**[L2][CO3]**	**[5M]**
	b	Describe the working of Mapper and Reducer phases.	**[L2][CO3]**	**[5M]**
3		Compare and evaluate old and new Hadoop MapReduce APIs.	**[L4][CO3]**	**[10M]**
4	**a**	Discuss the various failures involved in Map Reduce	**[L2][CO3]**	**[5M]**
	b	Illustrate the components of Hadoop MapReduce API.	**[L2][CO3]**	**[5M]**
5		Analyze the complete execution flow of a MapReduce program in Hadoop.	**[L4][CO3]**	**[10M]**
6		Demonstrate a simple MapReduce program using Weather dataset.	**[L3][CO3]**	**[10M]**
7	**a**	Apply the Combiner and Partitioner in a MapReduce program to improve performance and data distribution.	**[L3][CO4]**	**[5M]**
	b	Apply the Driver code in a MapReduce program to configure and execute a job, illustrating its purpose.	**[L3][CO4]**	**[5M]**
8	**a**	Analyze how Record Reader processes input data.	**[L4][CO4]**	**[5M]**
	b	Explain the data flow in a MapReduce job.	**[L2][CO4]**	**[5M]**
9		Create a complete MapReduce program including Driver, Mapper, and Reducer codes.	**[L6][CO4]**	**[10M]**
10		Analyze and evaluate scenarios where the use of Combiner and Partitioner significantly improves performance.	**[L5][CO4]**	**[10M]**

Course Code: 23CS1104 **R23**

UNIT –IV

Stream Memory and Spark

1	**a**	What is stream computing?	**[L1][CO5]**	**[2M]**
	b	What is stream data model?	**[L1][CO5]**	**[2M]**
	c	Define Apache Spark.	**[L1][CO5]**	**[2M]**
	d	What is Spark RDD?	**[L1][CO5]**	**[2M]**
	e	State any two features of Spark.	**[L1][CO5]**	**[2M]**
2	**a**	Explain stream data model and architecture.	**[L2][CO5]**	**[5M]**
	b	Describe the concept of stream computing.	**[L2][CO5]**	**[5M]**
3	**a**	Apply sampling techniques in data streams	**[L3][CO5]**	**[5M]**
	b	Apply filtering operations in a stream processing system	**[L3][CO5]**	**[5M]**
4		Demonstrate basic Spark RDD operations with examples.	**[L2][CO5]**	**[10M]**
5		Analyze stream memory concepts and explain how data streams are processed efficiently.	**[L4][CO5]**	**[10M]**
6	**a**	Explain methods for counting distinct elements in a data stream.	**[L2][CO5]**	**[5M]**
	b	Describe Spark architecture and its components.	**[L2][CO5]**	**[5M]**
7		Analyze the advantages of Spark over traditional batch processing.	**[L4][CO5]**	**[10M]**
8	**a**	Explain the steps involved in installing Apache Spark.	**[L2][CO5]**	**[5M]**
	b	Explain the concept and characteristics of Spark RDD.		**[5M]**
9		Explain Spark architecture in detail and evaluate the role of each component.	**[L4][CO5]**	**[10M]**
10		Design and explain a Spark application using RDD operations.	**[L6][CO5]**	**[10M]**

Course Code: 23CS1104 **R23**

UNIT –V
Pig & Applying Structure to Hadoop with Hive

1	**a**	Define Apache Pig.	**[L1][CO6]**	**[2M]**
	b	What are the modes of running Pig scripts?	**[L1][CO6]**	**[2M]**
	c	What is Apache Hive?	**[L1][CO6]**	**[2M]**
	d	What is HiveQL?	**[L1][CO6]**	**[2M]**
	e	What is meant by querying in Hive?	**[L1][CO6]**	**[2M]**
2	**a**	Explain Pig architecture with a neat diagram.	**[L2][CO6]**	**[5M]**
	b	Describe the Pig Latin application flow.	**[L2][CO6]**	**[5M]**
3	**a**	Explain basic Pig Latin operators with examples.	**[L2][CO6]**	**[5M]**
	b	Compare local and distributed modes of running Pig scripts.	**[L4][CO6]**	**[5M]**
4	**a**	Explain different Pig script interfaces.	**[L2][CO6]**	**[5M]**
	b	Demonstrate the working of Apache Hive architecture	**[L3][CO6]**	**[5M]**
5	**a**	Apply different Hive data types with suitable examples.	**[L3][CO6]**	**[5M]**
	b	Demonstrate creation and management of Hive databases and tables.	**[L2][CO6]**	**[5M]**
6		Explain Hive Data Manipulation Language (DML).	**[L2][CO6]**	**[10M]**
7		Analyze the role of Hive in querying and analyzing Big Data.	**[L4][CO6]**	**[10M]**
8		Analyze the Pig architecture and explain how Pig simplifies Hadoop programming.	**[L4][CO6]**	**[10M]**
9		Design and explain Pig Latin scripts for processing large datasets.	**[L6][CO6]**	**[10M]**
10		Create and explain Hive queries for database creation, data manipulation, and analysis.	**[L6][CO6]**	**[10M]**