Course Code: 23CS1208 **R23**

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

Subject with Code: BIG DATA (23CS1208) **Course & Branch**: B.Tech –CCC

Regulation: R23 **Year & Sem:** III-B.Tech & II-Sem

UNIT –I
INTRODUCTION TO BIG DATA AND HADOOP

1	a)	Define structured data with one example.	**[L1][CO1]**	**[2M]**
	b)	What is Velocity in Big Data?	**[L1][CO1]**	**[2M]**
	c)	Mention two challenges in Big Data Analytics	**[L1][CO1]**	**[2M]**
	d)	Name the two core components of Hadoop.	**[L1][CO1]**	**[2M]**
	e)	Define DataNode.	**[L1][CO1]**	**[2M]**
2		Discuss in detail about Apache Hadoop and History of Hadoop.	**[L2][CO1]**	**[10M]**
3	a)	Examine the different types of digital data with examples.	**[L4][CO1]**	**[5M]**
	b)	Examine the Significance of big data analytics.	**[L2][CO1]**	**[5M]**
4		Establish the evolution of Hadoop ecosystem with neat diagram.	**[L3][CO1]**	**[10M]**
5		Distinguish between structure, unstructured and semi-structure data with an examples.	**[L4][CO1]**	**[10M]**
6	a)	List out the challenges faced by big data.	**[L1][CO1]**	**[5M]**
	b)	Summarize Big Data in terms of three dimensions, volume, variety and velocity.	**[L3][CO1]**	**[5M]**
7		Discuss about the Analysis of data through Unix tools and Hadoop	**[L2][CO1]**	**[10M]**
8	a)	Define big data analytics. Identify the Classification of Analytics.	**[L2][CO1]**	**[5M]**
	b)	Illustrate in detail about Hadoop streaming.	**[L2][CO1]**	**[5M]**
9		Explain the of Hadoop framework with neat diagram.	**[L2][CO1]**	**[10M]**
10		Discriminate the Big Data in Healthcare, Transportation & Medicine.	**[L4][CO1]**	**[10M]**
11		Generalize the list of tools related to Hadoop.	**[L3][CO1]**	**[10M]**

Course Code: 23CS1208 **R23**

UNIT –II
HDFS(HADOOP DISTRIBUTED FILE SYSTEM)

1	a)	What is HDFS?	**[L1][CO2]**	**[2M]**
	b)	Explain data write operation in HDFS.	**[L1][CO2]**	**[2M]**
	c)	Differentiate Flume and Sqoop	**[L1][CO2]**	**[2M]**
	d)	What is serialization?	**[L1][CO2]**	**[2M]**
	e)	Define Avro schema.	**[L1][CO2]**	**[2M]**
2		Illustrate the concepts of HDFS.	**[L2][CO2]**	**[10M]**
3		What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with neat diagram.	**[L2][CO2]**	**[10M]**
4		Explain the block, name node and data node in Hadoop file system.	**[L2][CO2]**	**[10M]**
5		Determine the basic commands in Hadoop command line interface.	**[L4][CO2]**	**[10M]**
6	a)	Discuss about the data ingest operation using Sqoop.	**[L2][CO2]**	**[5M]**
	b)	Discuss about the data ingest operation using Flume.	**[L2][CO2]**	**[5M]**
7		Infer File read and File write operations in HDFS.	**[L2][CO2]**	**[10M]**
8		Establish the evolution of Hadoop ecosystem with neat diagram.	**[L3][CO2]**	**[10M]**
9		Elaborate the AVRO file format with diagram.	**[L4][CO2]**	**[10M]**
10	a)	Describe the dataflow process in Hadoop Distributed file System.	**[L2][CO2]**	**[5M]**
	b)	Analyze the features of Apache Hadoop.	**[L4][CO2]**	**[5M]**
11	a)	Justify File Based Data structures.	**[L6][CO2]**	**[5M]**
	b)	Differentiate Sqoop and Flume operation.	**[L2][CO2]**	**[5M]**

UNIT –III

MAP REDUCE

1	a)	Define a MapReduce	**[L1][CO3]**	**[2M]**
	b)	What is TaskTracker?	**[L1][CO3]**	**[2M]**
	c)	What is FIFO scheduler?	**[L1][CO3]**	**[2M]**
	d)	Name any two InputFormats.	**[L1][CO4]**	**[2M]**
	e)	Mention limitation of MapReduce.	**[L1][CO4]**	**[2M]**
2		Examine the Anatomy of a MapReduce Job Run.	**[L4][CO3]**	**[10M]**
3		Sketch neatly and Explain MapReduce Architecture in detail.	**[L3][CO3]**	**[10M]**
4		Explain in detail about Hadoop YARN Architecture.	**[L2][CO3]**	**[10M]**
5	a)	Discuss different types of failures in Classic MapReduce.	**[L2][CO3]**	**[5M]**
	b)	Discuss different types of failures in Yet Another Resource Negotiator.	**[L2][CO3]**	**[5M]**
6		Examine the different types of Job Scheduling process in Map Reduce.	**[L2][CO4]**	**[10M]**
7		Establish the evolution of Yet Another Resource Negotiator framework with neat diagram.	**[L4][CO4]**	**[10M]**
8	a)	Illustrate Shuffle operations in MapReduce.	**[L2][CO4]**	**[5M]**
	b)	Illustrate Sort operations in MapReduce.	**[L2][CO4]**	**[5M]**
9		Categorize different types of MapReduce input formats.	**[L4][CO4]**	**[10M]**
10		Justify types of output formats in MapReduce.	**[L3][CO4]**	**[10M]**
11		Discriminate the below features in MapReduce. i) Counters ii) Sorting iii) Joins	**[L4][CO4]**	**[10M]**

Course Code: 23CS1208 **R23**

UNIT –IV
HADOOP ECO SYSTEM-PIG

1	a)	Mention two advantages of Apache Pig.	**[L1][CO5]**	**[2M]**
	b)	Compare Pig and SQL.	**[L1][CO5]**	**[2M]**
	c)	What is a tuple?	**[L1][CO5]**	**[2M]**
	d)	Define STORE operator.	**[L1][CO5]**	**[2M]**
	e)	Define Pig.	**[L1][CO5]**	**[2M]**
2		Explain Apache Pig Architecture in detail.	**[L2][CO5]**	**[10M]**
3		Evaluate the Expressions and types in Pig Latin.	**[L4][CO5]**	**[10M]**
4		Compare the PIG with SQL Databases with an Example.	**[L4][CO5]**	**[10M]**
5		Examine the different execution modes available in Pig.	**[L3][CO5]**	**[10M]**
6		Explain Pig Latin datatypes and data model in detail.	**[L2][CO5]**	**[10M]**
7	a)	Explain about Arithmetic Operators in Pig Latin.	**[L2][CO5]**	**[5M]**
	b)	Find the Grouping and Joining Data in Pig Latin.	**[L2][CO5]**	**[5M]**
8		Examine the Pig Latin Relational Operators.	**[L3][CO5]**	**[10M]**
9		Develop Pig Latin Schemas and Functions.	**[L4][CO5]**	**[10M]**
10	a)	Explain about the data types in Pig Latin.	**[L2][CO5]**	**[5M]**
	b)	Illustrate the Merits and Demerits of Apache Pig.	**[L2][CO5]**	**[5M]**
11		Evaluate Data Processing Operators in Pig Latin.	**[L4][CO5]**	**[10M]**

Course Code: 23CS1208 **R23**

UNIT –V
HIVE, HBASE

1	a)	What is the purpose of Hive shell?	**[L1][CO6]**	**[2M]**
	b)	What is Hive Metastore?	**[L1][CO6]**	**[2M]**
	c)	Define Hive table	**[L1][CO6]**	**[2M]**
	d)	Name any two Hive services	**[L1][CO6]**	**[2M]**
	e)	Define Hbase.	**[L1][CO6]**	**[2M]**
2		Illustrate Hive table with example.	**[L2][CO6]**	**[10M]**
3		Explain Hive Query Language (HiveQL).	**[L2][CO6]**	**[10M]**
4		Explain about Hive shell architecture and components in a neat sketch.	**[L2][CO6]**	**[10M]**
5	a)	Deduce the various services offered by Hive.	**[L4][CO6]**	**[5M]**
	b)	Examine the Characteristics of HBase.	**[L4][CO6]**	**[5M]**
6		Discuss about Hive shell command line interface.	**[L2][CO6]**	**[10M]**
7	a)	Appraise about Hive query language in detail.	**[L4][CO6]**	**[5M]**
	b)	Illustrate Metastore Hive in detail.	**[L2][CO6]**	**[5M]**
8		Differentiate Hbase over RDBMS.	**[L4][CO6]**	**[10M]**
9		Explain with a neat diagram the architecture of Hbase.	**[L2][CO6]**	**[10M]**
10	a)	Categorize the joins in HiveQL.	**[L4][CO6]**	**[5M]**
	b)	Illustrate the HBase basics in detail.	**[L2][CO6]**	**[5M]**
11		Explain the core concepts of HBase.	**[L2][CO6]**	**[10M]**

Prepared by:Dr.J.Sridhar