



**SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY:: PUTTUR
(AUTONOMOUS)**

Siddharth Nagar, Narayanavanam Road – 517583

QUESTION BANK (DESCRIPTIVE)

Subject with Code: BIG DATA(20CS0538)

Course & Branch: B.Tech – CSE,CSM

Regulation: R20

Year & Sem: IV-B.Tech & I-Sem

**UNIT –I
INTRODUCTION TO BIG DATA AND HADOOP**

1		Discuss in detail about Apache Hadoop and History of Hadoop?	[L2][CO1]	[12M]
2	a)	Examine the different types of digital data with examples?	[L4][CO1]	[6M]
	b)	Summarize Big Data in terms of three dimensions, volume, variety and velocity.	[L2][CO1]	[6M]
3		Establish the evolution of Hadoop ecosystem with neat diagram.	[L3][CO1]	[12M]
4		Distinguish between structure, unstructured and semi-structure data with an examples.	[L4][CO1]	[12M]
5	a)	List out the challenges faced by big data.	[L1][CO1]	[6M]
	b)	Examine the Significance of big data analytics.	[L3][CO1]	[6M]
6		Discuss about the Analysis of data through Unix tools and Hadoop	[L2][CO1]	[12M]
7	a)	What is big data analytics? Identify the Classification of Analytics.	[L3][CO1]	[6M]
	b)	Illustrate in detail about Hadoop streaming	[L2][CO1]	[6M]
8	a)	What is Big Sheets? What can be done with big sheets?	[L1][CO6]	[6M]
	b)	Explain in detail about Infosphere Big Insights.	[L2][CO1]	[6M]
9	a)	Discriminate the Big Data in Healthcare, Transportation & Medicine.	[L5][CO1]	[6M]
	b)	Why organizations using big data for competitive advantage?	[L4][CO1]	[6M]
10	a)	How to implement IBM Big Data Strategy?	[L2][CO1]	[6M]
	b)	Generalize the list of tools related to Hadoop.	[L6][CO1]	[6M]

UNIT –II
HDFS(HADOOP DISTRIBUTED FILE SYSTEM)

1		Illustrate the concepts of HDFS.	[L3][CO2]	[12M]
2		What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with neat diagram	[L1][CO2]	[12M]
3		Explain the block, name node and data node in Hadoop file system	[L2][CO2]	[12M]
4		Determine the basic commands in Hadoop command line interface	[L3][CO2]	[12M]
5	a)	What is an interface? Establish the Hadoop system interfaces	[L3][CO2]	[6M]
	b)	Generalize about Hadoop Archives and its Limitations	[L2][CO2]	[6M]
6		Infer File read and File write operations in HDFS	[L2][CO2]	[12M]
7	a)	Discuss about the data ingest operation using Sqoop and Flume	[L2][CO2]	[6M]
	b)	Differentiate the compression and serialization operation in Hadoop I/O.	[L4][CO2]	[6M]
8		Elaborate the AVRO file format with diagram.	[L6][CO2]	[12M]
9	a)	Describe the dataflow process in Hadoop Distributed file System	[L3][CO2]	[8M]
	b)	Analyze the features of Apache Hadoop	[L4][CO2]	[4M]
10	a)	Justify File Based Data structures.	[L5][CO2]	[6M]
	b)	Differentiate Sqoop and Flume?	[L2][CO2]	[6M]

UNIT –III
MAP REDUCE

1		Examine the Anatomy of a MapReduce Job Run.	[L4][CO3]	[12M]
2		Sketch neatly and Explain MapReduce Architecture in detail.	[L3][CO3]	[12M]
3		Explain in detail about Hadoop YARN Architecture	[L2][CO3]	[12M]
4	a)	Discuss different types of failures in Classic MapReduce.	[L2][CO3]	[6M]
	b)	List out the different types of failures in Yet Another Resource Negotiator.	[L1][CO3]	[6M]
5	a)	Examine the different types of Job Scheduling process in Map Reduce.	[L3][CO3]	[6M]
	b)	Identify the Significance of YARN over MapReduce.	[L3][CO3]	[6M]
6		Illustrate Shuffle and Sort operations in MapReduce.	[L5][CO4]	[12M]
7	a)	Summarize Task Execution Environment Properties.	[L2][CO4]	[6M]
	b)	Discuss about Speculative Execution and its Properties.	[L2][CO4]	[6M]
8		Categorize different types of MapReduce input formats.	[L4][CO4]	[12M]
9		Justify types of output formats in MapReduce.	[L5][CO4]	[12M]
10		Discriminate the below features in MapReduce. a) Counters b) Sorting c) Joins	[L4][CO4]	[12M]

UNIT –IV
HADOOP ECO SYSTEM-PIG

1	a)	Illustrate the concept of grunt	[L3][CO2]	[5M]
	b)	Why Do We Need Apache Pig? Identify the features of PIG.	[L4][CO2]	[7M]
2		What is Pig? How to Install and execute PIG on Hadoop Cluster	[L2][CO5]	[12M]
3	a)	Compare the PIG with Databases with an Example	[L5][CO3]	[6M]
	b)	Evaluate the Expressions and types in Pig Latin.	[L4][CO4]	[6M]
4		Examine the different execution modes available in Pig	[L3][CO4]	[12M]
5		Construct User Define Functions in Pig Latin.	[L6][CO5]	[12M]
6	a)	Explain about Arithmetic Operators in Pig Latin .	[L2][CO3]	[6M]
	b)	Find the Grouping and Joining Data in Pig Latin.	[L3][CO3]	[6M]
7		Examine the Pig Latin Relational Operators.	[L4][CO2]	[12M]
8		Develop Pig Latin Schemas and Functions.	[L3][CO5]	[12M]
9	a)	Explain about the data types in Pig Latin.	[L2][CO2]	[6M]
	b)	Develop a program to calculate the maximum recorded temperature by year for the weather dataset in Pig Latin.	[L6][CO5]	[6M]
10	a)	Discriminate the Structures, Statements in Pig Latin	[L4][CO1]	[6M]
	b)	Evaluate Data Processing Operators in Pig Latin.	[L5][CO4]	[6M]

UNIT –V
HIVE, HBASE, BIG SQL

1		Illustrate Hive table with example.	[L3][CO5]	[12M]
2		Discuss about Hive shell command line interface.	[L2][CO5]	[12M]
3	a)	Draw a neat sketch of Hive architecture.	[L3][CO2]	[4M]
	b)	Explain about components of Hive architecture.	[L2][CO2]	[8M]
4	a)	Deduce the various services offered by Hive.	[L4][CO4]	[6M]
	b)	Examine the Characteristics of HBase	[L4][CO1]	[6M]
5	a)	Infer the advantages of Hive over traditional databases?	[L2][CO5]	[6M]
	b)	What are the operators and functions in HIVE?	[L1][CO2]	[6M]
6	a)	Appraise about Hive query language?	[L4][CO5]	[6M]
	b)	Review Metastore in Hive?	[L2][CO5]	[6M]
7		Differentiate Hbase over RDBMS.	[L4][CO1]	[12M]
8		Explain with a neat diagram the architecture of Hbase.	[L2][CO2]	[12M]
9	a)	Categorize the joins in HiveQL	[L4][CO5]	[6M]
	b)	Report the Implementation of queries on sorting and aggregation of data in Hive	[L6][CO3]	[6M]
10	a)	Explain about IBM Big SQL?	[L2][CO6]	[6M]
	b)	Assess how HBase is implemented at Streamy.com	[L4][CO6]	[6M]

Prepared by:Dr.R.Elankavi,T.Mekala Rani, P.Balaji