SIDDARTHA INSTITUTE OF SCIENCE AND TECHNOLOGY:: PUTTUR
(AUTONOMOUS)
Siddharth Nagar, Narayanavanam Road – 517583
QUESTION BANK (DESCRIPTIVE)

Subject with Code: BIG DATA ANALYTICS(18CS0538)   **Course & Branch**: B.Tech - CSE

Regulation: R18   **Year & Sem:** IV-B.Tech & I-Sem

UNIT –I
Introduction To Big Data And Hadoop

1	a)	Define data?	[L1][CO1]	**[2M]**
	b)	What are the four V's of big data.	[L1][CO1]	**[2M]**
	c)	Summarize the data types for Big data	[L2][CO1]	**[2M]**
	d)	Define Hadoop streaming	[L1][CO2]	**[2M]**
	e)	Formulate Hadoop development	[L6][CO2]	**[2M]**
2	a)	Examine the different types of digital data with examples?	[L4][CO1]	**[5M]**
	b)	Discuss Big Data in terms of three dimensions, volume, variety and velocity.	[L2][CO1]	**[5M]**
3		Establish the Evolution of Hadoop ecosystem with neat diagram.	[L3][CO2]	**[10M]**
4		Explain the difference between structure, unstructured and semi-structure data with an examples.	[L4][CO1]	**[10M]**
5	a)	List the Top challenges facing big data.	[L1][CO1]	**[5M]**
	b)	What is the Significance of big data analytics	[L1][CO1]	**[5M]**
6		Distinguish between Analysis of data through Unix tools and Hadoop Ecosystem	[L4][CO5]	**[10M]**
7	a)	What is big data analytics? Identify the Classification of Analytics	[L3][CO1]	**[5M]**
	b)	Illustrate in detail about Hadoop streaming	[L2][CO2]	**[5M]**
8	a)	What is Big Sheets? What can be done with big sheets?	[L1][CO6]	**[5M]**
	b)	Explain in detail about Infosphere Big Insights ?	[L2][CO6]	**[5M]**
9	a)	Discriminate the Big Data in Healthcare,Trasportation & Medicine.	[L5][CO1]	**[5M]**
	b)	Why business are using big data for competitive advantage?	[L4][CO1]	**[5M]**
10	a)	How to implement IBM Big Data Strategy?	[L2][CO1]	**[5M]**
	b)	Generalize the list of tools related to Hadoop.	[L6][CO2]	**[5M]**

UNIT –II
HDFS(Hadoop Distributed File System)

1	a)	What is the Hadoop file system	[L1][CO1]	**[2M]**
	b)	Enumerate about Data flow	[L1][CO1]	**[2M]**
	c)	Define data injection in sqoop	[L1][CO2]	**[2M]**
	d)	Distinguish between horizontal file format and vertical File format	[L2][CO2]	**[2M]**
	e)	Construct the need for a distributed file system	[L3][CO3]	**[2M]**
2		What are the advantages of Hadoop? Explain Hadoop Architecture and its Components with proper diagram	[L3][CO2]	**[10M]**
3		Explain the block, name node and data node in Hadoop file system	[L2][CO3]	**[10M]**
4		Determine the basic commands in Hadoop command line interface.	[L3][CO5]	**[10M]**
5	a)	What is an interface? Establish the Hadoop system interfaces	[L3][CO2]	**[5M]**
	b)	Discuss about the Hadoop Archives and its Limitations	[L2][CO2]	**[5M]**
6		Describe the File read and File write operations in HDFS	[L1][CO5]	**[10M]**
7	a)	Discuss about the data ingest operation using sqoop and flume	[L2][CO2]	**[5M]**
	b)	Differentiate the compression and serialization operation in Hadoop I/O.	[L4][CO2]	**[5M]**
8		Elaborate the AVRO file format with a diagram	[L6][CO3]	**[10M]**
9	a)	What is data serialization?	[L3][CO3]	**[3M]**
	b)	Demonstrate the File Based Data structures.	[L2][CO2]	**[7M]**
10	a)	Analyze the features of Apache Hadoop .	[L4][CO6]	**[5M]**
	b)	How does Hadoop work?	[L2][CO2]	**[5M]**

UNIT –III
Map Reduce

1	a	What is Shuffling in MapReduce?.	[L1][CO1]	**[2M]**
	b	Define MapReduce.	[L1][CO1]	**[2M]**
	c	Compare Shufling and Sorting in MapReduce.	[L2][CO1]	**[2M]**
	d	List the parameters of mappers and reducers?	[L1][CO2]	**[2M]**
	e	Find the role of combiner and partitioner in map reduce application?	[L3][CO2]	**[2M]**
2	a)	Examine the Anatomy of a MapReduce Job Run.	[L4][CO4]	**[4M]**
	b)	Construct the Classic MapReduce Job Run with a neat diagram.	[L6][CO5]	**[6M]**
3		Estimate the Significance of YARN over Classic MapReduce Job Run.	[L5][CO3]	**[10M]**
4		What are the different types of failures in a)Classic MapReduce b)YARN	[L1][CO1]	**[10M]**
5	a)	Examine the different types of Job Scheduling process in Map Reduce.	[L3][CO4]	**[5M]**
	b)	Describe the Default MapReduce Job.	[L3][CO4]	**[5M]**
6		Describe the Shuffle and Sort operations in Map side and Reduce side	[L1][CO3]	**[10M]**
7	a)	What are the Properties in Task Execution Environment.	[L1][CO4]	**[5M]**
	b)	Discuss about Speculative Execution and its Properties.	[L2][CO4]	**[5M]**
8		Categorize the different types of input formats in MapReduce.	[L4][CO2]	**[10M]**
9		Examine the different types of output formats in MapReduce.	[L3][CO2]	**[10M]**
10		Contrast the below features in MapReduce. a) Counters b) Sorting c) Joins	[L4][CO3]	**[10M]**

UNIT –IV
Hadoop Eco System-Pig

1	a	Define Pig Latin.	[L1][CO1]	**[2M]**
	b	Illustrate and Give two examples of user defined functions.	[L2][CO2]	**[2M]**
	c	What is Grunt?	[L1][CO2]	**[2M]**
	d	Compare any two execution modes of pig.	[L2][CO5]	**[2M]**
	e	What are pig Latin relational operators	[L1][CO2]	**[2M]**
2	a)	Identify the features of PIG.	[L2][CO3]	**[2M]**
	b)	How to Install and execute PIG on Hadoop Cluster	[L2][CO5]	**[8M]**
3	a)	Compare the PIG with Databases with an Example	[L5][CO3]	**[5M]**
	b)	Evaluate the Expressions and types in Pig Latin.	[L4][CO4]	**[5M]**
4	a)	Why Do We Need Apache Pig?	[L4][CO2]	**[5M]**
	b)	Examine the different execution modes available in Pig	[L3][CO4]	**[5M]**
5		Construct User Define Functions in Pig Latin.	[L6][CO5]	**[10M]**
6	a)	Explain about Arithmetic Operators in Pig Latin .	[L2][CO3]	**[5M]**
	b)	Find the Grouping and Joining Data in Pig Latin.	[L3][CO3]	**[5M]**
7		Examine the Relational Operators in Pig Latin .	[L4][CO2]	**[10M]**
8		Develop the Schemas and Functions in Pig Latin	[L3][CO5]	**[10M]**
9	a)	Explain about the data types in Pig Latin.	[L2][CO2]	**[5M]**
	b)	Develop a program to calculate the maximum recorded temperature by year for the weather dataset in Pig Latin.	[L6][CO5]	**[5M]**
10	a)	Discriminate the Structures, Statements in Pig Latin	[L4][CO1]	**[5M]**
	b)	Evaluate Data Processing Operators in Pig Latin.	[L5][CO4]	**[5M]**

UNIT –V
Hive, Hbase, Big SQL

1	a	List out 5 hive shell commands.	[L1][CO1]	**[2M]**
	b	Sketch about Hbase.	[L3][CO2]	**[2M]**
	c	Examine Metadata	[L4][CO1]	**[2M]**
	d	Report about Big SQL	[L1][CO6]	**[2M]**
	e	Critique about the advantages of Hive query language	[L5][CO3]	**[2M]**
2		Discuss about Hive shell command line interface.	[L2][CO5]	**[10M]**
3	a)	Examine about Hive architecture with a neat diagram	[L3][CO2]	**[5M]**
	b)	Explain about components Hive architecture	[L2][CO2]	**[5M]**
4	a)	Deduce the various services offered by Hive.	[L4][CO4]	**[5M]**
	b)	Examine the Characteristics of HBase	[L4][CO1]	**[5M]**
5	a)	Infer the advantages of Hive over traditional databases?	[L2][CO5]	**[5M]**
	b)	what are the operators and functions in HIVE?	[L1][CO2]	**[5M]**
6	a)	Appraise about Hive query language?	[L4][CO5]	**[5M]**
	b)	Review Metastore in Hive?	[L2][CO5]	**[5M]**
7		Differentiate between of Hbase over RDBMS.	[L4][CO1]	**[10M]**
8		Explain with a neat diagram the architecture of Hbase.	[L2][CO2]	**[10M]**
9	a)	Categorize the joins in HiveQL	[L4][CO5]	**[5M]**
	b)	Report the Implementation of queries on sorting and aggregation of data in Hive	[L6][CO3]	**[5M]**
10	a)	Explain about IBM Big SQL?	[L2][CO6]	**[5M]**
	b)	Assess how HBase is implemented at Streamy.com	[L4][CO6]	**[5M]**

Prepared by:
G. Ravi Kumar