Course Code: 19MC9142 **R19**

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

QUESTION BANK (DESCRIPTIVE)

Subject with Code : Big Data Analytics (19MC9142) **Course & Branch**: MCA

Regulation: R19 **Year & Sem:** III-MCA & I-Sem

UNIT –I

INTRODUCTION TO BIG DATA

1	Discuss the following in detail. a) Conventional challenges in big data b) Nature of Data	[L6][CO1]	[12M]
2	Discuss about intelligent data analysis and nature of data.	[L6][CO1]	[12M]
3	Build the steps involved in support vector based inference methodology.	[L3][CO1]	[12M]
4	Define and explain Statistical Inference.	[L2][CO1]	[12M]
5	What are the different inferences in big data analytics.	[L1][CO1]	[12M]
6	Analyze about the bootstrapping and its importance.	[L4][CO1]	[12M]
7	What is sampling and sampling distribution give a detailed analysis.	[L1][CO1]	[12M]
8	Define and explain the following. a) Intelligent Data Analysis b) Analysis Vs Reporting.	[L5][CO1]	[12M]
9	Clearly explain the prediction error.	[L5][CO1]	[12M]
10	Discuss any five characteristics of Big Data.	[L6][CO1]	[12M]

Course Code: 19MC9142 **R19**

UNIT –II

INTRODUCTION TO STREAM CONCEPTS

1	a) What is a data stream? b) Discuss 14 insights of Info sphere in data stream.	[L1][CO2] [L6][CO2]	[02M] [10M]
2	Explain the different applications of data streams in detail.	[L2][CO2]	[12M]
3	Clearly Explain the stream model and architecture.	[L5][CO2]	[06M]
4	Identify how to count ones in a window using DGIM algorithm.	[L3][CO2]	[12M]
5	Discuss the following. a) Counting distinct elements in a stream. b) Finding most popular elements using decaying window.	[L6][CO2]	[12M]
6	What are filters in Big Data? Explain Bloom Filter with example	[L1][CO2]	[12M]
7	Define Decaying window and explain how its performed in data analytics.	[L2][CO2]	[12M]
8	Explain the following. a) FM algorithm and its application b) AMS algorithm and its applications	[L5][CO2]	[06M]
9	What is Real Time Analytics? Discuss about RTAP applications.	[L1][CO2]	[12M]
10	Analyze the three categories of Prediction methodologies.	[L4][CO2]	[12M]

Course Code: 19MC9142 **R19**

UNIT-III

HISTORY OF HADOOP

1	a) What is Hadoop? Explain its components. b) How do you analyze the data in hadoop.	[L1][CO3] [L1][CO3]	[05M] [07M]
2	Explain the following a) Mapper class b) Reducer class c) Scaling out	[L2][CO3]	[12M]
3	Listout and explain the failures in Mapreduce.	[L4][CO3]	[12M]
4	Explain the map reduce data flow with single reduce and multiple reduce.	[L5][CO3]	[12M]
5	How Hadoop streaming is suited with text processing explain.	[L1][CO3]	[12M]
6	Define HDFS. Discuss namenode, datanode and block. Explain HDFS operations in detail.	[L6][CO3]	[12M]
7	Analyze the concept of developing the Map Reduce Application.	[L4][CO3]	[12M]
8	How map reduce job works with classic java stream.	[L3][CO3]	[12M]
9	Clearly Explain how map reduce jobs run on YARN.	[L5][CO3]	[12M]
10	Discuss the various types of map reduce & its formats.	[L6][CO3]	[12M]

Course Code: 19MC9142 **R19**

UNIT-IV

SETTING UP HADOOP CLUSTER

1	What is Cluster? Explain the setting up a Hadoop cluster	[L1][CO4]	[12M]
2	a) What are the different types of Hadoop configuration files? Discuss.	[L1][CO4]	[06M]
	b) What are control scripts? Explain the start.dfs.sh script, Start.mapred.sh.	[L1][CO4]	[06M]
3	What are the additional configuration properties to set for HDFS	[L6][CO4]	[12M]
4	Explain three step Kerberos ticket exchange protocol	[L2][CO4]	[12M]
5	Define is benchmarking how it works in Hadoop.	[L1][CO4]	[12M]
6	a) How will you define commissioning new nodes and decommissioning old nodes?	[L1][CO4]	[06M]
	b) Build the steps for upgrading HDFS. .	[L3][CO4]	[06M]
7	Discuss administering Hadoop with its checking point process diagram	[L6][CO4]	[12M]
8	a) How to run proxy & Running map reduce job.	[L1][CO4]	[06M]
	b) Explain Data node directory structure	[L5][CO4]	[06M]
9	List out and explain Important Hadoop daemon properties?	[L4][CO4]	[12M]
10	How does security is done in Hadoop. Justify.	[L5][CO4]	[12M]

Course Code: 19MC9142 **R19**

UNIT-V

APPLICATIONS ON BIG DATA

1	a) What is PIG ? Explain its installing process. b) Explain two execution types or modes in PIG.	[L1][CO5] [L2][CO5]	[06M] [06M]
2	Explain Grouping, Join, CoGroup, Cross & Group in data.	[L2][CO5]	[12M]
3	Clearly explain the process of installing HIVE & features of HIVE.	[L5][CO5]	[12M]
4	Identify How will you query the data in HIVE?	[L3][CO5]	[12M]
5	Briefly discuss about HBASE.	[L6][CO5]	[12M]
6	What is Zookeeper? Explain its features with applications.	[L4][CO5]	[12M]
7	Explain in detail IBM infosphere Big insights and Streams.	[L2][CO5]	[12M]
8	Discuss the visual data analysis techniques in detail.	[L6][CO5]	[12M]
9	Briefly explain Interaction techniques with its applications.	[L2][CO5]	[12M]
10	What is HiveQL? Explain its features.	[L1][CO5]	[12M]

Prepared by Ms. P. SUKANYA, Asst. Professor, Department of MCA