SIDDHARTH INSTITUTE OF ENGINEERING & TECHNOLOGY :: PUTTUR

Siddharth Nagar, Narayanavanam Road – 517583

MODEL QUESTION BANK (DESCRIPTIVE)

Subject with Code :BIG DATA ANALYTICS(18MC9120)   **Course & Branch**: MCA

Year & Sem:II-MCA& II-Sem   **Regulation:** R18

UNIT –I

INTRODUCTION TO BIG DATA

1. Discuss the following in detail.
   a) Conventional challenges in big data   6M
   b) Nature of Data   6M
2. Describe the steps involved in support vector based inference methodology. 12M
3. Write a short note on Statistical Inference.   12M
4. Define and explain the different inferences in big data analytics.   12M
5. Describe the bootstrapping and its importance.   12M
6. What is sampling and sampling distribution give a detailed analysis.   12M
7. Define and explain the following.
   a) Intelligent Data Analytics   6M
   b) Analysis Vs Reporting.   6M
8. Describe the prediction error and regression techniques.   12M
9. Describe any five characteristics of Big Data.   12M
10. Define Arcing classifier , Bagging predictors and explain in detail.   12M

.

UNIT –II

INTRODUCTION TO STREAM CONCEPTS

1. a) What is a data stream?   02M

    b) Discuss 14 insights of Info sphere in data stream.   10M

2. Explain the different applications of data streams in detail.   12M

3. Explain the stream model and Data stream management system architecture.   12M

4. Explain how to count ones in a window using DGIM algorithm   12M

5. Write a short note on the following.
    a) Counting distinct elements in a stream.   06M
    b) Finding most popular elements using decaying window.   06M
6. What are filters in Big Data? Explain Bloom Filter with example   12M

7. Define Decaying window and explain how its performed in data analytics.   12M

8. Explain the following.

    a) FM algorithm and its application   06M

    b) AMS algorthim and its applications   06M

9. What is Real Time Analytics? Discuss their technologies in detail   12M

10. Explain the three categories of Prediction methodologies.   12M

UNIT-III

HISTORY OF HADOOP

1. a) What is Hadoop? Explain its components.   05M

    b) How do you analyze the data in hadoop.   07M

2. Explain the following

    a) Mapper class   05M

    b) Reducer class   05M

    c) Scaling out   02M

3. Explain the failures in Mapreduce.   12M

4. Explain the map reduce data flow with single reduce and multiple reduce. 12M

5. How Hadoop streaming is suited with text processing explain.   12M

6. Define HDFS. Describe namenode, datanode and block. Explain HDFS operations

    in detail.   12M

7. Write in detail the concept of developing the Map Reduce Application.   12M

8. How map reduce job works with classic java stream.   12M

9. Explain how map reduce jobs run on YARN.   12M

10. Discuss the various types of map reduce & its formats.   12M

UNIT-IV

SETTING UP HADOOP CLUSTER

1. What is Cluster? Explain the setting up a Hadoop cluster   12M

2. a) What are the different types of Hadoop configuration files? Discuss.   06M

    b) What are control scripts? Explain the start.dfs.sh script, Start.mapred.sh.   06M

3. What are the additional configuration properties to set for HDFS   12M
4. Explain three step Kerberos ticket exchange protocol   12M
5. What is benchmarking how it works in Hadoop.   12M

6. a) How will you define commissioning new nodes and decommissioning old nodes?   08M
    b) Write the steps for upgrading HDFS.   .   04M
7. Discuss administering Hadoop with its checking point process diagram   12M

8. a) How to run proxy & Running map reduce job.   08M

    b) Explain Data node directory structure   02M

    c) Set Log levels & its metrics   02M

9. What are the Important Hadoop daemon properties?   12M

10. How does security is done in Hadoop.Justify.   12M

UNIT-V
APPLICATIONS ON BIG DATA

1. a)What is PIG ? Explain its installing process. 05M

  b)Explain two execution types or modes in PIG. 07M

2. Explain Grouping, Join, CoGroup, Cross & Group in data. 12M

3. Explain the process of installing HIVE & features of HIVE. 12M

4. How will you query the data in HIVE? 12M

5. Give a detail note on HBASE. 12M

6. What is Zookeeper? Explain its features with applications. 12M

7. Explain in detail IBM infosphere Big insights and Streams. 12M

8. Discuss the visual data analysis techniques in detail. 12M

9. Give a detail note on Interaction techniques with its applications. 12M

10. What is HiveQL? Explain its features. 12M

Prepared by: Ms. P.Sukanya, Assistant Professor, Dept. of MCA, SIETK , Puttur