

SYNOPTIC

Big Data Analytics  
Management  
8/12/2018

SET 01

Q 1 A) Advantages of Big data Analytics as compared to traditional analytical model and Advantages of Hadoop in particular over RDBMS [10 M]

Ans : All written as given 1 mark each → Scalability, no-preprocessing like RDBMS, Unstructured data, supports different data types, data formats. , protection against failure, Big data, simple Interface to carry out Analysis,

Q 1 B) Illustrate Hadoop Architecture. Give functionality of each component ie. NN, DN, JT, TT [10 M]

Ans :

Correct diagram 2 M + all above mentioned blocks correctly explained 2 M each

Q2 A) Write a Map Reduce program for word count of large data. Give the output at each stage [10M]

Ans : Correct Mapper code 5 M + correct reducer code 5 M

OR

Q 2A ) MR program of transaction [10M]

Ans : Correct Mapper code 5 M + correct reducer code 5 M

Q2 B) Explain natural join by map reduce with pseudo code and example [10M]

Ans : natural join explanation 2 M + for big data 2 M + example 2 M + pseudocode fully correct 4 M

OR

Q 2 B) Explain grouping and aggregation by by map reduce with pseudo code and example

Ans : Grouping and aggregation 2 M + for big data 2 M + example 2 M + pseudocode fully correct 4 M

Q3 A) Describe stream clustering algorithm. Explain initialization, merging buckets and answering queries. [10 M]

Introduction 1 M + Initializing Buckets [2M] + Merging buckets [2M] + Answering Queries [2M]

Q3 B) what is sentiment analysis? Explain in brief. [10 M]

Ans : Overview 4M + Sentiment Analysis in big data 4 M + retrieving , storing data in HDFS 2 M

Q 4 A) Give the analysis of Flajolet-Martin algorithm. [10 M]

Ans : Explanation 3 M + Algorithm 4 M + example 3 M

Q 4 B) Describe data stream model in details with examples. [10 M]

Ans : Explanation 6 M + examples 4 M

OR

Give different techniques to combat link spam [10 M]

Ans : Introduction Link Spam Combating technique 2 M + TrustRank 4M + Spam Mass 4 M

Q 5 A) what are the different types of load scenarios in Hadoop. Explain with examples.

Ans : Load scenarios (Data at rest + Data in Motion + Streaming Data + Data from a web server + Data from a data warehouse ) to be written for 10 M

B) Describe Sqoop along with its working? Give its import command. [10 M]

Ans : Sqoop working and overview 6 M + Import command 4 M

OR

Explain working of Flume with the help of diagrams. [10 M]

Ans : Overview 2 M + working 4 M + diagram 4 M