Big Data Analytics Question Bank:

UNIT I :

1.Explain V's of big data analytics.
2.Write short note on drivers of Big Data.
3.Explain different applications of big data.
4.State and explain data privacy protection with neat diagram default the product knowledge hubing bigdata.
5.Write a note on location based services on big data.
6.explain real time adaptive analysis and decision engine.
7.Explain in detail MPP platform.
8.Explain Unstructured data analysis and reporting.

UNIT II:

1.Explain feature of R language.
2.Explain different phrases of mapreduce with example.
3.What is HDFS and explain Mapreduce Architecture.
4.List and explain different components of Hadoop.
5.Explain in detail the data flow of mapreduce with diagram.
6.Explain the limitations of mapreduce.
7.Explain data mining technique which are used to perform data modelling in R.
8.Mention different Hadoop installation mode.

UNIT III :

1.Explain architecture of RHIPE.
2. Explain R Hadoop in detail.
3. Explain architecture of R Hadoop.
4. Explain working of R-Hadoop withexample.
5.Explain the hstable reader function in Hadoop streaming.
6.Explain the reader function for Hadoop streaming.
7.Explain the Hadoop streaming compnents.
8.Explain the format of Hadoop streaming commands with each line?.

UNIT IV :

1.Explain Data Analytics project life cycle stages.
2.Explain how data analysis problem for calculating the frequency of stock market change can be solved using Mapreduce.
3.Write case study predicting the action sales price of heavy equipment to create blue book for bulldozers.
4.Explain Poisson – Approximation resampling technique on the map of the Map Reduce task.
5.How will data analysis help to identify the category of a web page of website, which may categorize popularity wise as High, Medium,or Low(Regular),based on the visit count of the pages.
6.Write steps to build and run map reduce alogorithm with R and Hadoop integration for web page categorization problem.
7.Explain pre-processing and performing analysis over any data.

8.Explain how mapreduce proble is designed for computing the frequency of stock market changes.

UNIT V :

1.What is Resilient Distributed Dataset (RDD)? Explain transformations and actions in RDD. Explain RDD operations in brief?
2.Why Spark is preferred over Hadoop? Explain the limitations of Hadoop?
3.Explain how Spark overcomes the limitations of Hadoop.
4.Briefly explain the core components in Spark.
5.Explain the architecture of Spark.
6.What is Spark Context in Apache Spark?
7.What is a Directed acyclic graphs (DAG) in Spark, and how does it work?
8.What are Spark Data Frames? Why do we use them in Spark?
9.Explain Apache Spark RDD Operations in detail.
10.What are different types of RDD transformation? Explain functions in RDD transformation.
11.What are RDD actions? When they are used? Explain Spark actions.
12.What are the deployment modes in Spark? What is difference between client and cluster mode deployment?
13.What are the components of Spark architecture?
14.What is Spark core? What are the various functions of Spark core? Which is a component on the top of Spark core?
15.What are the components of Spark Streaming? What is Spark Streaming used for?

UNIT VI :

1. What is machine learning? Explain types of machine-learning algorithms.
2. Explain Supervised Machine Learning Algorithm.
3. Explain how Linear regression is performed using with R and Hadoop?
4. Explain how logistic regression is performed using with R and Hadoop?
Explain Unsupervised Machine Learning Algorithm.
6. Explain steps to performing clustering with R and Hadoop.
7. Explain Steps to generate recommendations in R.
8. What is recommendation algorithm? Explain two different types of Recommendations Algorithms.
9. How do you create a recommendation algorithm with R and Hadoop?
10. How one can use R and Hadoop together to generate recommendations from big datasets?