

BDA Question Bank

1. Define Big Data Analytics
2. Explain with a neat diagram Big Data Architecture design layer
3. Explain MapReduce Programming Model along with example(Example 4.1 and 4.2 from notes)
4. Explain MapReduce Process for client submitting a job
5. Explain MapReduce steps with neat diagram and apply the same on an example.
6. How do the task trackers cope up with node failure?
7. Illustrate application of mapreduce on Matrix application with steps
8. Illustrate features of Hive with advantages and limitation
9. What are the steps of Hive installation.
10. Difference between Hive database and RDBMS.
11. Explain Hive data types and file formats
12. Discuss HiveQL DDL with example.
13. Discuss partitioning of table.
14. Write short note on bucketing
15. What are the features of apache pig. Compare Apache Pig and Mapreduce and Apache Pig and Hive.
16. Discuss Pig Architecture.
17. Discuss Pig Latin Data Model
18. Write short note on Artificial intelligence
19. Explain the following i. Linear and non-linear ii. Outlier iii. Variance iv.linear and non-linear
20. Explain Probabilistic Distribution with Example(6.2).
21. Short note on ANOVA test
22. Discuss linear regression
23. How to do prediction using Regression Analysis
24. Discuss KNN
25. Different techniques for finding similarities.
26. Explain Market Basket Model
27. Define text mining. Explain text mining process phases with example.
28. Short note on SVM
29. Discuss web usage mining
30. How is the discriminative classifier used?
31. Short note on Hub and Authorities
32. Explain the following terms
i. Cluster Computing ii. Resource Management iii. Ecosystem
33. Differentiate MapReduce v1 and MapReduce v2.
34. With neat Diagram, explain Hadoop main components and ecosystem components.
35. List and explain features of Hadoop
36. Write a short note on

a. a) Zookeeper

b) Oozie

37. Discuss the framework for distribution of job based on client application
38. With neat diagram, explain various system roles in Hadoop deployment
39. List & Explain core components of Hadoop.
40. What are the key benefits of HDFS Namenode Federation
41. How do Job-Tracker and Tasktracker function?
42. Explain the four important components of Apache Flume.
43. With neat diagram, explain Hadoop 2 Execution Model.
44. Explain the features of distributed computing architecture.
45. Discuss HDFS safe mode
46. Explain the features of distributed computing architecture.
47. Explain Brewers CAP theorem
48. How to increase flexibility for data manipulation
49. Compare NoSQL databases with SQL databases
50. Explain Trigger, View and Schedules in SQL databases
51. With neat diagram, explain CAP theorem in Big Data Solution
52. Illustrate Not Only SQL with example
53. What are the Characteristics of Graph DataBases
54. What is MangoDB?
55. Design document structure of a blog site for Mango BD.
56. How does an HTML object and XML based web service relate with tabular data stores.
57. What are the advantages of Key-value Store
58. Write short note on RC file format
59. Explain different XML document Pattern
60. Discuss Big Data table store