Bigdata Assignment

Handwritten Assignment: All are required to submit the handwritten solution of one question from the list attached.

*Every student has to solve one question. There are 60 questions. You have to take mod of your roll no with 60 and solve the question accordingly. ****___Your answer should be of minimum 5 pages (A4 size)___ ****. You have to submit the hard copy to the undersigned during your offline MST to be scheduled from 6th November 2023.*

***Only in pages and in good handwriting with neat diagrams (if any) ***

1. Why Big Data? Illustrate the necessity of Big Data analysis with the help of suitable case studies.
2. Explain the evolution of big data as a problem. What are the drivers responsible for generation of big data? Explain them in detail.
3. A bookmarking site permits to bookmark, review, and rate and search various links on any topic. Analyze social bookmarking sites like stumbleupon.com or reddit.com to find insights.
4. Analyze the tourism data and identify the ones which are difficult in storing in RDBMS and the need for Big Data technologies for storing and analyzing them.
5. Why linked lists, stacks and queues are called as linear data structures and explain the operations performed on stacks and queues with examples.
6. What is the use of generic methods and classes supported by java?
7. Explain the basic building blocks of Hadoop and neat sketch.
8. Explain the various operational nodes of Hadoop cluster configuration.
9. Distinguish between the old and new versions of Hadoop API and Map Reduce frame work.
10. What is Map Reduce? Why it is required? How the algorithm will address the problem of big data?
11. Illustrate significance of Writable interface along with writable comparable and comparators w.r.to implementing and serialization.
12. sketch the architecture of pig and illustrate it in detail.
13. What are the data types supported by HIVE querry language. Explain with example.
14. What is the use of SerDe's (Serializer & Deserializer) in HIVE and explain various types of SerDes'ssupported by HIVE.
15. Discuss the architecture of HIVE with a neat sketch.
16. How Big Data Analytics can be useful in the development of smart cities.
17. What is Big data? Discuss it in terms of volume and velocity. What are the advantages of Hadoop? Explain the Hadoop Components with proper diagram.

18. Explain Storage mechanism in HBase. In Map Reduce how Job Scheduling is done in case of the Fair Scheduler?
19. Draw and explain HDFS Architecture. Explain the function of NameNode and DataNode. What is a Secondary Namenode? Is it a substitute to the Namenode?
20. Explain following commands of HDFS with syntax and at least one with example of each.
    (i) get (ii) cp (iii) chown.
21. Explain Metastore in Hive.
22. Explain the concept of Blocks and Heartbeat Message in HDFS Architecture. What are the benefits of block transfer?
23. What is Zookeeper? List the benefits of it. Explain the 5 P's of Data science in brief.
24. Explain Spark components in detail. Also list the features of spark.
25. Explain following for MongoDB. (i) Indexing (ii) Aggregation.
26. Explain how HBase uses Zookeeper to Build Applications with Zookeeper.
27. What is link analysis? Illustrate page rank algorithm with suitable example.
28. Explain working of following phases of Map Reduce with one common example. (i) Map Phase (ii) Combiner Phase (iii) Shuffle and Sort Phase (iv) Reducer Phase
29. Discuss YARN (Yet Another Resource Negotiator).
30. Explain Hadoop ecosystem.
31. Discuss Hadoop limitations in detail with suitable examples.
32. Describe: a. NameNode b. DataNode c. Job Tracker d. Task Tracker.
33. How traditional file system differs from hadoop distributed file system.
34. Explain: a. Amazon DynamoDB b. MongoDB c. Neo4j d.Google's Big Table
35. What are NoSQL data architectural patterns? Explain them in detail.
36. Explain CAP. How is CAP different from ACID property in databases.
37. Explain why NoSQL is scheme-less.
38. State an example and explain shared and shared nothing architecture.
39. Explain the four ways by which big data problems are handled by NoSQL.
40. Discuss Hyperlink-Induced Topic Search Algorithm and its concept.
41. Explain topic sensitive PageRank.
42. Illustrate the PageRank computation.
43. Explain social network analysis. Introduce the various applications of social network mining.
44. Write a python program for implementing fully connected social network of 100 nodes using python or R.
45. Discuss the different types of social networks.
46. What is clustering? Explain with suitable example. Discuss the methods used for clustering of social graphs for identification of communities.
47. What is recommender system? Illustrate how recommendation systems used to recommend friends in online social network Facebook.

48. Explain Partitioners and combiners? Why they are so important? Throw some light on resource management process done by Node manager and Resource manager?
49. Solve the word count problem with suitable diagram and write process for the same in details: para is given for the same find words: NOSQL, the, is, data

The Oracle NoSQL Database is a distributed key-value database. It is designed to provide highly reliable, scalable and available data storage across a configurable set of systems that function as storage nodes. Data is stored as key-value pairs, which are written to particular storage node(s), based on the hashed value of the primary key. Storage nodes are replicated to ensure high availability, rapid failover in the event of a node failure and optimal load balancing of queries. Customer applications are written using an easy-to-use Java/C API to read and write data. NoSQL databases are becoming a major part of the database landscape today, and with their handful of advantages, they can be a real game changer in the enterprise arena. However, NoSQL isn't ripe yet, and professionals in the industry need to approach it with caution. This is because it lacks the maturity that SQL databases like MySQL offer. If your application doesn't fall into the category of the likes of Google, Yahoo, Facebook or Wikipedia, you should reconsider your options for using NoSQL and stick with MySQL instead. Not only is there a major skills gap with finding NoSQL professionals, but issues like analytics, performance reporting and migration also need to be considered

50. Write Query statement of the following in MongoDB CREATE, UPDATE, INSERT, DELETE, DROP DB, DROP TABLE, and INSERTMANY; also explain how they are different from SQL query statements.
51. Explain the following: Homophiliy, Transitivity, Bridges, and Cliques with examples.
52. Differentiate the following; Betweenness Centrality, Closeness Centrality and Eigenvector Centrality.
53. What are small worlds? Explain Average and longest distance with suitable example.
54. What are the reasons for preferential Attachment? How Ego networks are identified?
55. What are recommender systems? Why we should use them? What are the advantages of recommender systems over traditional ones?
56. Explain with suitable example User-Based Collaborative Filtering? How it is different from Content Based Filtering?
57. What are Hybrid Recommendation systems? Why they are needed?
58. Explain Clustering with suitable diagram and example?
59. Explain Reciprocity with suitable diagram and example?
60. How to identify sets of key players? How edge weights related to relationship strength?