BDA

Module 1:

1. Compare traditional databases & big data system [5M] [2x]
2. How big data problems are handled by Hadoop system [5M] [3x]
3. Mention four characteristic of big data. Elaborate these charateristic with respect to social media [5M] [4x]
4. What is Hadoop and Why it Matters [5M]
5. Draw the Hadoop ecosystem and briefly explain it's component [10M] [3x] [4x]

Module 2:

1. Explain the arch. of HDFS [10 Marks]
2. Write a Map Reduce pseudo code for Matrix vector multiplication [10M] [3x]
3. Write a Map Reduce pseudo code for a word count problem [10M] [2x]
4. Explain relational Algebra Operation in MapReduce [10M]
5. What is function of Map Tasks in MapReduce framework? Explain with help of example [5M]

6. List the main Component of Map-Reduce execution Pipeline [10M] [2x] [3x]

7. Explain How Hadoop goals are covered in HDFS [10M]

Module 3:

1. What are the core business drivers behind the NoSQL movement? [5M]

2. Explain NoSQL arch. pattern such as key-value stores & graph stores [10M]

3. Discuss NoSQL's solution for handling big data [5M]

4. Explain the variation of NoSQL arch pattern, focussing on column family stores [10M]

5. Compare master-slave & peer-to-peer models in NoSQL system [5M]

6. Explain Column family store & Graph Store NoSQL arch pattern with example [10M]

7. Explain the four type of NoSQL database [10]

8. List & explain the core business drivers behind the NoSQL movement [5M] 3x

Pass 2:

① After every merge, new pts. merged are considered as representative for new cluster.

② Finally, cluster merging will stop when target k (no. of clusters) is achieved.

~~M-3~~ ~~M-4~~

Q) ☆ Explain DGIM algo (10)

Q) Issues in Data stream query processing (10) & (5)

Q) ☆ Flajolet Master numerical (10) ~~[Comps May'24 & 23]~~

Q) Explain DGIM algo. for counting ones in a stream with example. (10).

Q) ☆ Girvan – Newman → determine communities (10) ~~[Comps May 24 & 23 & 22]~~

Q) List down all 6 constraints that must be satisfied for representing a stream by buckets using DGIM algo with examples (10) (5).

Q) ~~What~~ Bloom Filter for stream data mining (5)

OR

Explain Bloom Filter with example (5)

Q) With neat sketch explain architecture of DSMS (10).

M-5 –

Q) ⭐ Ex Explain cure algo (10) (AI DS)

Q) Explain with eg. 2 major classes of distance measures (10). (AI DS)

Q) Short note: Dist. measures for Big data (5)

M-6 –

Q) What is Recommender System? Explain its types (10).

Q) What is Social NLW? Give varieties of Social NWs & need for social NLW graph. (10).

Q) Explain structure of web with suitable diagram (10)

Q) Explain Collaborative Filtering system. How is it ~~as~~ diff. from content based system (10).

Q) What is clique percolation method. Write an algo on CPM. Also show how CPM finds clique from following graph. Explain with steps (10).

Q) ⭐ Explain ~~Page~~ PageRank algo (10).

Q) ⭐ Girvan Newman Numericals (10)

Q) Describe collaborative filtering in recommendation system (10).

Q) Define collaborative filtering. Using an example of an e-commerce site like flipkart, or amazon describe how it can be used to provide recommendation to users. (10).

Q) How recommendation is done based on properties of the product? Explain with the help of an ~~eee~~ example (10).