Big Data Analytics - Final Important Questions (Harsh Gupta)

BIG DATA ANALYTICS (BDA) - Most Important Questions (Module-wise)

MODULE 1 - Introduction to Big Data & Hadoop

1. Explain 5V's of Big Data.
2. Compare Traditional Data and Big Data.
3. Draw and explain Hadoop Ecosystem.
4. Explain Architecture of Big Data.
5. What is Hadoop and Why it Matters?

MODULE 2 - HDFS & MapReduce

1. Explain HDFS Architecture.
2. Explain MapReduce Components and Execution Steps.
3. Write a MapReduce Pseudocode for Matrix Multiplication with example.
4. List Main Components of MapReduce Execution Pipeline.
5. Explain Advantages and Limitations of Hadoop.

MODULE 3 - NoSQL Databases

1. Explain CAP Theorem. How is it different from ACID?
2. Explain Four Types of NoSQL Databases with examples.
3. Differentiate between SQL and NoSQL.
4. Explain Column Family Store & Graph Store.

MODULE 4 - Mining Data Streams

1. Explain Issues in Data Stream Query Processing.
2. Explain Bloom Filter with example.
3. Explain DGIM Algorithm with example.
4. Explain Flajolet-Martin Algorithm.
5. Explain Exponentially Decaying Windows.

MODULE 5 - Finding Similar Items & Clustering

1. Explain all Distance Measures with example.
2. Explain CURE Algorithm.
3. Explain BDMO Approach.
4. Explain Two Major Classes of Distance Measures.

MODULE 6 - Real-Time Big Data Models

1. Explain PageRank Algorithm with example.
2. Explain Recommender System and its Types.
3. Explain Collaborative Filtering vs Content-based System.
4. What is a Social Network? Give types and need for Social Graph.
5. Explain Clique Percolation Method (CPM) with example.

TOP 10 QUESTIONS TO REVISE BEFORE EXAM

---

1. 5Vs of Big Data
2. HDFS Architecture
3. MapReduce Execution Steps
4. Matrix Multiplication using MapReduce
5. CAP Theorem vs ACID
6. Bloom Filter with Example
7. DGIM Algorithm
8. Distance Measures (any three)
9. PageRank Algorithm with Example
10. Recommender System (Types + Collaborative vs Content-based)