BDAV Practical Viva – Full Questions & Answers

HDFS – Viva Questions & Answers

1. What is HDFS?
HDFS is Hadoop Distributed File System used to store very large files across machines.

2. What is NameNode?
NameNode stores metadata. It is the master.

3. What is DataNode?
DataNode stores actual data in blocks.

4. What is block size?
Default block size is 128 MB.

5. What is replication factor?
Default replication is 3.

6. Command to see files:
hdfs dfs -ls /user/cloudera

7. Upload file:
hdfs dfs -put file.txt /user/cloudera

8. Remove file:
hdfs dfs -rm file.txt

9. Why HDFS is fault tolerant?
Because of replication.

10. Difference between HDFS and local FS?
HDFS is distributed & fault tolerant.

Hadoop / YARN Viva Answers

1. YARN = Yet Another Resource Negotiator.

2. ResourceManager allocates resources.

3. NodeManager runs tasks.

4. Container = environment where tasks run.

5. MapReduce has Map & Reduce phases.

6. Mapper outputs key-value pairs.

7. Reducer aggregates results.

8. Input Split = piece of data for mapper.

9. Hadoop 2 uses YARN.

10. Hadoop is used for Big Data processing.

Hive Viva Answers

1. Hive is a data warehouse tool.

2. Used for SQL-like querying.

3. Create DB: CREATE DATABASE db;

4. Create table:
CREATE TABLE emp(id INT, name STRING, salary INT)
ROW FORMAT DELIMITED FIELDS TERMINATED BY ','

5. Managed table deletes data.

6. External table keeps data safe.

7. Load data:
LOAD DATA LOCAL INPATH 'emp.csv' INTO TABLE emp;

8. Fields terminated by ',' means CSV format.

9. Show tables: SHOW TABLES;

10. Partitioning = splitting table for fast search.

Pig Viva Answers

1. Pig is a scripting tool.

2. Grunt = Pig shell.

3. LOAD loads data.

4. DUMP shows output.

5. Counters appear only when DUMP executes a job.

6. Relation = table.

7. Join example:

j = join users by name, clicks by name;

8. Bag = collection of tuples.

9. PigStorage(',') reads CSV.

10. DUMP shows results; STORE saves to HDFS.

Sqoop Viva Answers

1. Sqoop transfers data between MySQL and Hadoop.

2. Used for import/export.

3. Import:
sqoop import --connect jdbc:mysql://localhost/db --table emp --username root --password cloudera

4. Export:
sqoop export --connect jdbc:mysql://localhost/db --table emp --export-dir /emp/out

5. Incremental import = only new rows.

6. Default mappers = 4.

7. --target-dir = output folder.

8. import-all-tables = imports whole DB.

9. JDBC string = MySQL connection.

10. Sqoop metastore stores job info.

General Big Data Viva Answers

1. Big Data = large, fast, varied data.

2. 5Vs: Volume, Velocity, Variety, Veracity, Value.

3. Distributed system = many machines working together.

4. SQL vs NoSQL: tables vs flexible.

5. Hadoop = cheap, scalable, fast Big Data system.