HDFS Commands, Operations, User Management & SSH – Detailed Answers

16. Commands to Insert and Retrieve Data into HDFS

HDFS provides command-line utilities to store (insert) and retrieve data from the distributed file system.

Insert Data into HDFS:
hdfs dfs -put localfile /hdfs/path
This command uploads a file from the local file system into HDFS. If the file already exists, it will throw an error unless overwritten using -f option.

Retrieve Data from HDFS:
hdfs dfs -get /hdfs/path localpath
This command downloads files from HDFS to the local file system.

17. HDFS Read and Write Operations

HDFS read and write operations are designed to handle large files efficiently in a distributed environment.

Write Operation: When a client writes data to HDFS, it first contacts the NameNode to get metadata and block locations. The data is then written to DataNodes in a pipeline fashion and replicated as per configuration.

Read Operation: During read operation, the client requests block locations from the NameNode and reads data directly from the nearest DataNode.

18. Steps for Adding User and SSH Access

Adding a user and enabling SSH access is essential for managing Hadoop clusters securely.

- Create a new user account.
- Set password for the user.
- Generate SSH key pair using ssh-keygen.
- Copy public key to authorized_keys file.
- Set correct permissions for .ssh directory.
- Test password-less SSH login.

19. Command to Create User Account

Linux provides commands to create and manage user accounts.

useradd username
This command creates a new user account. After creating the user, set a password using the passwd command.

passwd username
This command sets or changes the password for the user account.

20. HDFS Commands

HDFS supports several commands to manage files and directories.

- hdfs dfs -ls – Lists files and directories in HDFS.
- hdfs dfs -mkdir – Creates directories in HDFS.
- hdfs dfs -rm – Deletes files from HDFS.
- hdfs dfs -du – Displays disk usage.
- hdfs dfs -df – Shows free space information.
- hdfs dfs -cat – Displays file content.
- hdfs dfs -chmod – Changes file permissions.
- hdfs dfs -chown – Changes ownership of files.