

Postlab:-

1. What are the main components of a Hadoop application?

HDFS (Hadoop Distributed File System):

Stores large files across multiple machines with fault tolerance using replication.

YARN (Yet Another Resource Negotiator):

Manages cluster resources and job scheduling.

MapReduce:

A programming model used for distributed data processing (map = split, reduce = aggregate).

Hadoop Common:

Provides essential Java libraries and utilities used by other modules.

2. Difference between NameNode, Backup Node, and Checkpoint Node:

Component	Function	Real-Time Sync	Failure Recovery Role
NameNode	Manages file system metadata like file names, directories, and block locations.	Yes	Acts as the master; essential for HDFS operation.
Backup Node	Maintains an in-memory, up-to-date copy of metadata from the NameNode.	Yes	Can immediately take over if NameNode fails.
Checkpoint Node	Periodically downloads and merges fsimage and edits, then sends a new fsimage to NameNode.	No	Reduces NameNode startup time, not used for failover.

3. Explain the use of cat, du, du -s:

- `cat` (concatenate):
Used to view the contents of files in the terminal. Example: `cat file.txt`
- `du` (disk usage):
Shows the space used by files and directories. Example: `du myfolder/`
- `du -s` (summary):
Displays the total size of a folder, instead of listing all subdirectories. Example: `du -s myfolder/`