

Devanshu Surana

1032210755, Roll no : 50

Panel C, Batch BDT 1-B2

BDT Lab Assignment 4

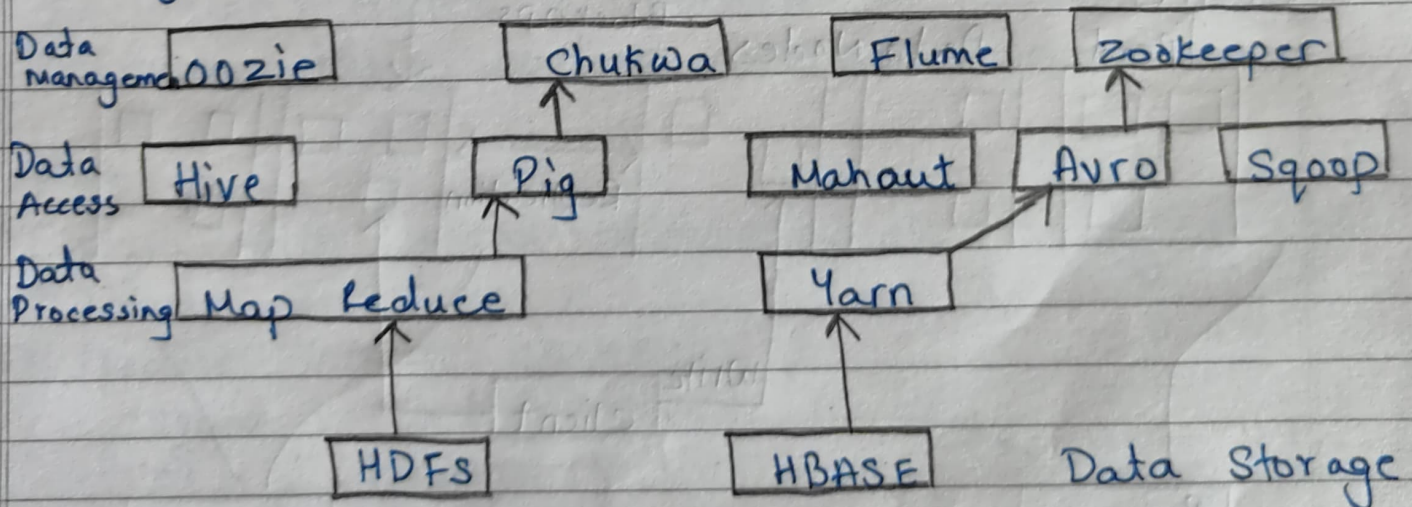
Problem statement:

Install Hadoop and perform basic hadoop commands on it.

Objectives:

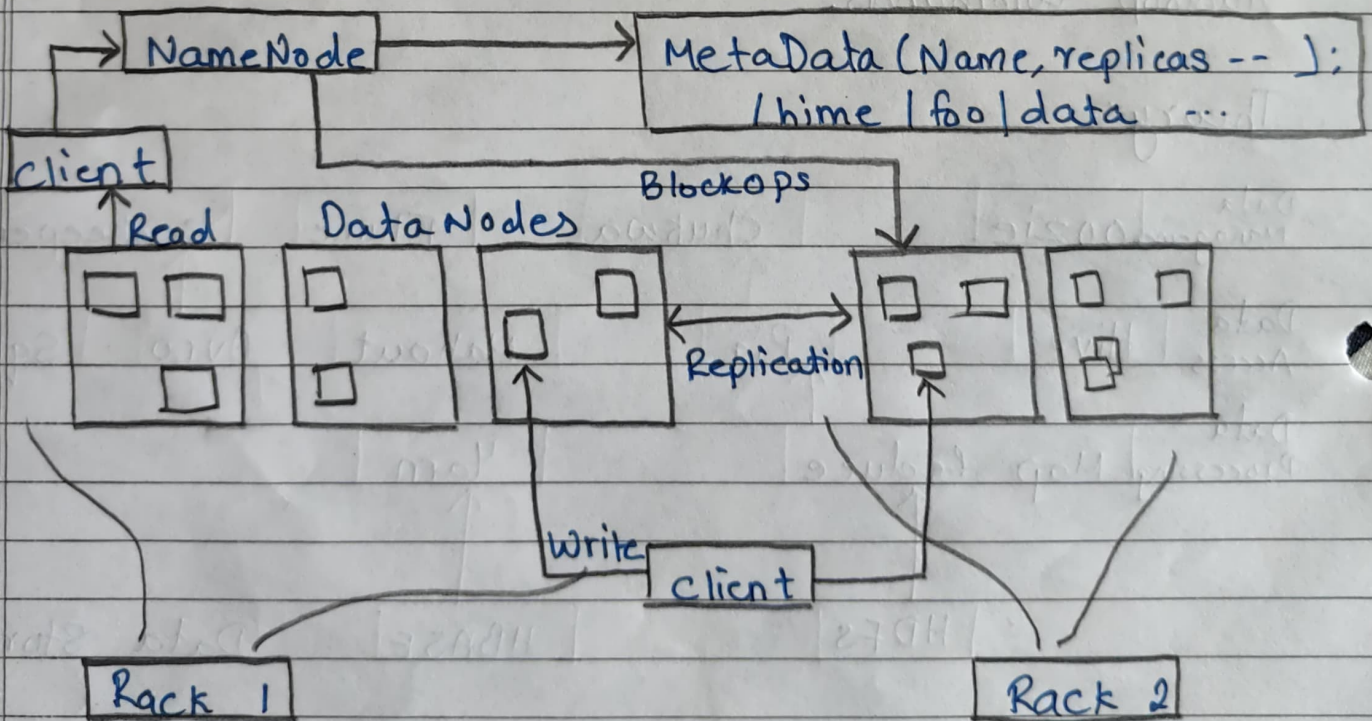
1. To learn concepts of Hadoop ecosystem
2. To learn how to install hadoop and perform basic Hadoop commands.

Theory:



Hadoop Ecosystem

- Hadoop is an open-source framework for processing and storing big data. It was created by Doug Cutting and Mike Cafarella in 2005. The project was originally developed to support the Nutch search engine. Cutting named the project after his son's toy elephant. Hadoop was inspired by Google's MapReduce, which splits an application into small fractions to run on different nodes. Cutting joined Yahoo in 2006 to scale the Hadoop project to thousands of nodes. It was originally called the Nutch Distributed File system and was developed as part of the Nutch project in 2004. It officially became part of Apache Hadoop in 2006.



HDFS Architecture

The Hadoop Distributed File System (HDFS) uses a master-slave architecture. The master device controls one or more slave devices. The HDFS cluster consists of a single Name Node. The Name Node is the master node and handles all the blocks on the DataNodes. The DataNodes are the slave nodes.

The HDFS architecture includes the following elements: NameNode, Secondary NameNode, DataNode, Checkpoint Node, Backup Node, Blocks, Write Operation, Read operation.

HDFS is fault-tolerant and is managed through the replication process. Stores files as data blocks.

Platform : 64-bit Open Source Linux/Windows.

Conclusion: Hence, I learned to install Hadoop and perform basic hadoop commands on it.

FAQ's

1. Explain with syntax and example of any 10 basic Hadoop Commands.

Ans. 1. Hadoop Version check:

Syntax : `hadoop version`

ex : `hadoop version`

2. List files and Directories in HDFS:

Syntax : `hadoop fs -ls [path]`

ex : `hadoop fs -ls /user/username`

3. Create a Directory in HDFS:

Syntax: `hadoop fs-mkdir [path]`

ex: `hadoop fs-mkdir /user/username/new-directory`

4. Copy Local file to HDFS:

Syntax: `hadoop fs-copyfromlocal [hdfs-path]`

ex: `hadoop fs-copyfromlocal localfile.txt /user/username/ame/hdfs-path/`

5. Copy file from HDFS to local file system:

Syntax: `hadoop fs-copytolocal [hdfs-path] [local-path]`

6. ~~Read~~ Read file:

Syntax: `hadoop fs-cat [hdfs-path]`

7. Delete file:

Syntax: `hadoop fs-rm [-r] [hdfs-path]`

ex: `hadoop fs-rm [/user/username/hdfs-path/file.txt]`

8. Run a MapReduce Job:

Syntax: `hadoop jar [jarfile] [mainclass] [inputpath] [output path]`

9. View Job Status and logs:

Syntax: `hadoop job-list`

To view job logs: `hadoop job-logs [job-id]`

10. Set hadoop config Properties:

Syntax: `hadoop config [property-name] [property-value]`

2. State the use of Name Node and Data Node

Ans. Data Nodes are the work horses of the filesystem, while name nodes keep the index of which block is stored in which data Node, datanodes store the actual data.

3. State the different applications of Hadoop.

Ans. - Stream processing
- fraud detection
- Content management
- Government Agencies
- Financial Sectors.