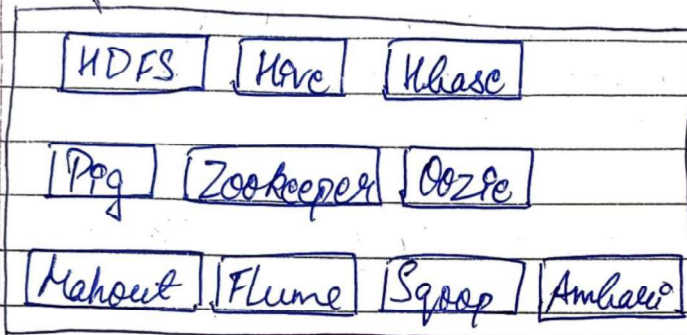


Hadoop - [Store + Process]

Components $\begin{cases} \text{HDFS} \\ \text{Map Reduce} \end{cases}$

Ecosystem



① HDFS

Stores the data files close to original items types.
It concentrates only on storage.
It distributes data across several nodes.
~~It is~~ It is naturally redundant.

② Hbase

Hadoop Database - It compares with RDBMS and it supports structured data storage.

③ Hive

Hive Query Language - Stores unstructured data.

④ Pig

Pig interpreter works like map reduce.
It has its own scripts [Pig Latin Commands]

⑤ Zookeeper

It is the coordinator.

⑥ Oozie

It is the scheduler.

⑦ Mahout

Machine Learning & Data Mining approaches maintained.

⑧ Flume (Chukua)

How to transfer data between distributed system.

Data Collector / Log Data Collector.

⑨ Sqoop

Transfer bulk amount of data between Hadoop and Structured data.

⑩ Ambari

Manage / Monitor / Process entire ecosystem.

MapReduce

It is a computational framework.

It splits a task across multiple nodes.

It processes the data parallelly.

Linux Commands

1) Clear

- The terminal screen is cleared

2) Pwd

- Working directory path is displayed

3) ls

- List all the files in the directory

4) who

- Displays information of all users currently logged in server

5) whoami

- Displays username of current user

6) date

- Displays current date and time

7) cat

- Saves the echo into file

- Prints the contents of file

8) rm -r

- Delete a directory

9) touch

- Create empty files

10) cp

- Copy a file

11) mv

- Move a file from one location to another

12) rename

- Rename a large group of files

13) head

- Display first few lines of a file

14) tail

- Display bottom few lines of a file

15) `tac`

- Display contents of file in reverse order

16) `su`

- Provides administrative access to another user

17) `id`

- Display the user id (UID) and group id (GID)

18) `useradd`

- Add or remove a user on a Linux system

19) `passwd`

- Create and change password for a user.

20) `groupadd`

- Create a user group

21) `cut`

- Select a specific column of a file

22) `wc`

- Count the lines, words and characters in a file

23) `find`

- Find a particular file within a directory

24) `locate`

- Search a file by file name.

25) `cal`

- Display the current month's calendar with current date highlighted.

① Display size of file-

`ls -h`

- Scales file sizes and directory sizes into KB, MB, GB or TB

② Identify the number of people

`cat /etc/passwd | wc -l`

- No. of lines in `/etc/passwd` corresponds to number of users

③ Command to create and open any pdf or text file

`nl file-name.txt`

- Displays file content with line numbers

`evince filename.pdf`

- Opens pdf file in GNOME document viewer