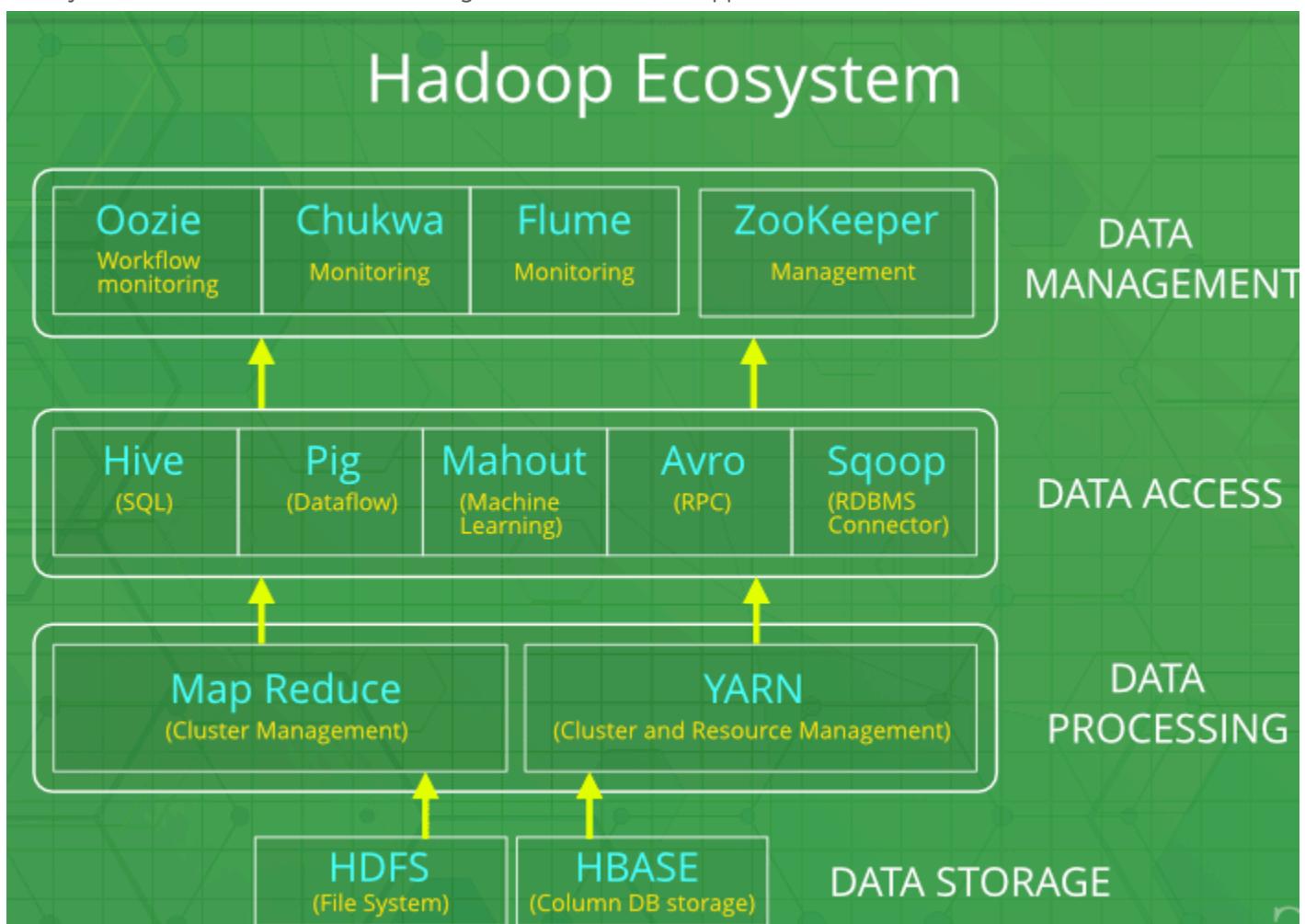


Overview of Hadoop

HDFS Overview

HDFS (Hadoop Distributed File System) is a distributed storage system in Hadoop designed to store large datasets reliably and to stream those datasets at high bandwidth to user applications.

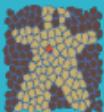




Non-relational
DBMS



Search Platform



Apache Giraph
Graph Processing



SQL
Interface
Apache Hive



Apache Pig
Scripting
Language



Application
Programming
Interface
(API)



MapReduce

Processing
Framework Layer

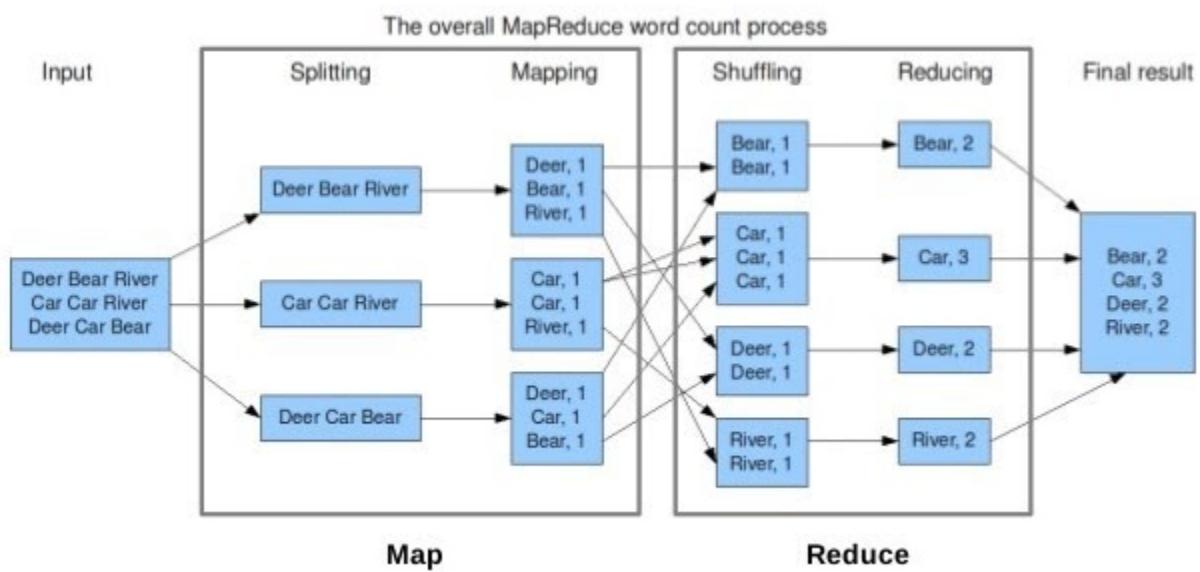
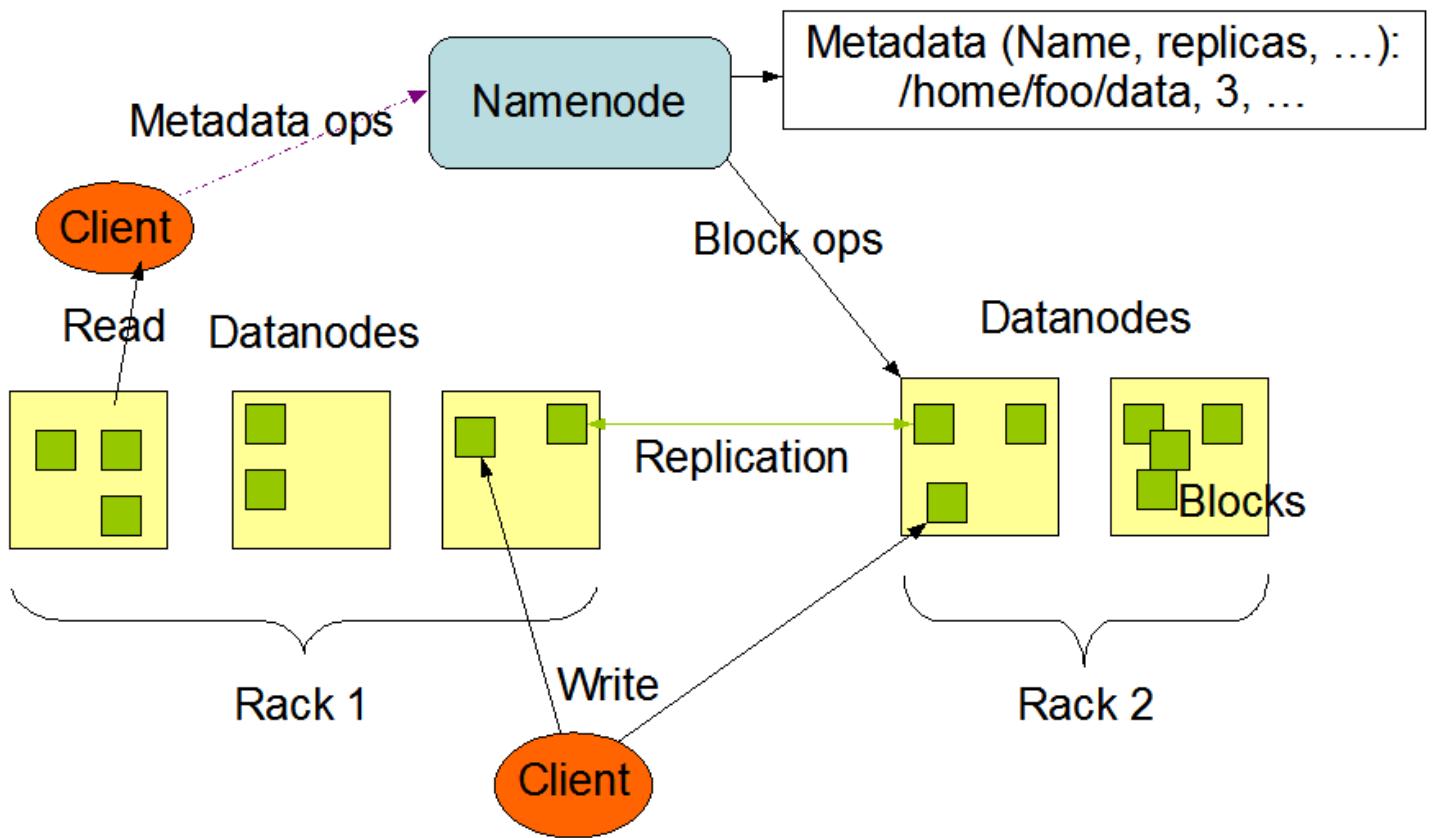
YARN

Cluster Resource Management

HDFS

Distributed Storage Layer

HDFS Architecture



Two main command-line interfaces:

- hdfs dfs → Recommended, works with Hadoop 2.x+.
- hadoop fs → Older, still works for compatibility.

◆ HDFS Commands with Explanation

1. Basic Directory Operations

Command	Description	Example
hdfs dfs -ls /	List files/directories in HDFS	hdfs dfs -ls /user
hdfs dfs -ls -R /	Recursive list	hdfs dfs -ls -R /user/data
hdfs dfs -mkdir /path	Create a directory	hdfs dfs -mkdir /user/dani
hdfs dfs -mkdir -p /path/subpath	Create parent directories if not exist	hdfs dfs -mkdir -p /user/dani/input
hdfs dfs -rm /path/file	Delete a file	hdfs dfs -rm /user/dani/data.txt
hdfs dfs -rm -r /path	Delete directory recursively	hdfs dfs -rm -r /user/dani/data
hdfs dfs -rmdir /path	Remove empty directory	hdfs dfs -rmdir /user/old

2. File Upload & Download

Command	Description	Example
hdfs dfs -put localfile /hdfs/path	Upload file from local to HDFS	hdfs dfs -put data.txt /user/dani
hdfs dfs -put -f localfile /hdfs/path	Force overwrite if file exists	hdfs dfs -put -f data.txt /user/dani
hdfs dfs -copyFromLocal localfile /hdfs/path	Same as -put	hdfs dfs -copyFromLocal data.txt /user/dani
hdfs dfs -get /hdfs/path localpath	Download from HDFS to local	hdfs dfs -get /user/dani/data.txt ./
hdfs dfs -copyToLocal /hdfs/path localpath	Same as -get	hdfs dfs -copyToLocal /user/dani/data.txt ./
hdfs dfs -getmerge /hdfs/dir localfile	Merge multiple HDFS files into one local file	hdfs dfs -getmerge /user/dani/output result.txt

3. File Read & View

Command	Description	Example
hdfs dfs -cat /path/file	View file contents	hdfs dfs -cat /user/dani/data.txt
hdfs dfs -tail /path/file	Show last KB of file	hdfs dfs -tail /user/dani/log.txt
hdfs dfs -text /path/file	View compressed file contents	hdfs dfs -text /user/dani/data.gz

4. File Management

Command	Description	Example
hdfs dfs -mv /source /dest	Move file	hdfs dfs -mv /user/dani/data.txt /user/dani/archive/
hdfs dfs -cp /source /dest	Copy file	hdfs dfs -cp /user/dani/data.txt /user/dani/backup/
hdfs dfs -du /path	Disk usage	hdfs dfs -du /user/dani
hdfs dfs -du -h /path	Human-readable disk usage	hdfs dfs -du -h /user/dani
hdfs dfs -count /path	Count files and dirs	hdfs dfs -count /user/dani
hdfs dfs -count -q /path	Show quota and space	hdfs dfs -count -q /user/dani

5. Permissions

Command	Description	Example
hdfs dfs -chmod 755 /path	Change permissions	hdfs dfs -chmod 755 /user/dani/data.txt
hdfs dfs -chown user:group /path	Change owner	hdfs dfs -chown dani:hadoop /user/dani/data.txt
hdfs dfs -chgrp group /path	Change group	hdfs dfs -chgrp hadoop /user/dani/data.txt

📌 HDFS Command Cheat Sheet (Quick Copy)

bash

List files

```
hdfs dfs -ls /path
hdfs dfs -ls -R /path # recursive
```

Directory operations

```
hdfs dfs -mkdir /path
hdfs dfs -mkdir -p /path/subpath
hdfs dfs -rm /path/file
hdfs dfs -rm -r /path/dir
hdfs dfs -rmdir /path
```

File upload & download

```
hdfs dfs -put localfile /hdfs/path
hdfs dfs -get /hdfs/path localpath
hdfs dfs -copyFromLocal localfile /hdfs/path
hdfs dfs -copyToLocal /hdfs/path localpath
hdfs dfs -getmerge /hdfs/dir localfile
```

Read file

```
hdfs dfs -cat /path/file
hdfs dfs -tail /path/file
hdfs dfs -text /path/file
```

Manage files

```
hdfs dfs -mv /src /dest
hdfs dfs -cp /src /dest
hdfs dfs -du /path
hdfs dfs -du -h /path
hdfs dfs -count /path
hdfs dfs -count -q /path
```

Permissions

```
hdfs dfs -chmod 755 /path
hdfs dfs -chown user:group /path
hdfs dfs -chgrp group /path
```

References:

https://hadoop.apache.org/docs/r1.2.1/hdfs_design.html

<https://hive.apache.org/>

<https://spark.apache.org/>