HDFS

* It provides distributed architecture for extremely large scale storage ,which can be extended by scaling out.
* HDFS has write once ,read often model of data access.
* It works with parallelism,which is the ability to process the individual blocks of these large files in parallel.
* The HDFS central metadata server is called Name node

HADOOP CLUSTER

It’s made up of two classes of servers

Slave node/data node

* Data -------> Stored and processed
* Raw data is stored in the form of data blocks
* It runs a background process (Daemon),keep track of slices of data that the system stores on its computer

Master node

* They govern the management of Hadoop clusters
* It contains metadata that maps data blocks to files stored in HDFS

Some Hadoop file system commands

$ hadoop fs -ls /

Found 6 items

drwxrwxrwx - hdfs supergroup 0 2017-10-23 10:29 /benchmarks

drwxr-xr-x - hbase supergroup 0 2022-01-05 23:14 /hbase

drwxr-xr-x - solr solr 0 2017-10-23 10:32 /solr

drwxrwxrwt - hdfs supergroup 0 2022-01-05 23:26 /tmp

drwxr-xr-x - hdfs supergroup 0 2020-05-22 02:53 /user

drwxr-xr-x - hdfs supergroup 0 2017-10-23 10:31 /var

$ hadoop fs -ls /user

Found 9 items

drwxr-xr-x - cloudera cloudera 0 2020-06-04 09:04 /user/cloudera

drwxr-xr-x - hdfs supergroup 0 2020-05-22 02:53 /user/hdfs

drwxr-xr-x - mapred hadoop 0 2022-01-05 23:27 /user/history

drwxrwxrwx - hive supergroup 0 2017-10-23 10:31 /user/hive

drwxrwxrwx - hue supergroup 0 2022-01-06 20:26 /user/hue

drwxrwxrwx - jenkins supergroup 0 2017-10-23 10:30 /user/jenkins

drwxrwxrwx - oozie supergroup 0 2017-10-23 10:30 /user/oozie

drwxrwxrwx - root supergroup 0 2017-10-23 10:30 /user/root

drwxr-xr-x - hdfs supergroup 0 2017-10-23 10:31 /user/spark

$ mkdir arjun\_dir

$ dir

anime.java Documents input.txt Public

arjun\_dir Downloads kerberos rakeshdata

avro eclipse lib rakeshdata1

cloudera-manager emp.java Music sparkjars\_exec

cm\_api.py enterprise-deployment.json parcels Templates

data1 express-deployment.json parquet\_write Videos

Desktop external\_jars part\_dir workspace

devices.json external-unified Pictures zeyo\_tab.java

$ vi bigdata.txt

$ cat bigdata.txt

Big data is the way of describing data problems that are unsolvable by traditional tools