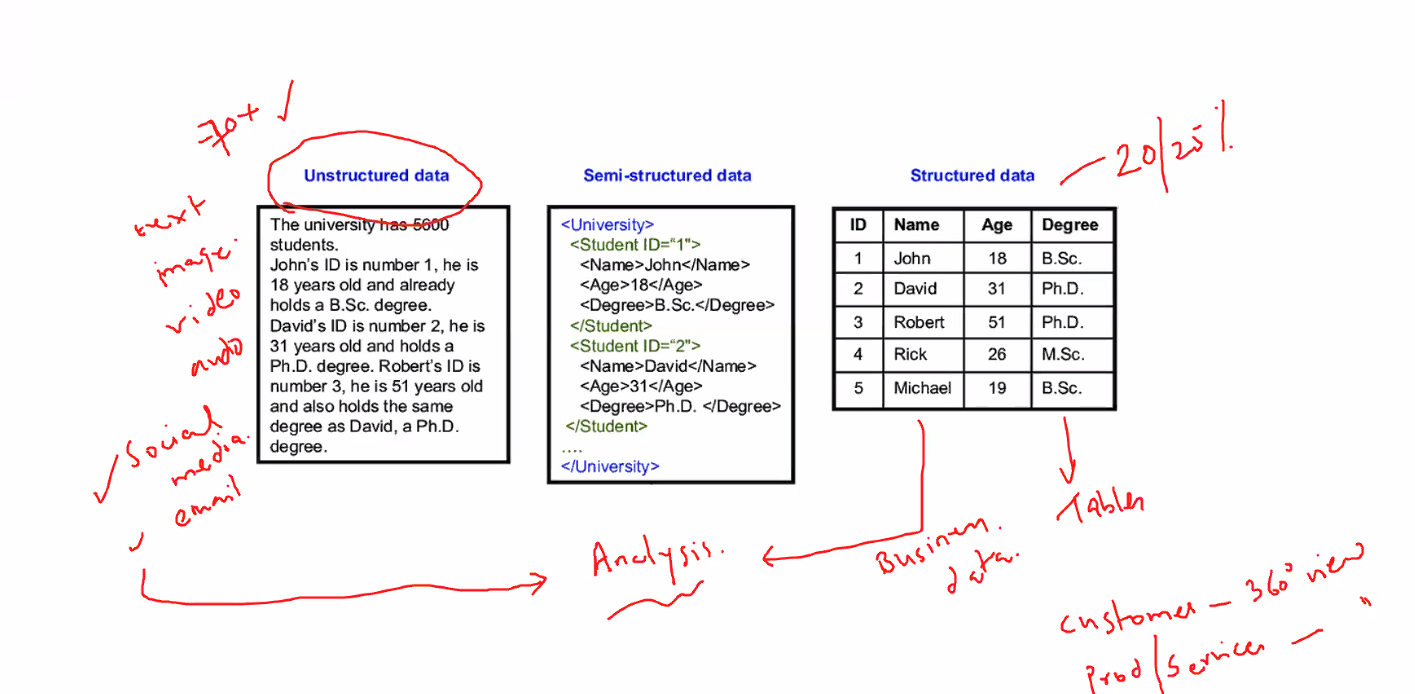
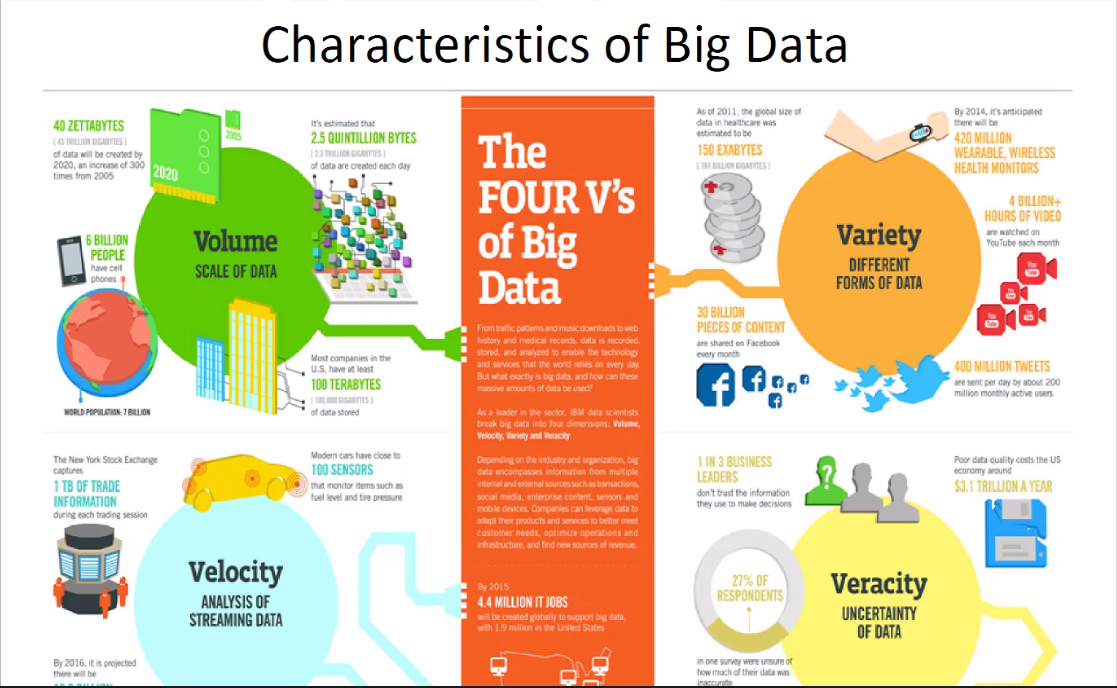
**Big Data**

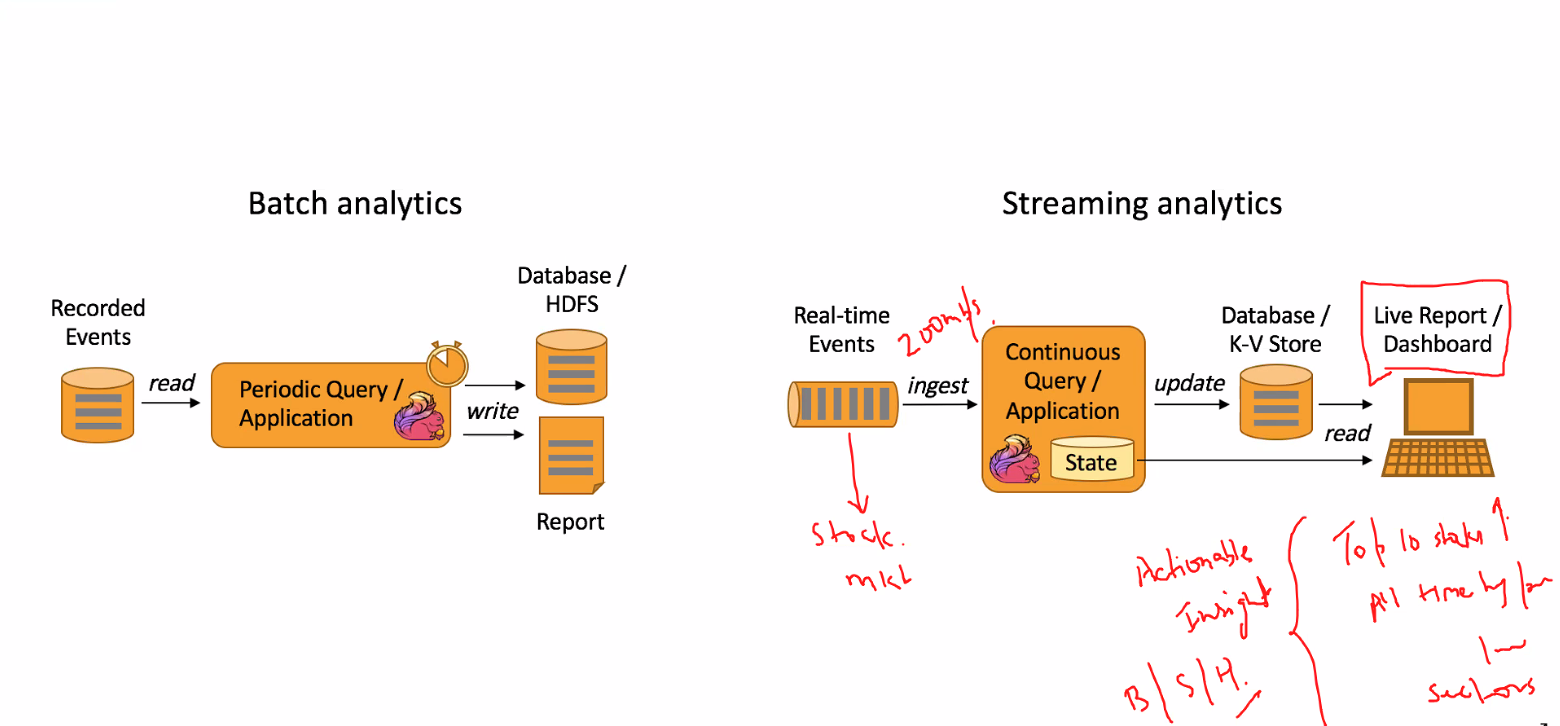
pwd

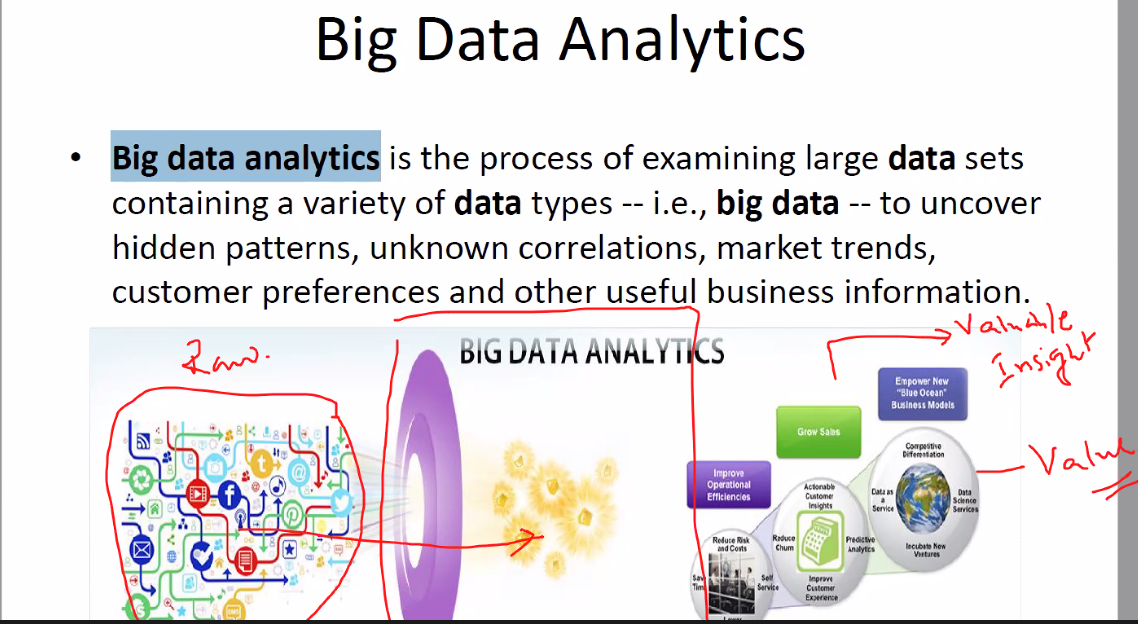
ls

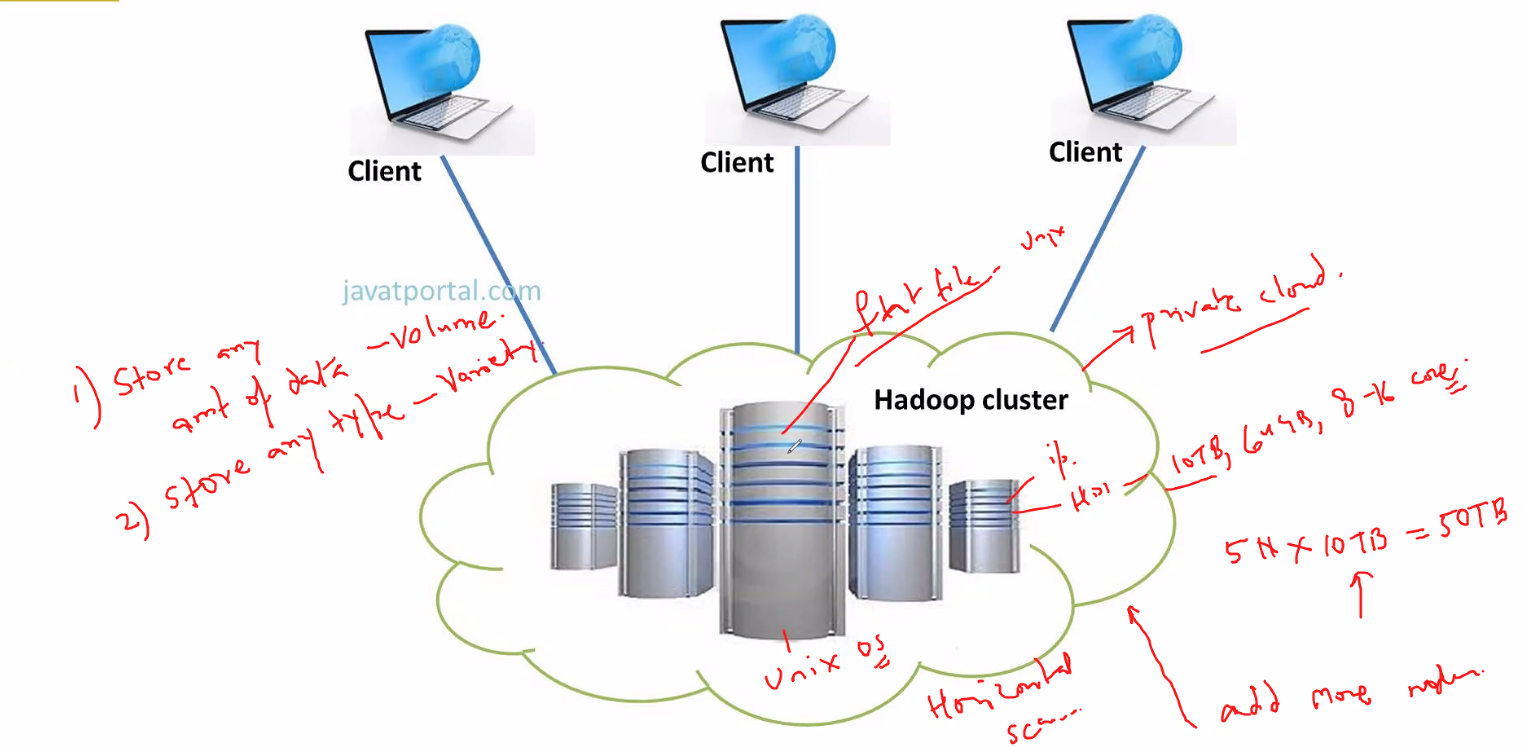




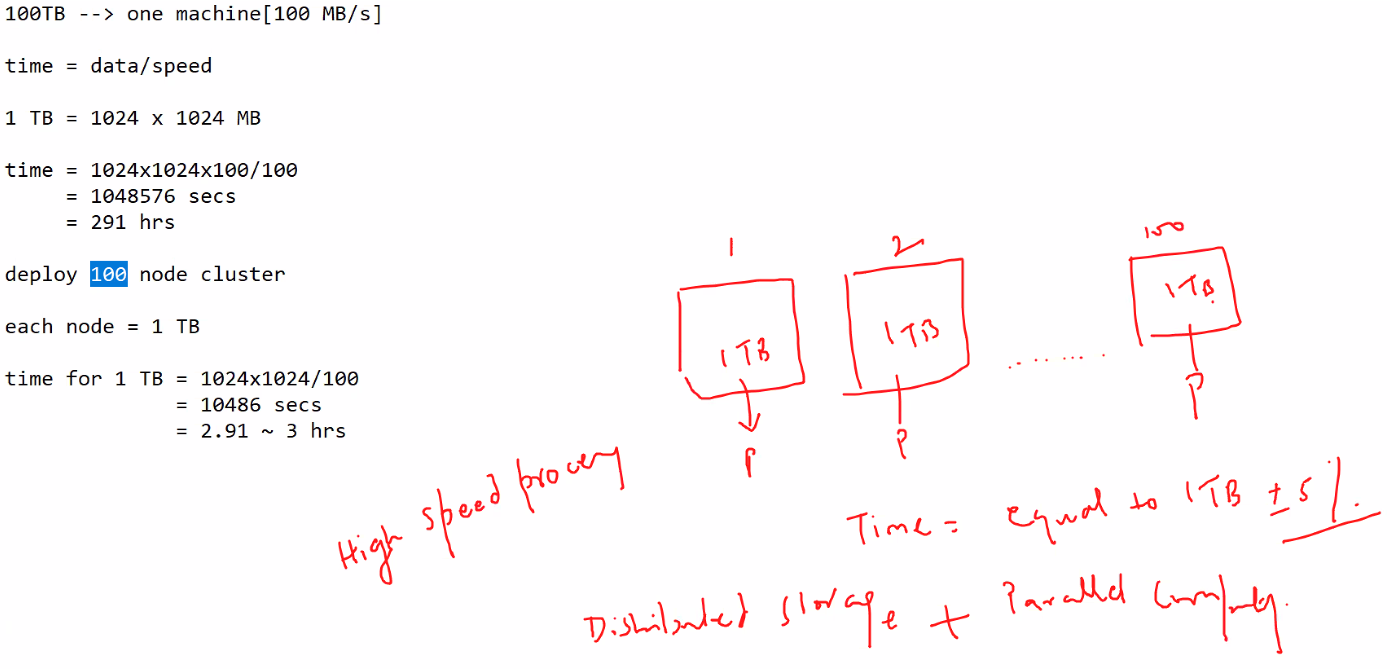
Veracity: Cleansing, preprocessing, validating

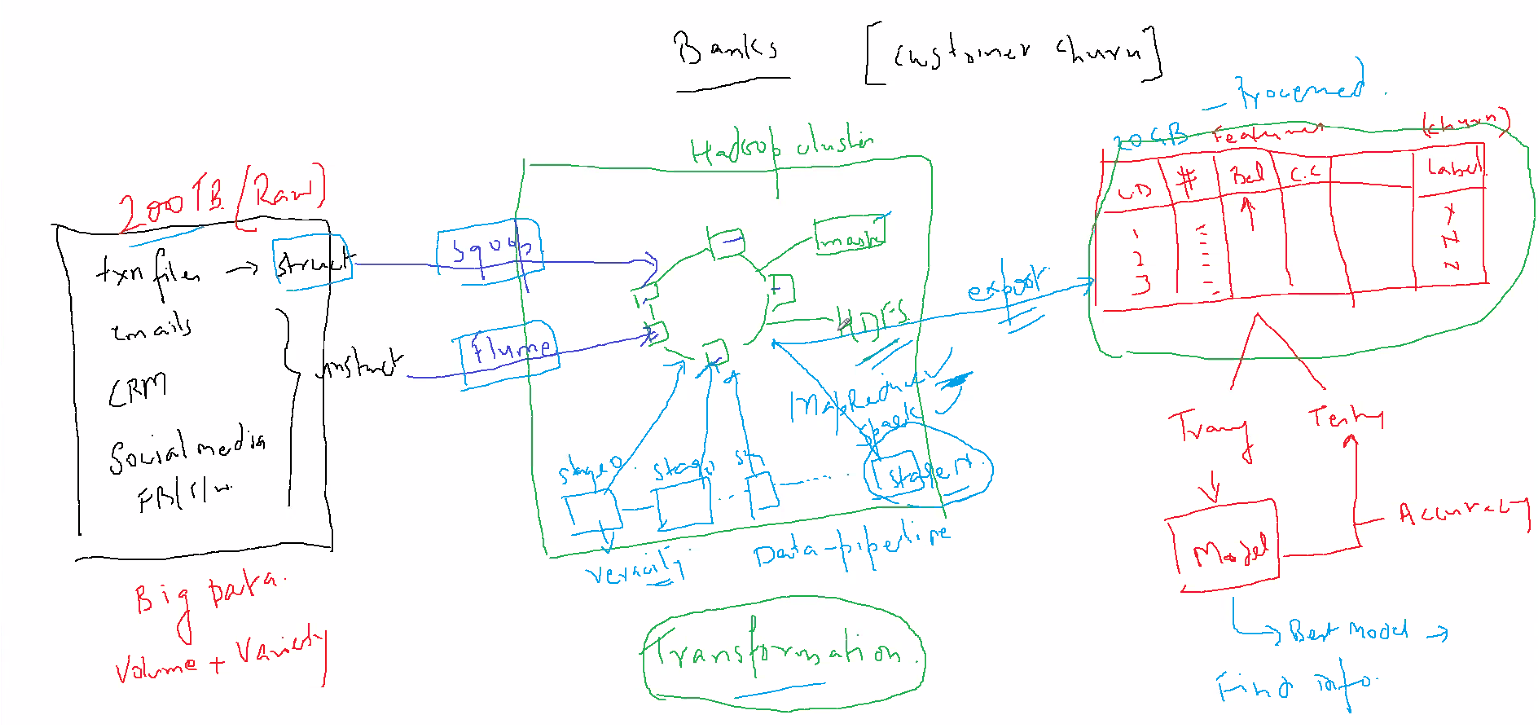






Hadoop -> Transform Unstructured data to structured processed data (processed in distributed storage and parallel computing is used)





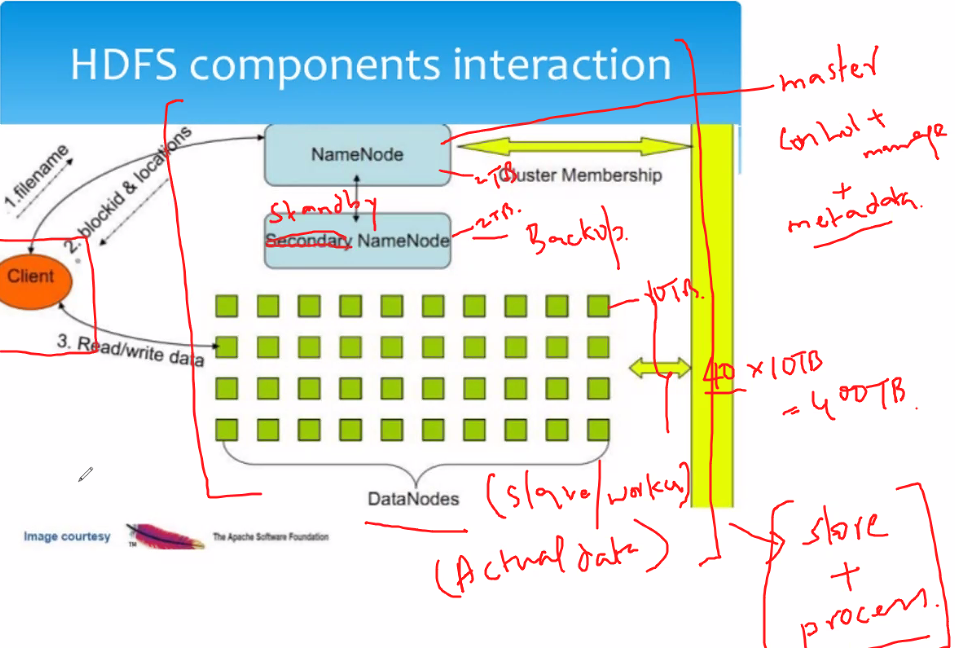
Sqoop => pull structured data faster

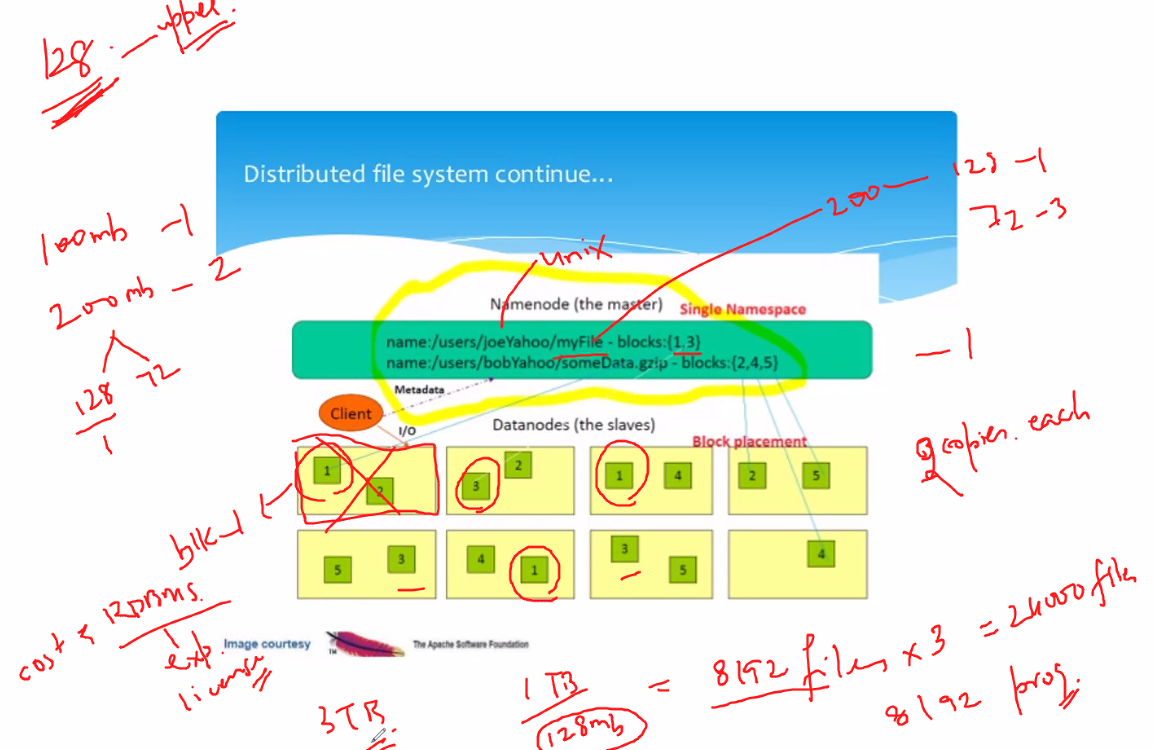
Flume => pull unstructured data faster

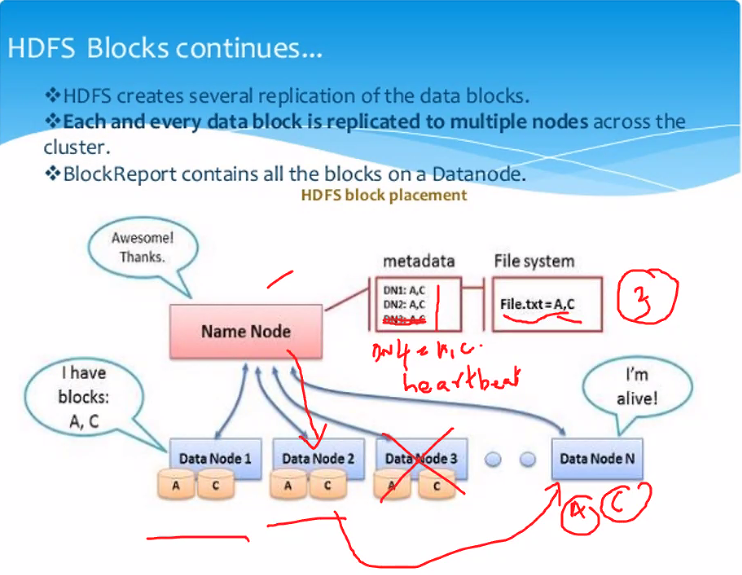
MapReduce/Spark => Algorithms to process distributed data

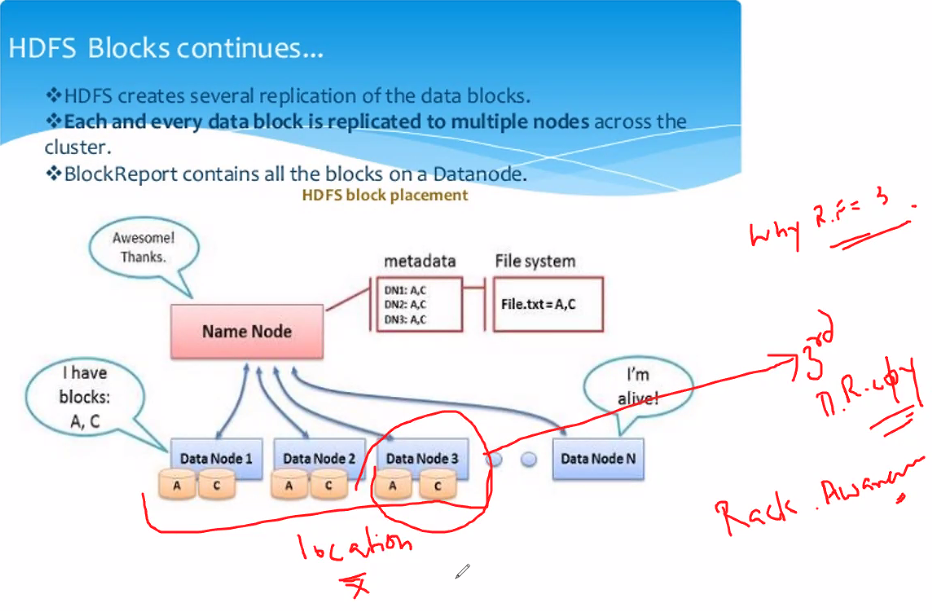
Yarn MapReduce written in Java

Hive => SQL query to Java





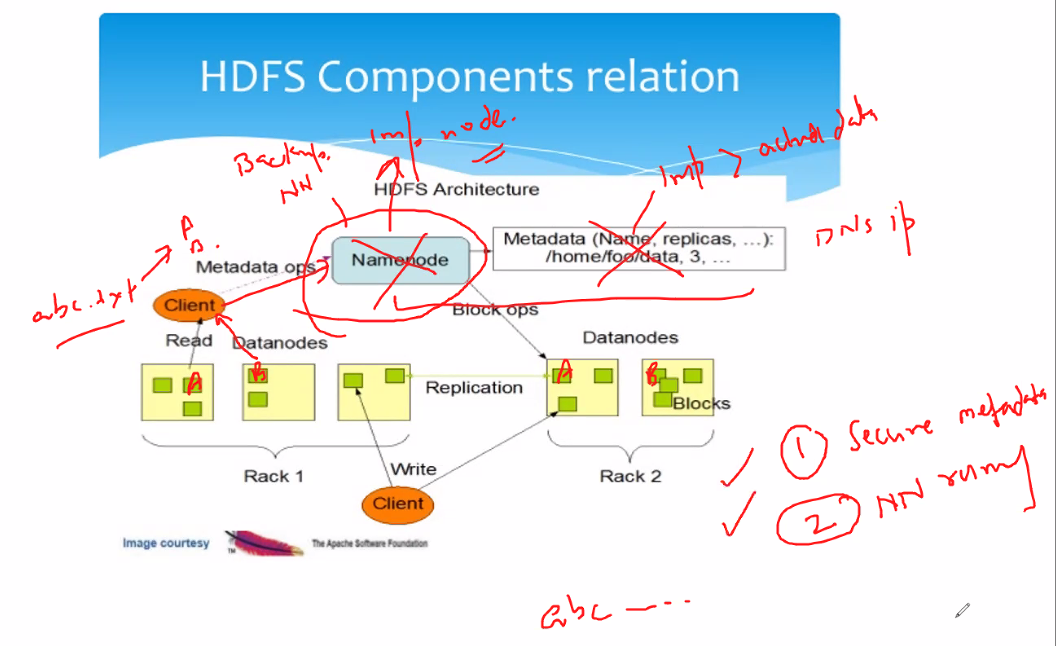


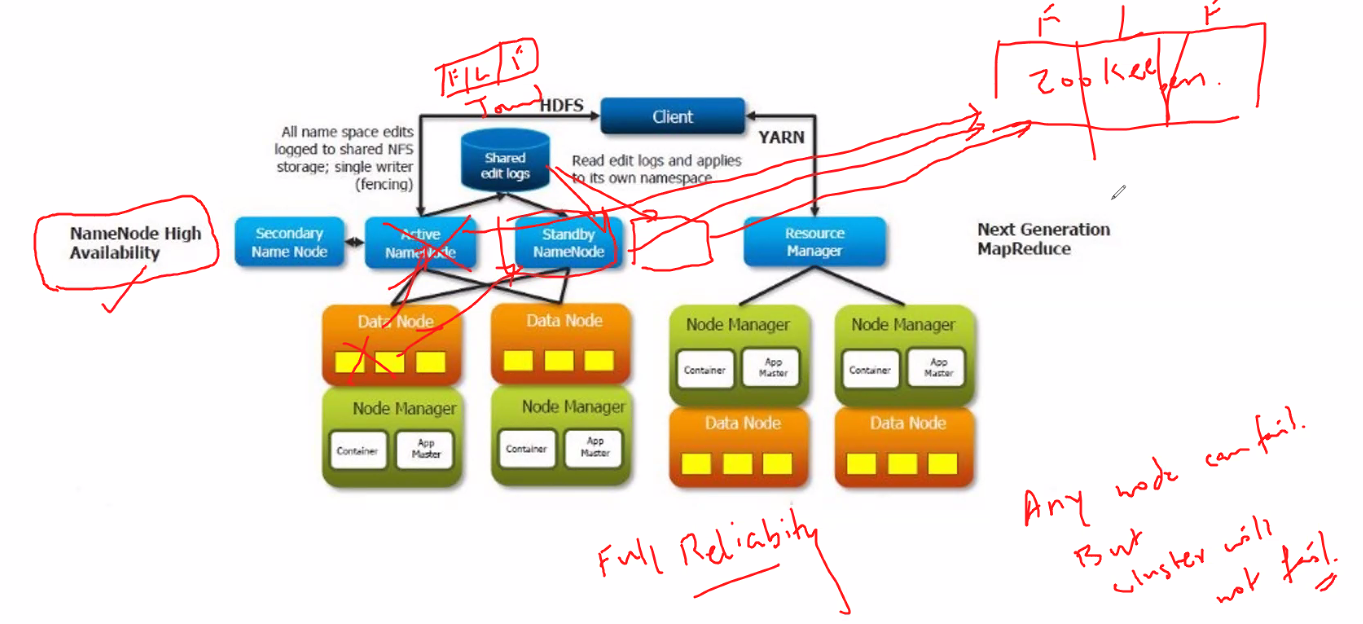


3rd copy => Different rack/location

DR => Disaster recovery

RF => Replication factor





**Linux commands**

lsb\_release –a

pwd => current working directory

machinename@hostname

~ denotes home/user directory

cd /

pwd

ls

ls –l => current directory files and folders

d represents directory

- represents files

cd home => change present directory to home

ls

ls -l

cd cloudera

pwd

mkdir folder1

ls

ls -l

cd folder1

ls

gedit file1.txt

ls -l

user, group & others

r – read => 4

w – write => 2

x – execute => 1

chmod 764 file1.txt

ls -l

chmod 724 file1.txt

ls -l

chmod 777 file1.txt

ls -l

chmod 700 file.txt

ls -l

7 = 4+2+1 (rwx)

user – group – others (permissions)

gedit (graphic edit system)

gedit file1.txt

ls -l

chmod 500 file1.txt

ls -l

gedit file1.txt

chmod 700 file1.txt

gedit file1.txt

cat file1.txt (read file)

chmod 300 file1.txt

ls -l

cat file1.txt

chmod 700 file1.txt

cat file1.txt

nano file1 (create new file without extension different from file1.txt)

Ctrl+X to exit

Enter Y for yes to save

ls -l

flat file system (no data type)

file names are case sensitive

nano File1

ls -l

cat File1.txt

pwd

cd ..

mkdir folder2

cp folder1/file1.txt folder2

cd folder2

ls -l

cd ..

mv folder1/file1 folder2

cd folder1

ls -l

cd ..

ls -l

cd folder2

ls –l

rm file1

ls –l

cd..

rm –R folder2

cd folder2

hadoop fs –ls /

hadoop fs –mkdir /user/cloudera/imart

ls –l NYSE.csv

ls –l eclipse.gz

hadoop fs –put NYSE.csv /user/cloudera/imart

hadoop fs –put eclipse.gz /user/cloudera/imart