

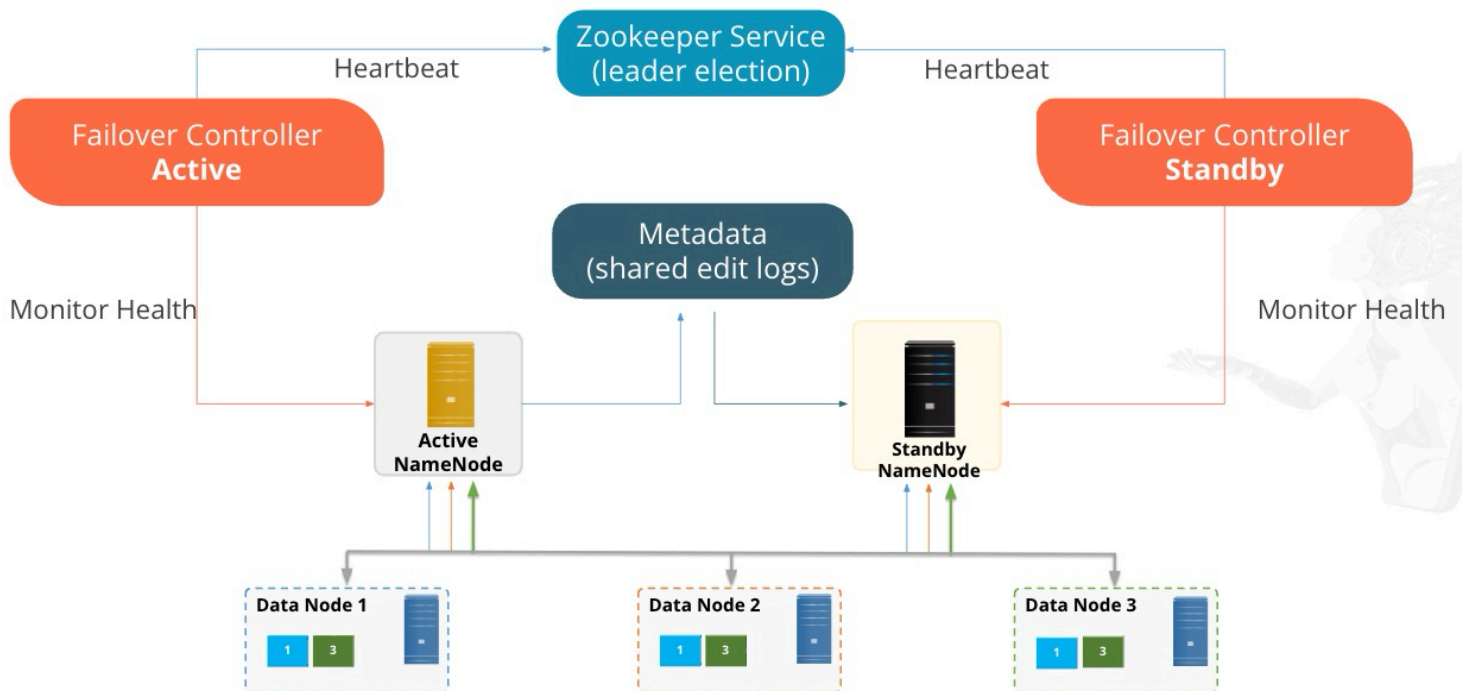
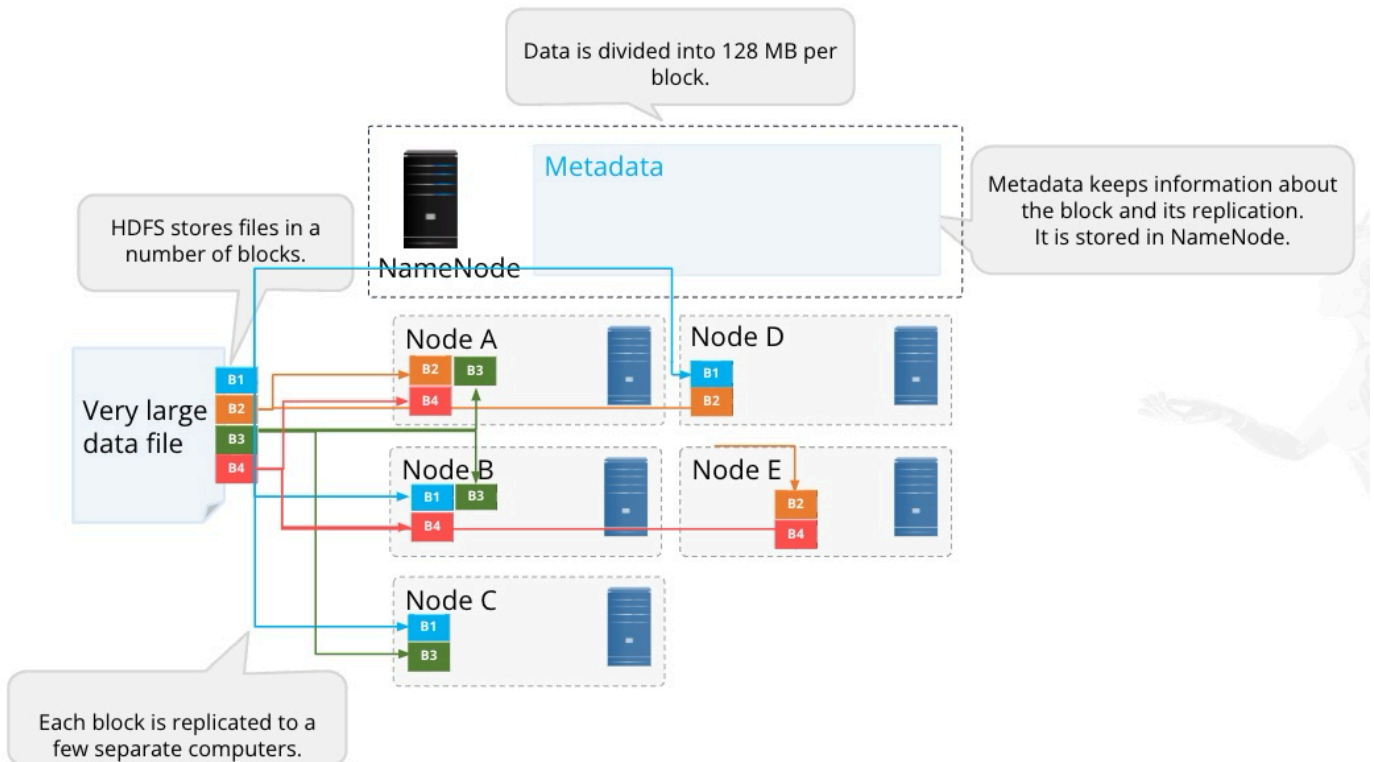
# Agenda

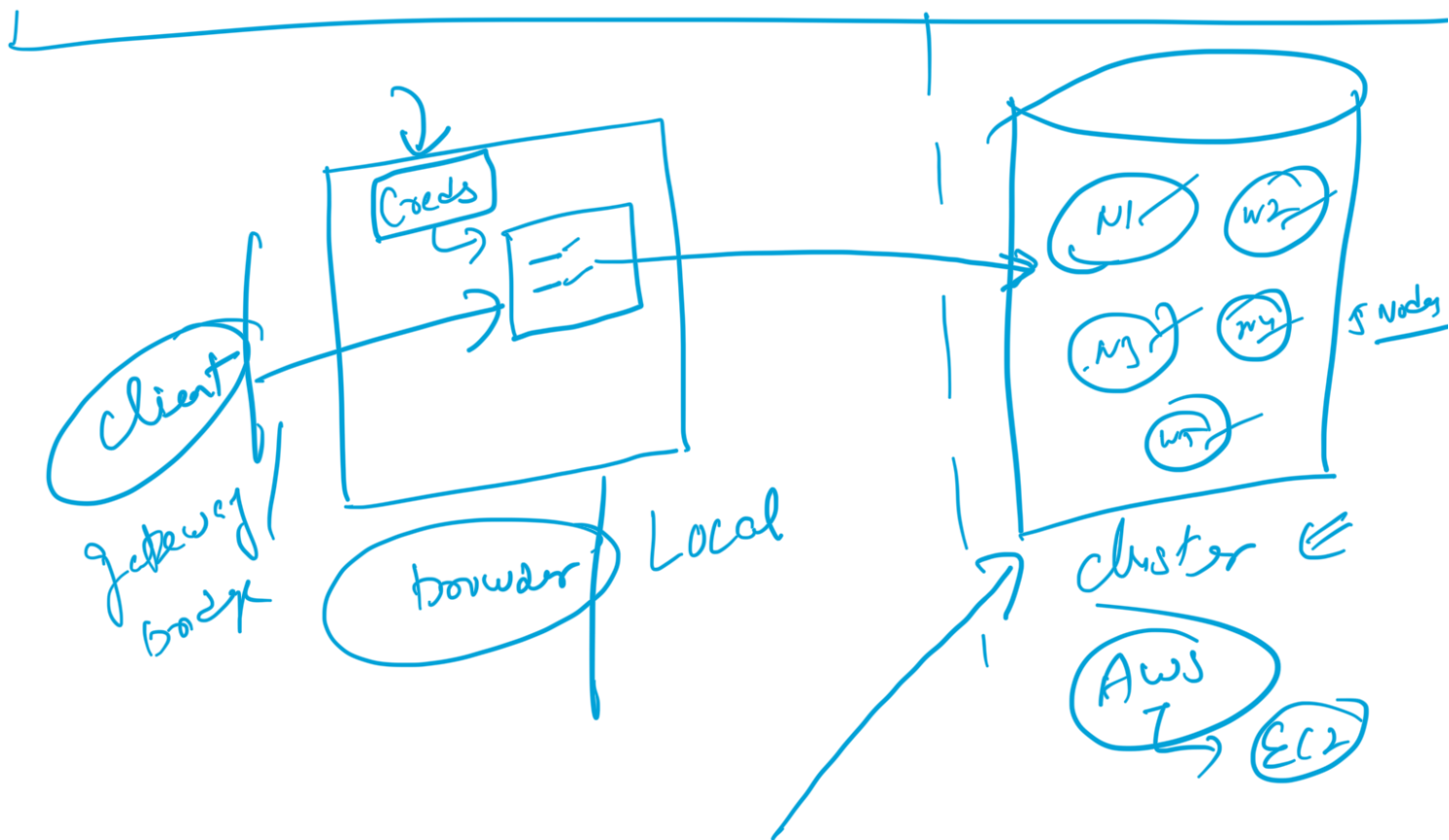
HDFS

YARN

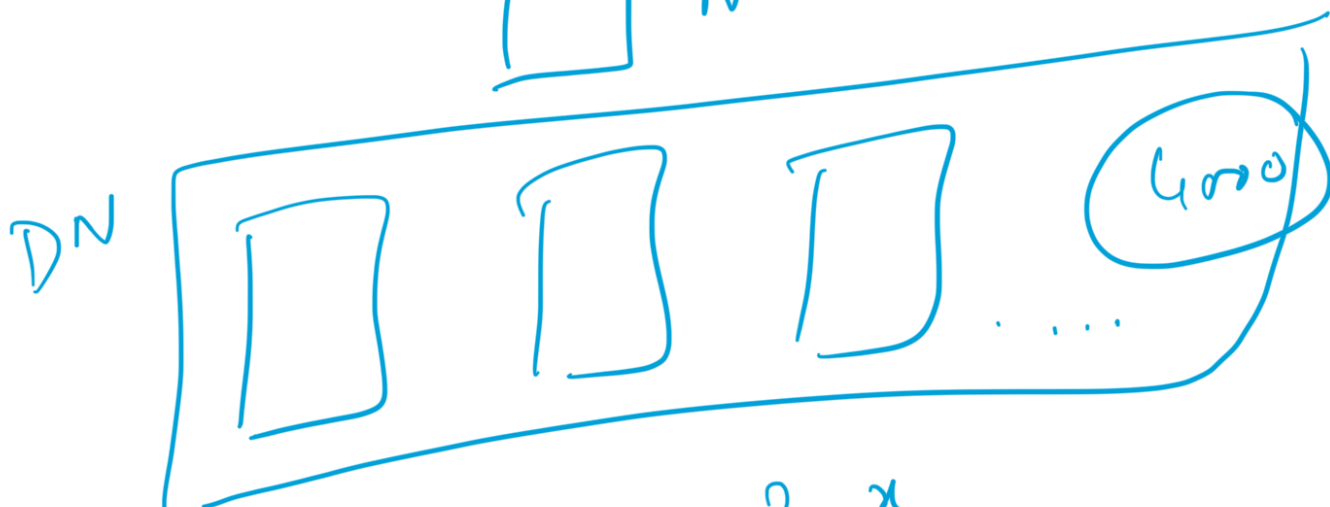


# HDFS Storage



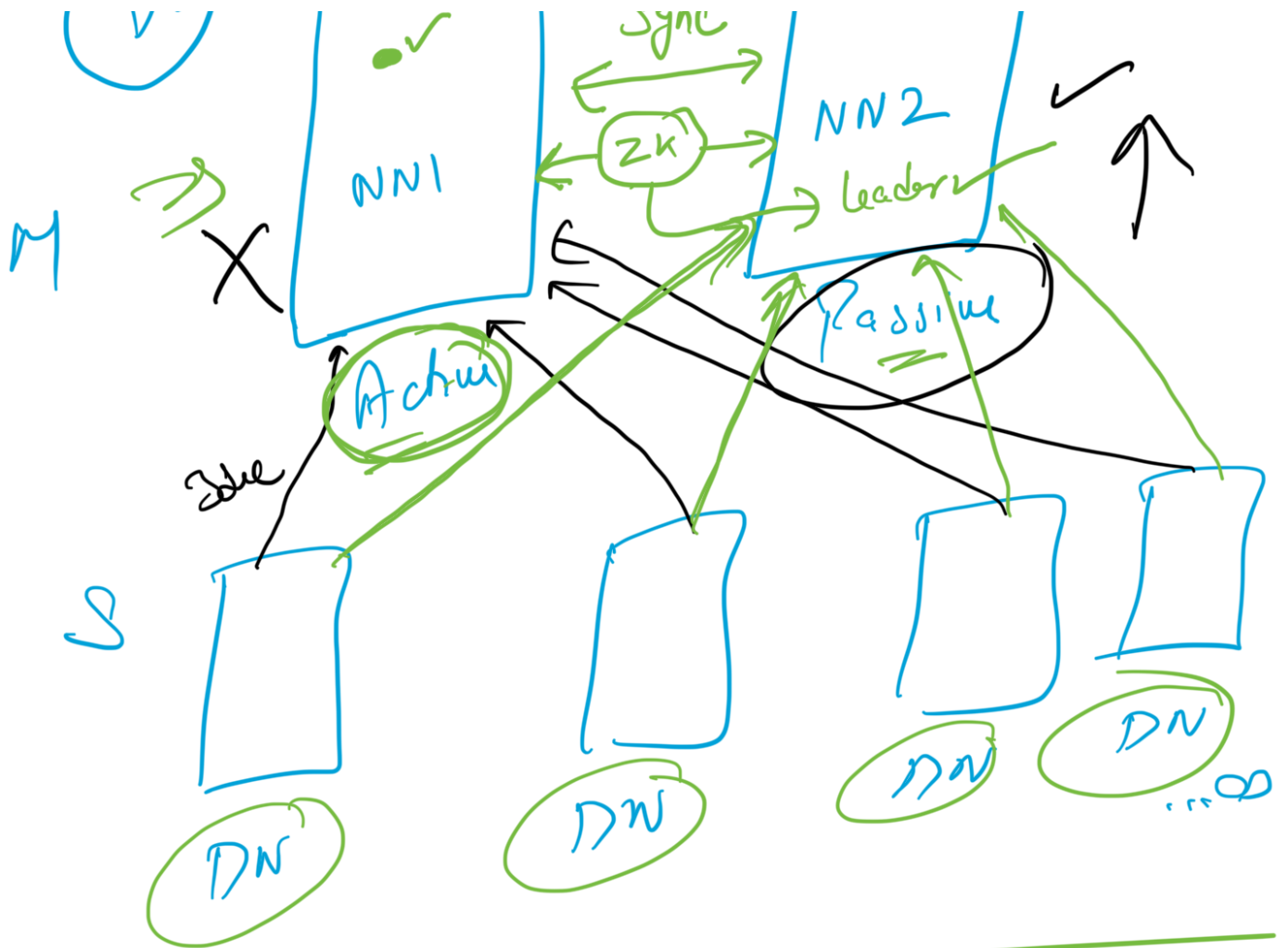


Hadoop V1.x



Hadoop 2.x

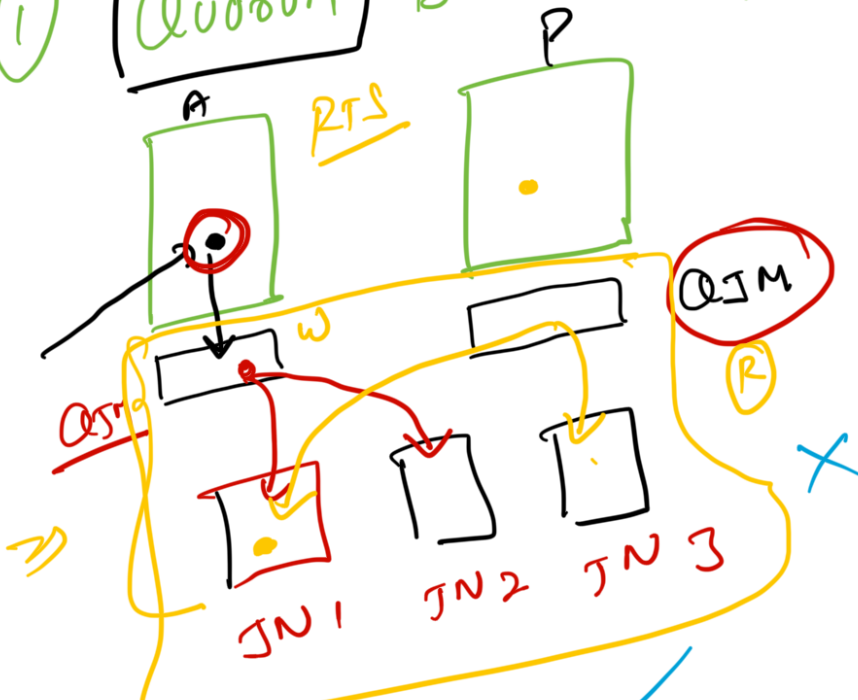




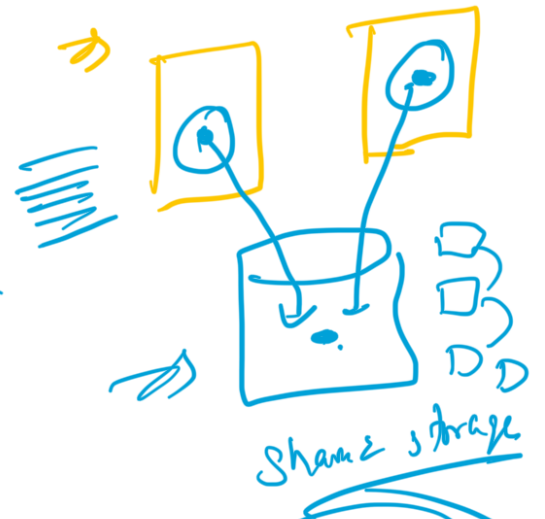
## Syncing

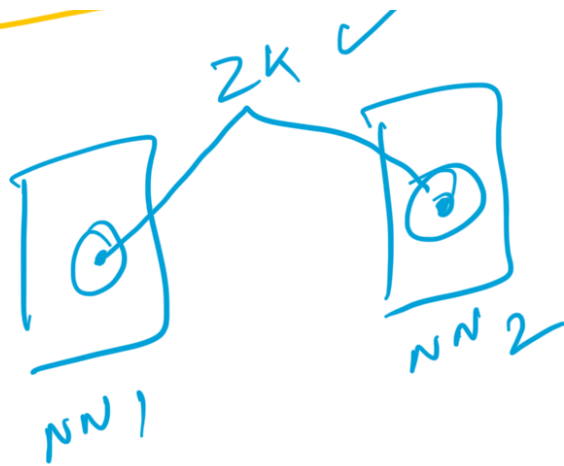
① Quorum

Based Storage



② Shared Storage using NFS.





X (RAID)  
 Redundant Array of Independent Disks

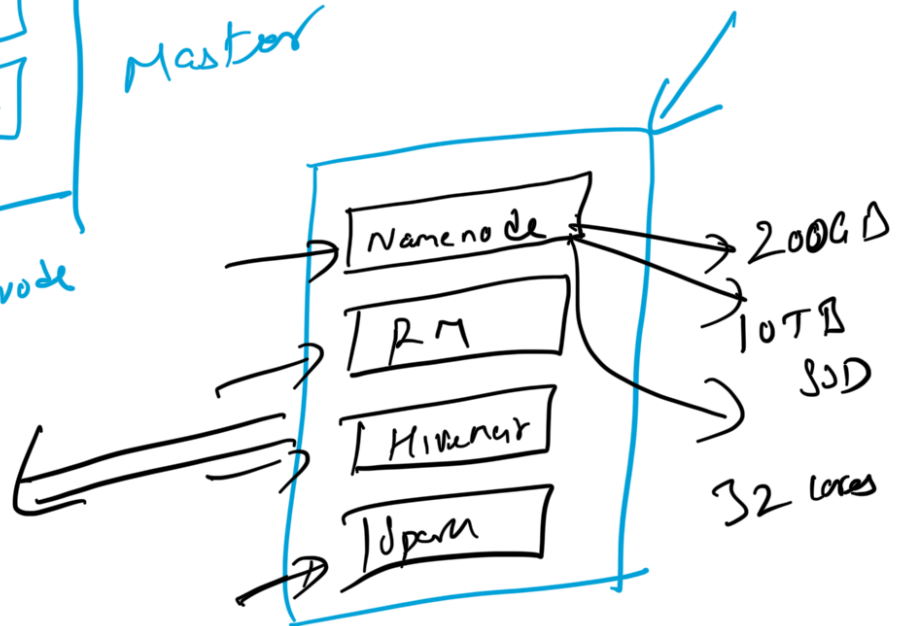


① namenode → is a misnomer ✓  
 → is a Service



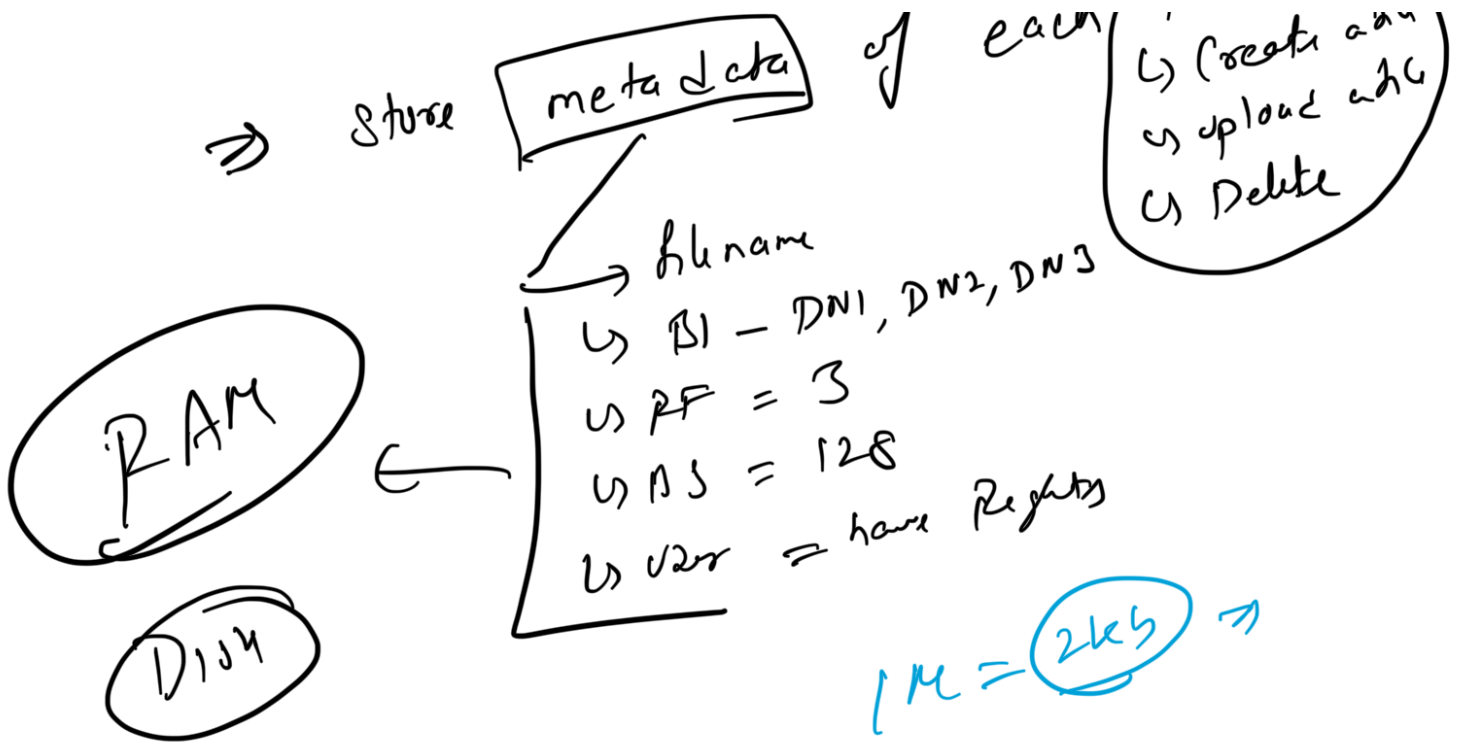
Master

namenode



Master → IN



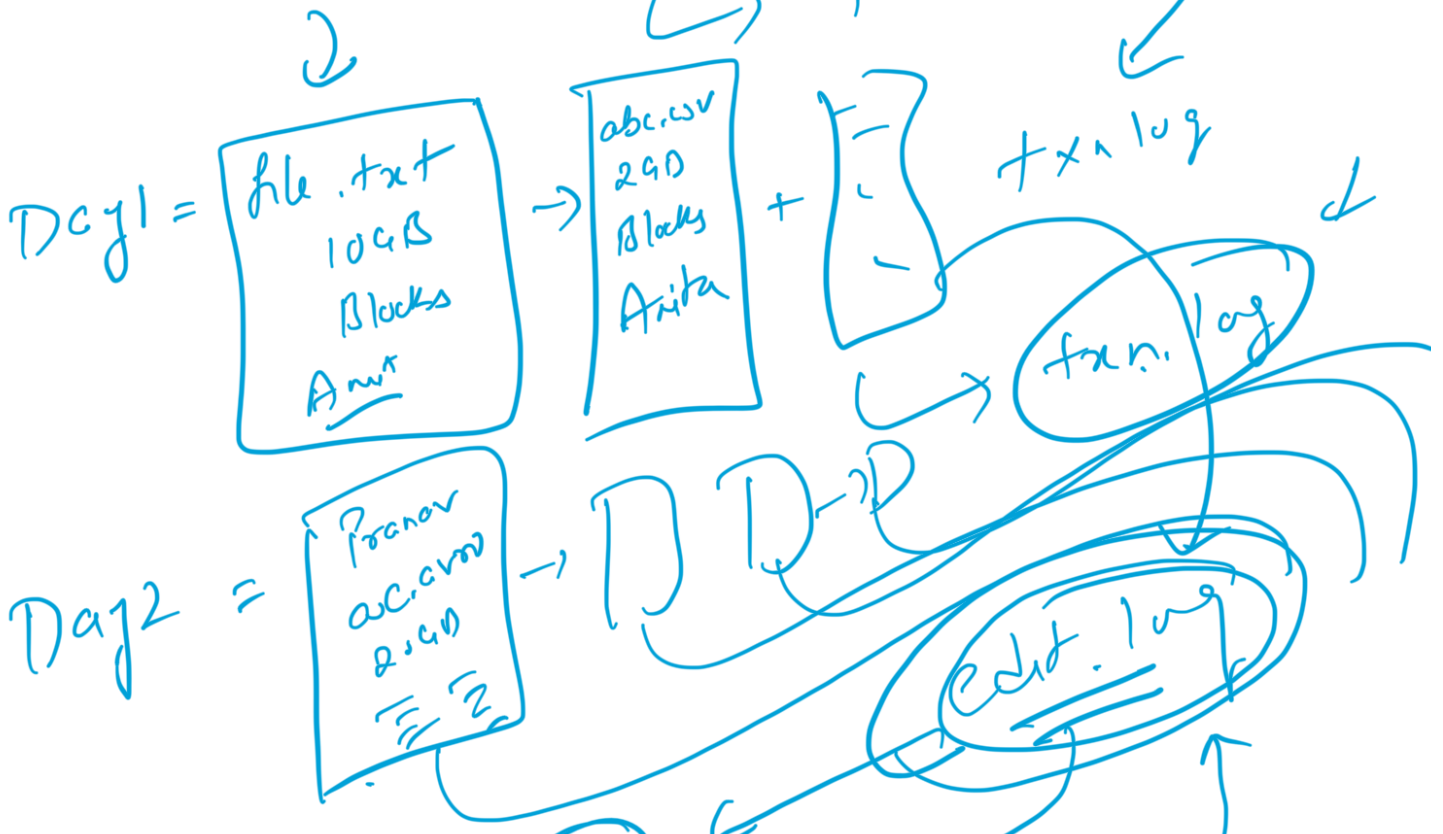


Metadata =

LOAD →

Name node operations :-

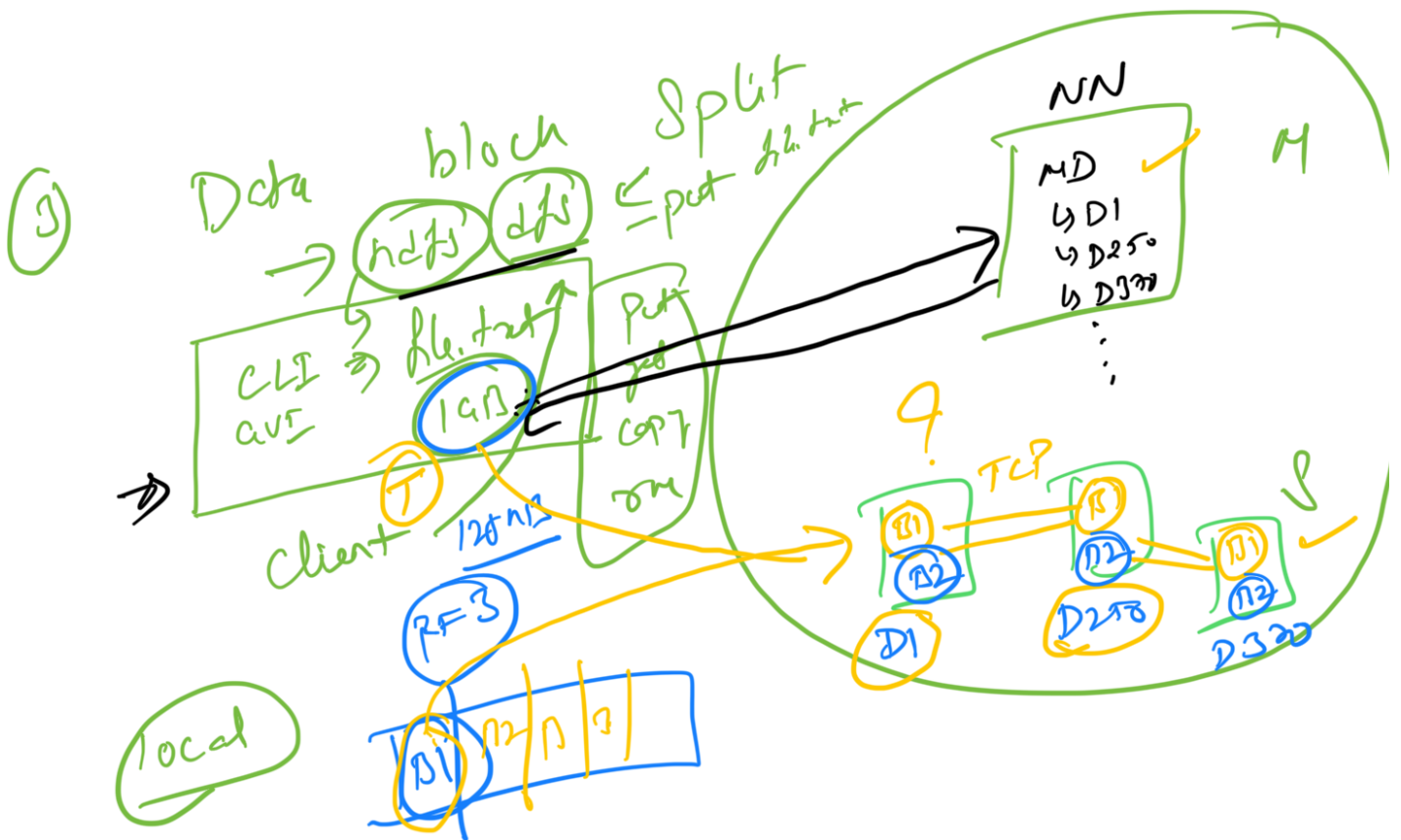
Edit log  
FS Image





Memory ←  
 1 user / unit | abc.log

- (2) Datanodes  
 ↳ multiple instance server  
 ↳ are responsible for storing and maintaining data/block blocks.



How does Replication happen?

→ Topology.py → IP → ①  
 ... and ... ②  
 ... ③

→ Topology . n.

Algo

