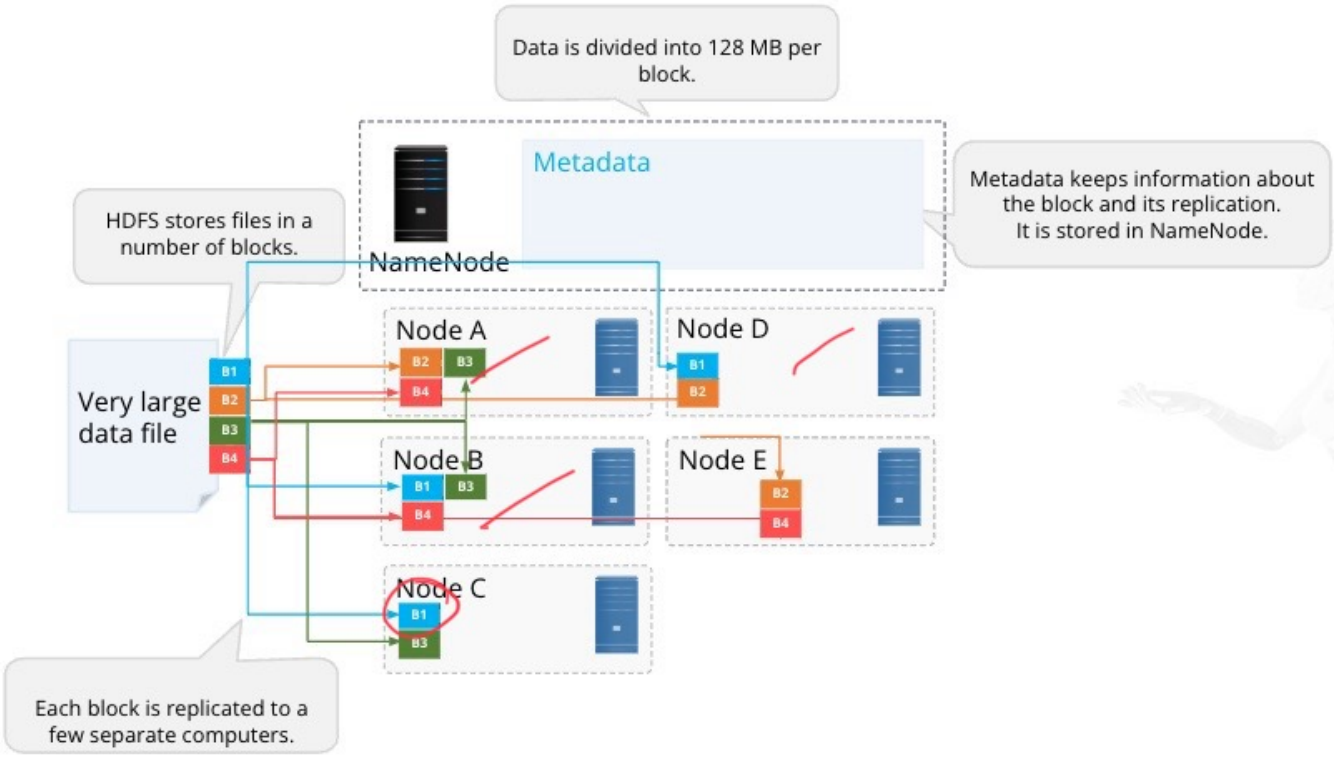
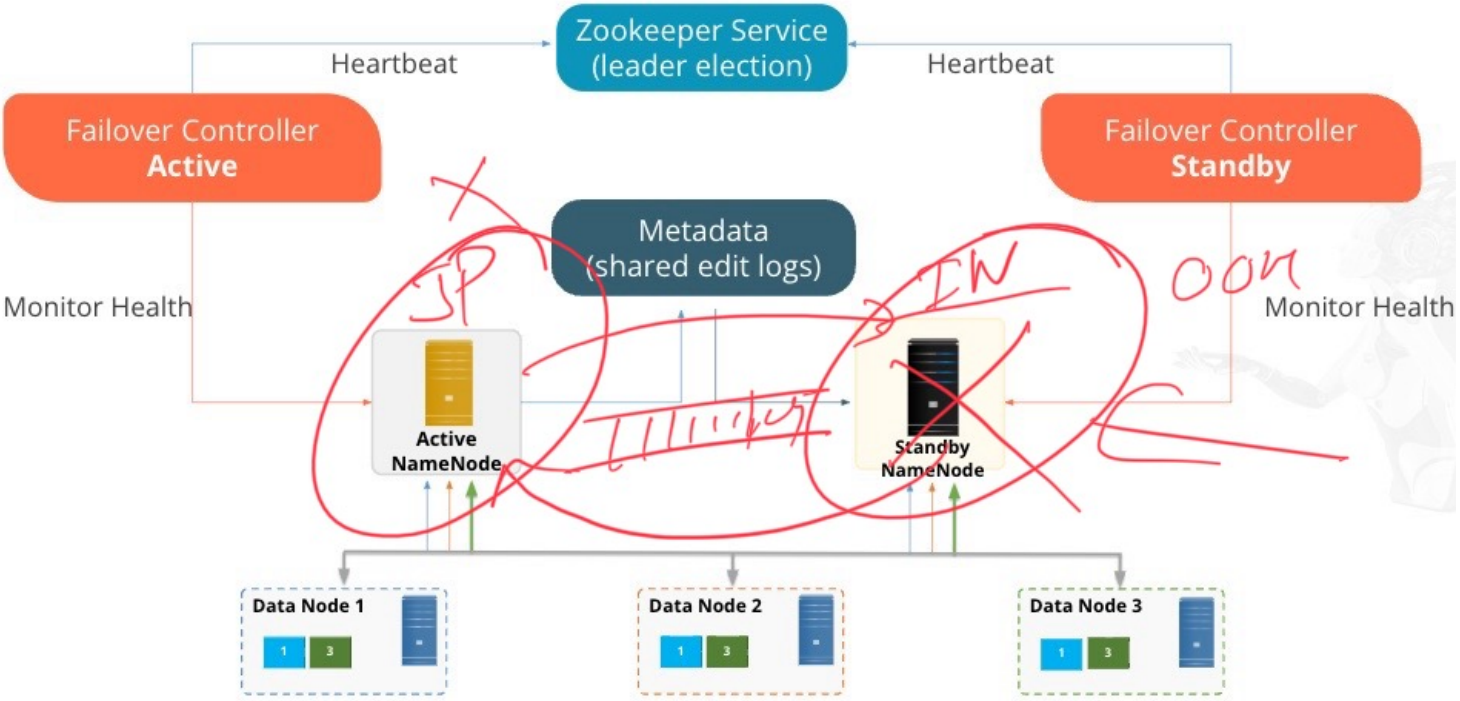
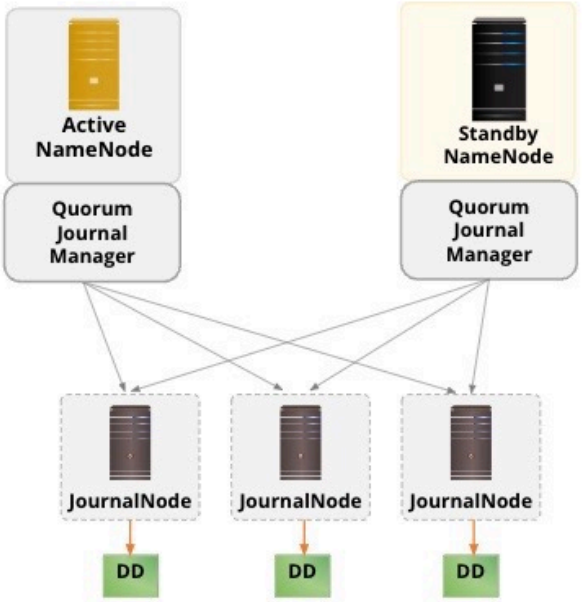
Understanding Big Data Architecture
HDFS in details
YARN

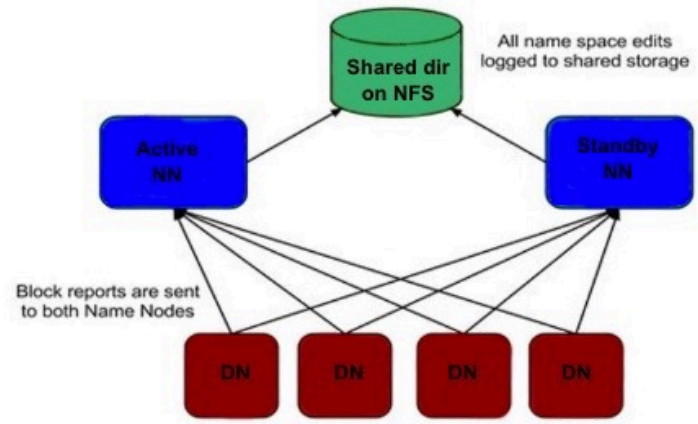


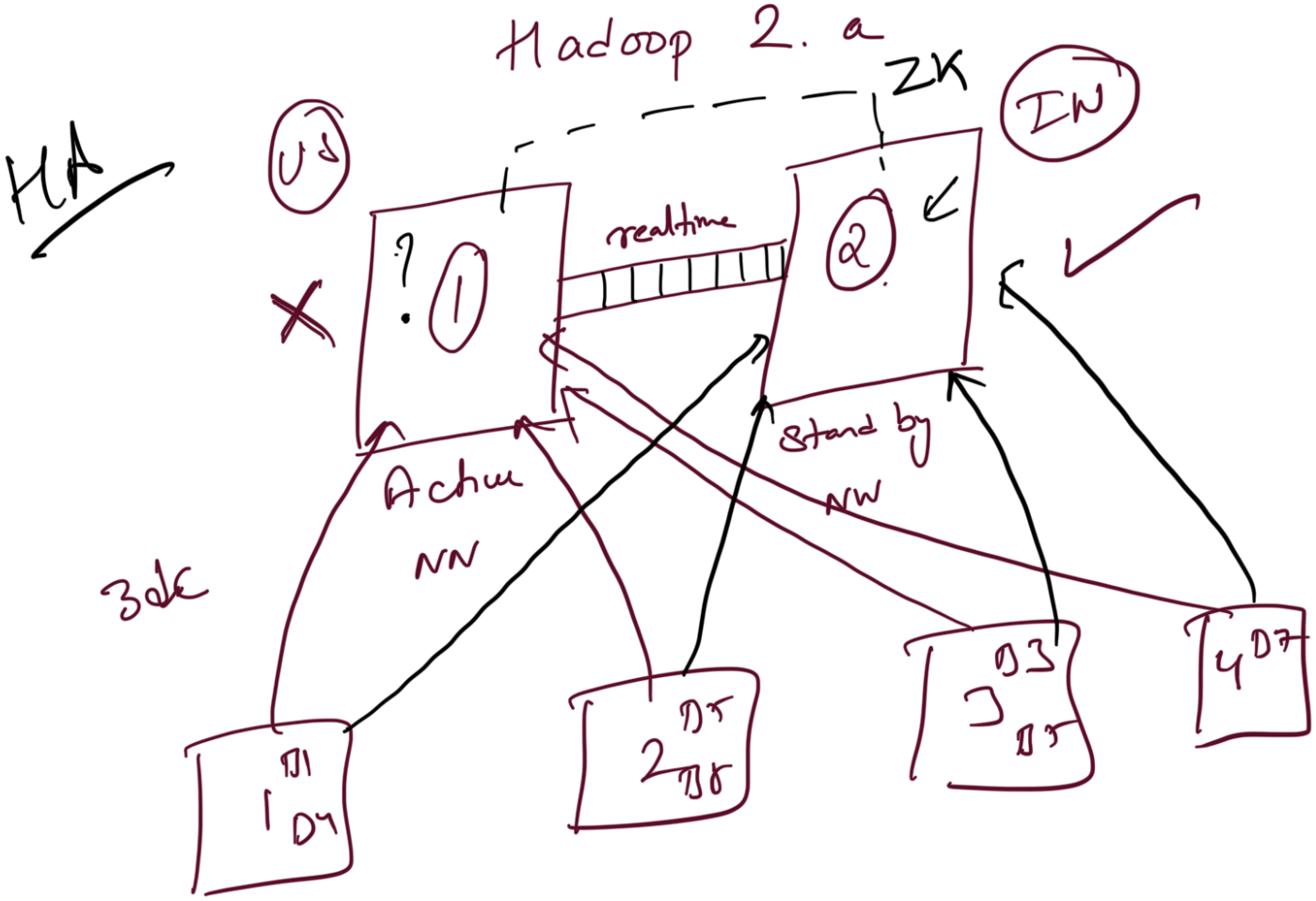


Quorum-based Storage



Shared storage using NFS



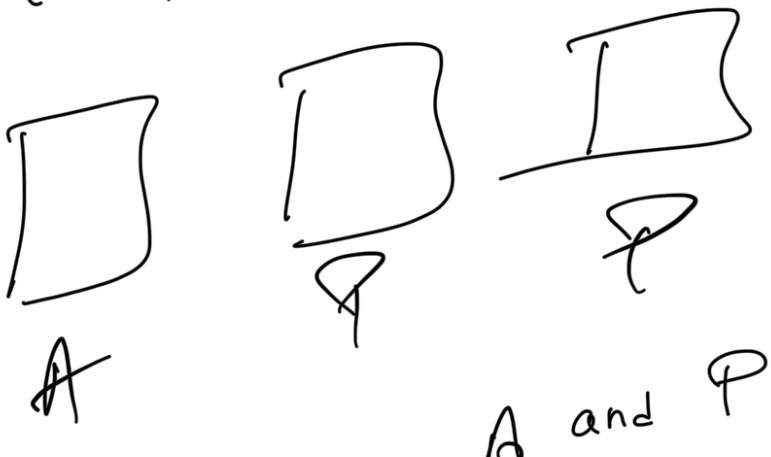


Topology

↳ IP1 ⇒ Masts

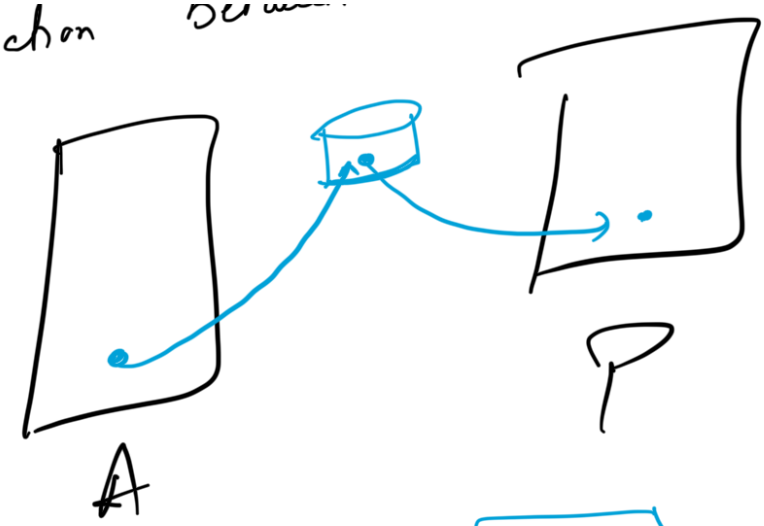
↳ IP2 ⇒ Stand by

Hadoop 3.x

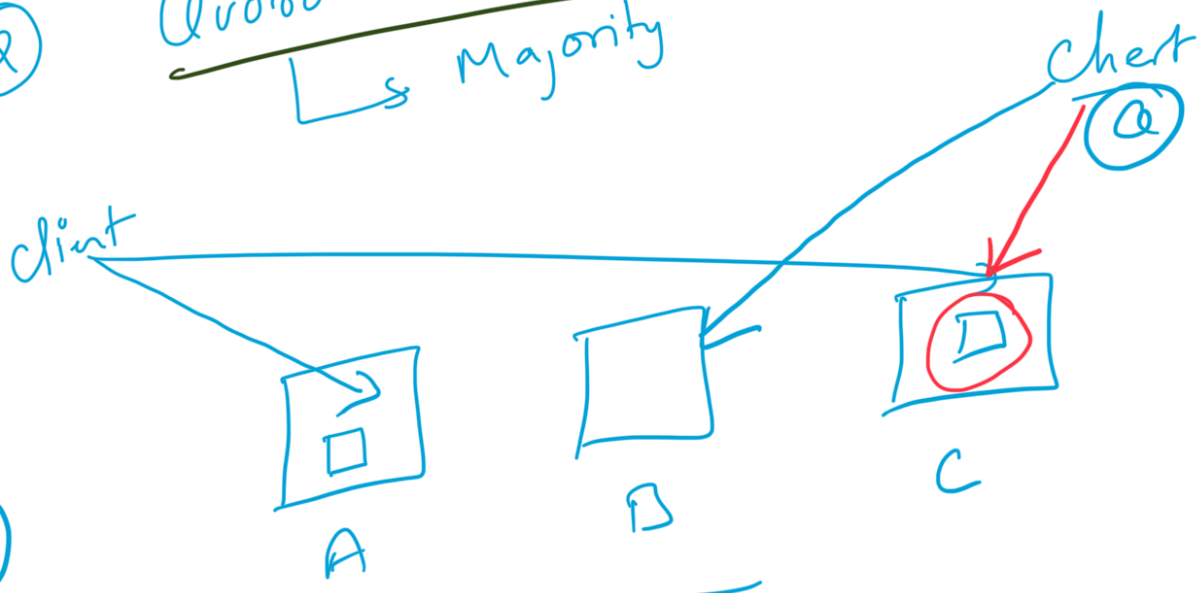


A and P

⇒ Communication



1. Shared Storage using NFS
2. Quorum Based Storage
   - Majority



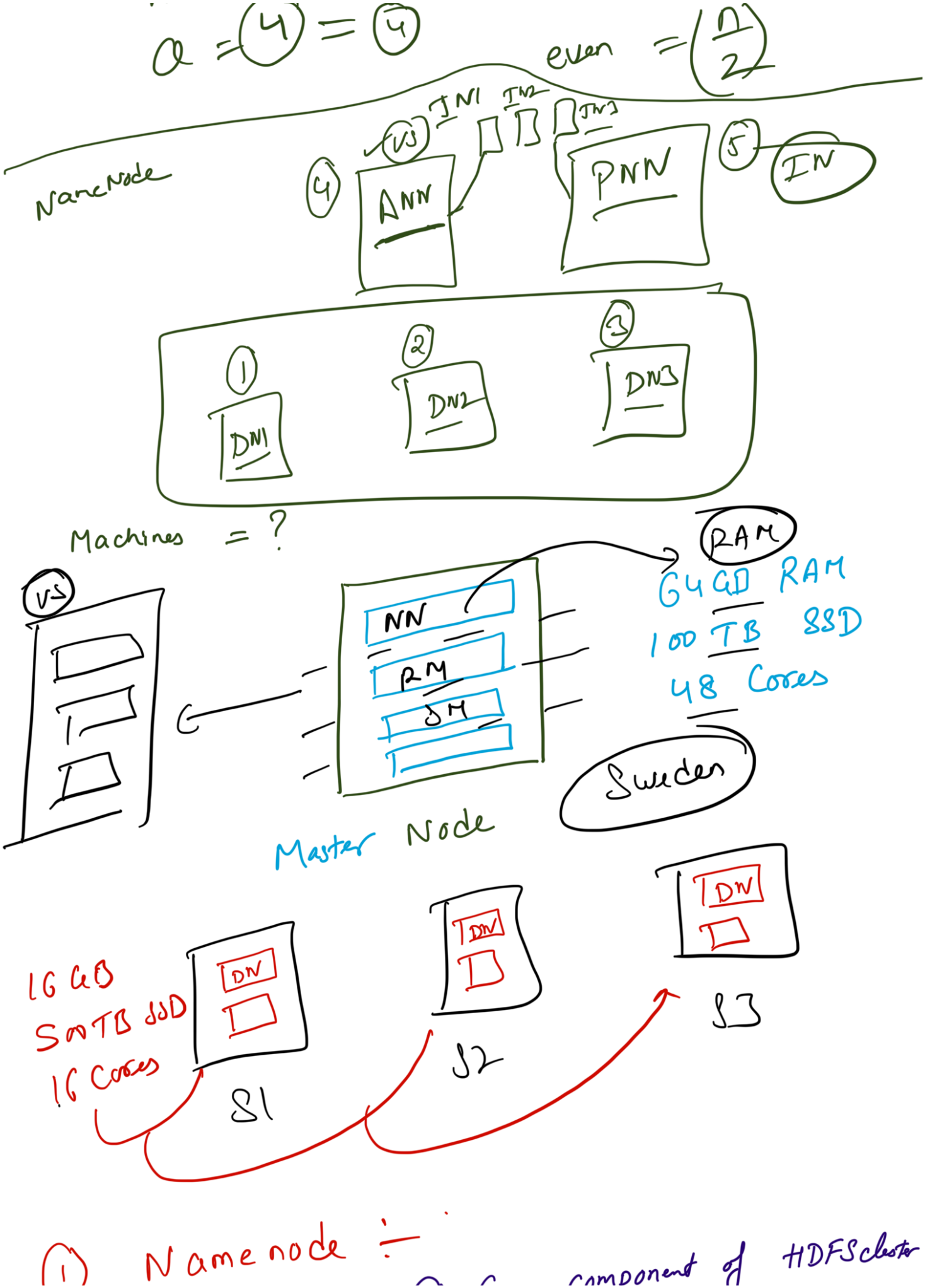
write

$$Q = \frac{n+1}{2} \Rightarrow 2$$

QJM
- replicate data into majority of nodes

$N = 7 \rightarrow 8$ (1W)

$$odd = \left(\frac{n+1}{2}\right)$$

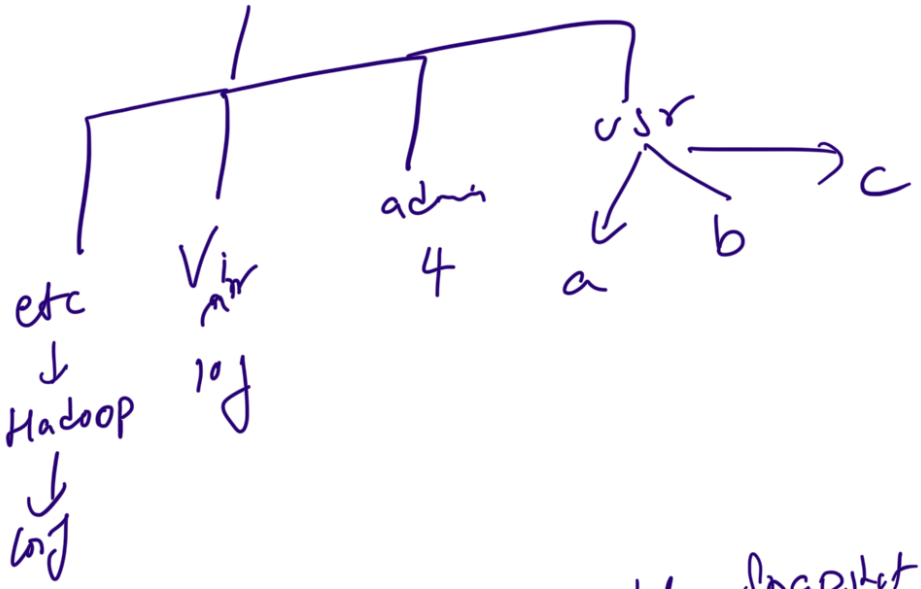


(1) Namenode :- component of HDFS cluster

(a) Core

(b) Stores metadata of each file that is being Created.

metadata? → FSImage
→ Editlogs



(1) FSImage :- store a Complete Snapshot of file System metadata

(2) Editlogs :- Sequential tan log that records every metadata changes in HDFS.

↓

File Creation, delete, Modified, Permission, file creation

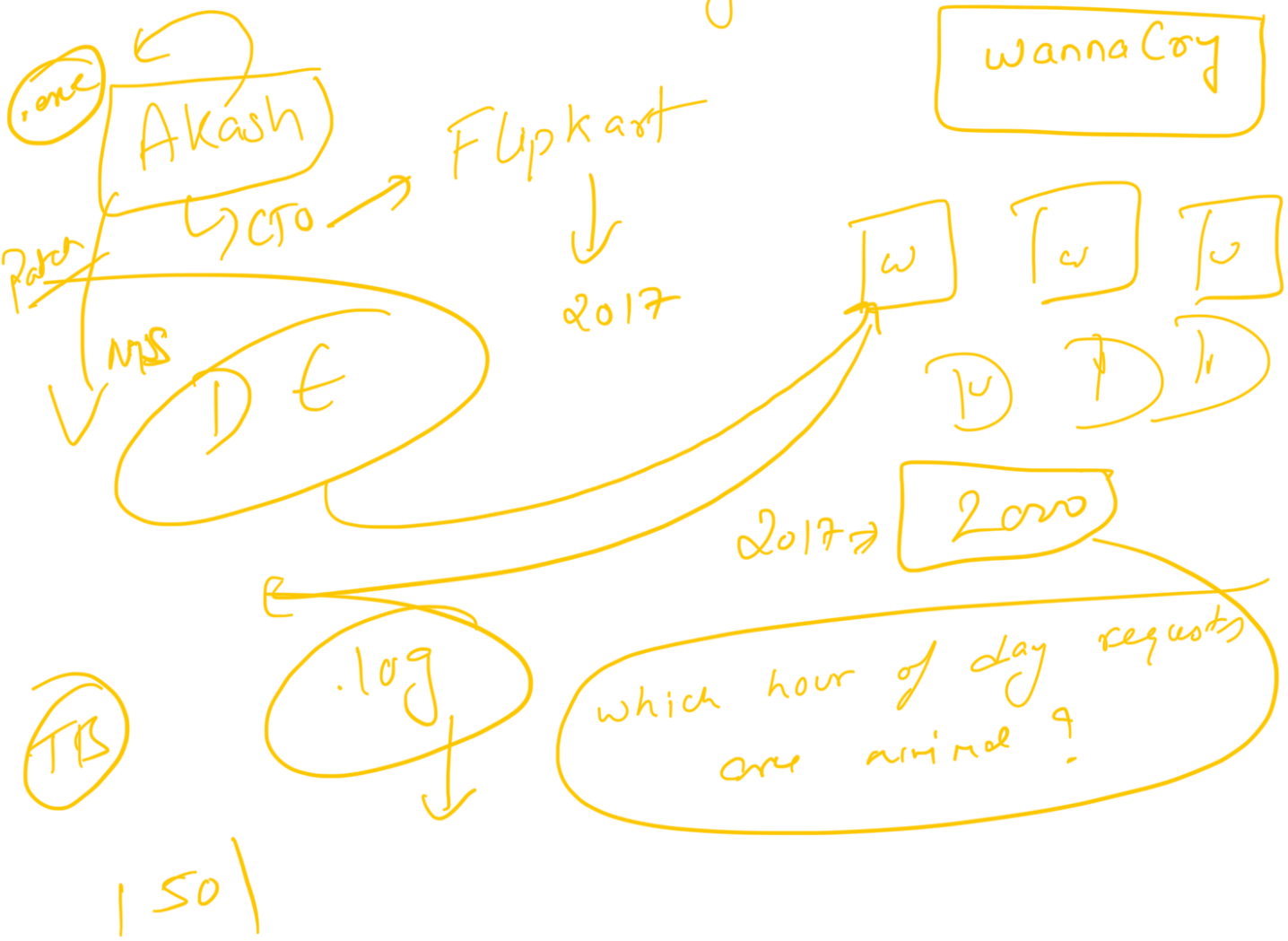
NN = FS + EL

NN =

(2) Datanode → are responsible for storing & maintaining the data blocks.

(3) Zookeeper :- to elect the leader in case of failure.

(4) JournalNode :- to keep passive NN in Sync with Achae NN.



150