

---

---

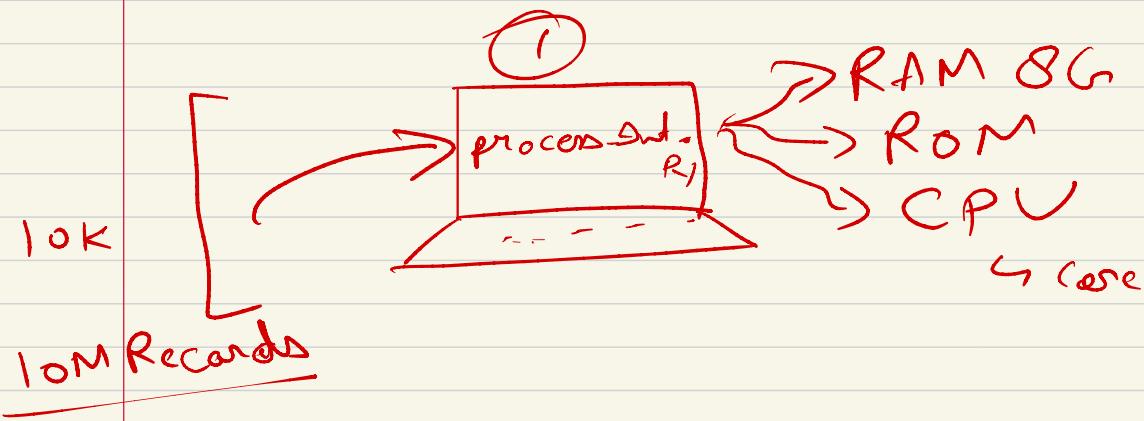
---

---

---

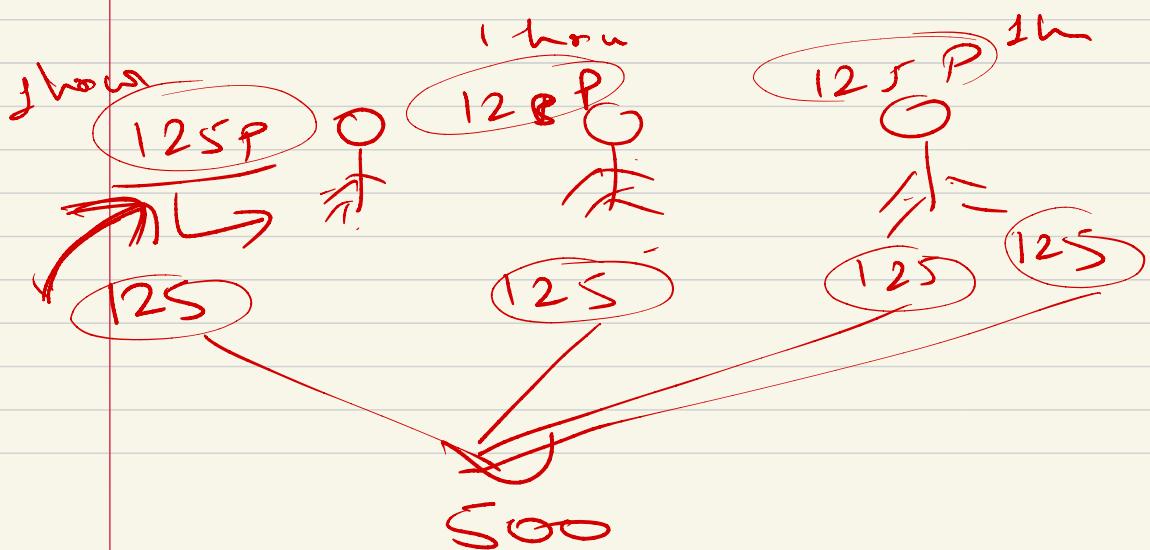


# → Distributed Computation

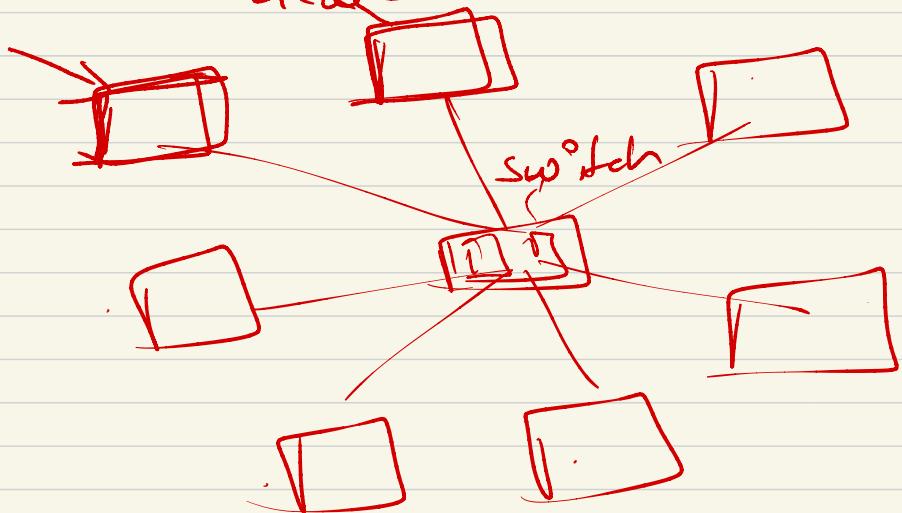


500 pages → Shows  
Input → OK  
Task 2 -  
Count total  
pages in the  
book

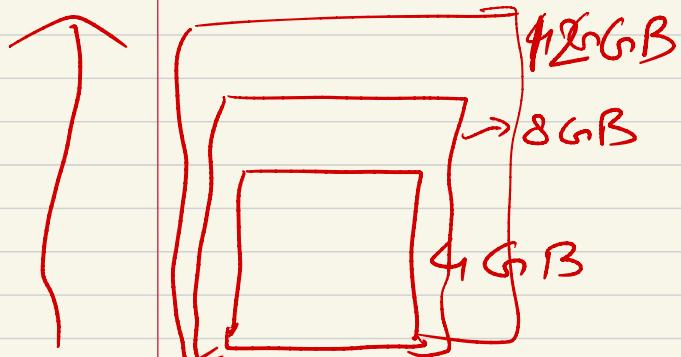
125P      1 hour



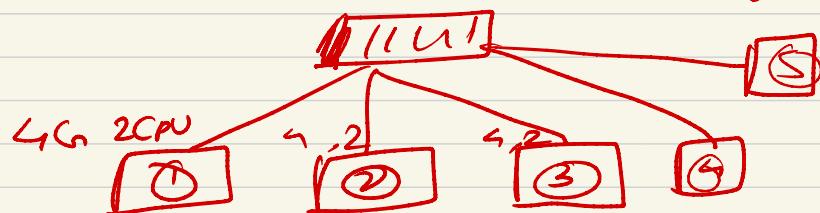
## 8GB RAM 4Core CPU Cluster



## Vertical Scaling

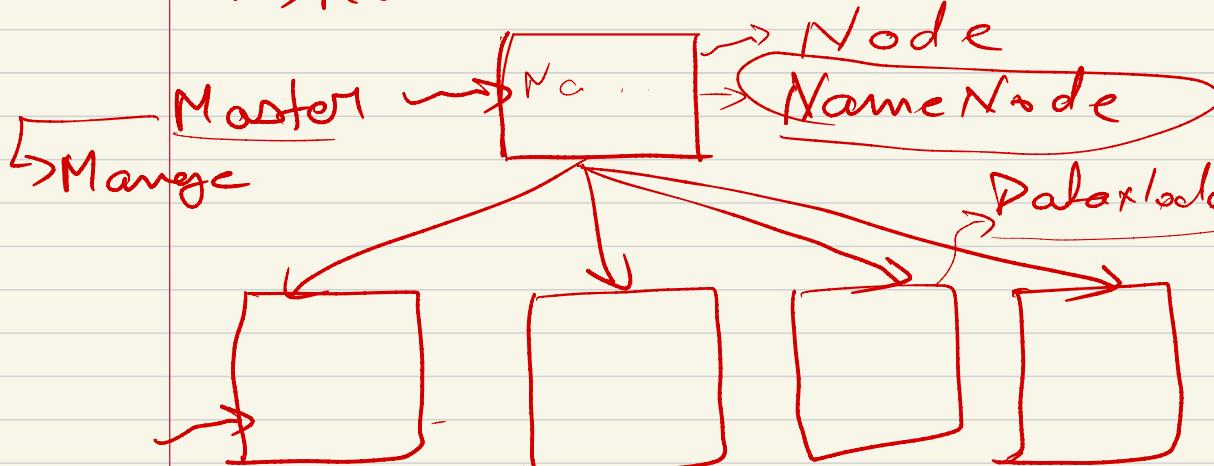


→ Horizontal Scaling



# Hadoop Cluster

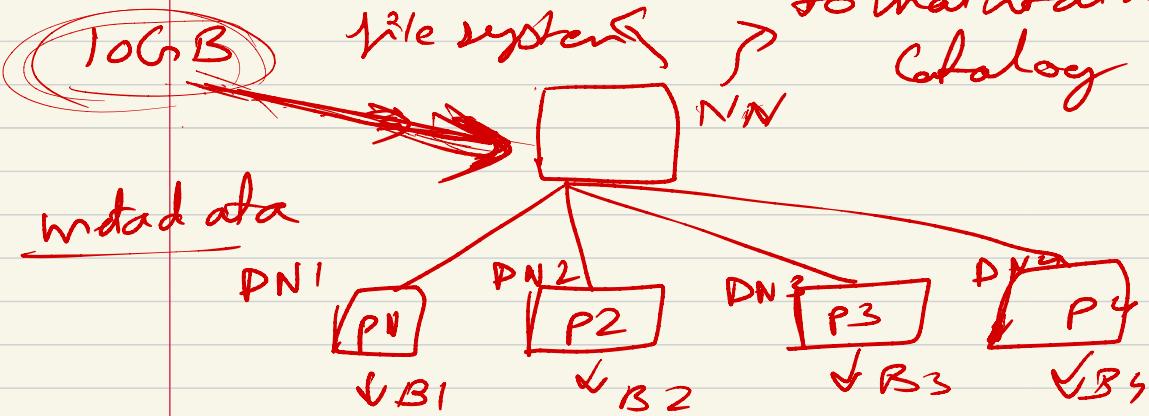
↳ Master - Slave Arch



## Slave

- ↳ Store distributed data
- ↳ Process local data

metadata of file system → maintain catalog



=> cp source-file dest-file

/home/shashank/Desktop/file.txt

syntax: hdfs or hadoop fs

▷ hdfs src-file destination

▷ hdfs -put src-file destination

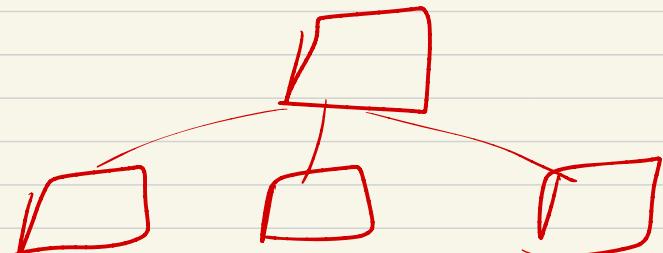
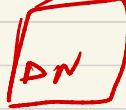
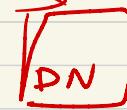
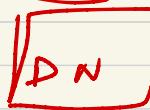
Metadata →  
start Namenode



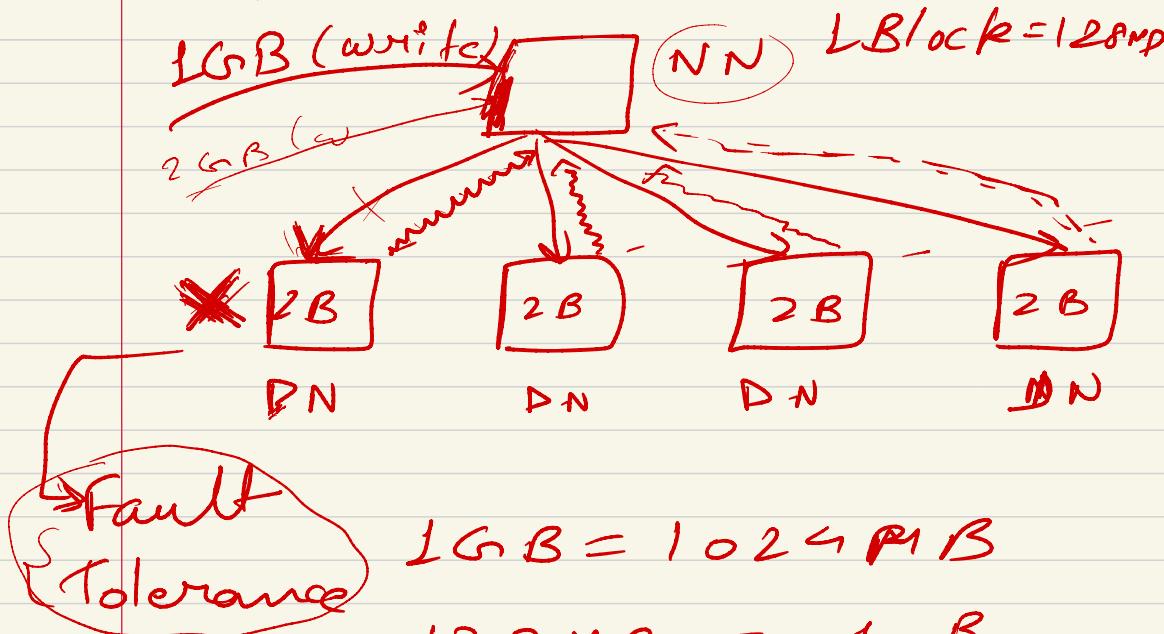
Namenode → Data Service



Data nodes  
service DN



① Block: Small unit of data to store. 128 MB



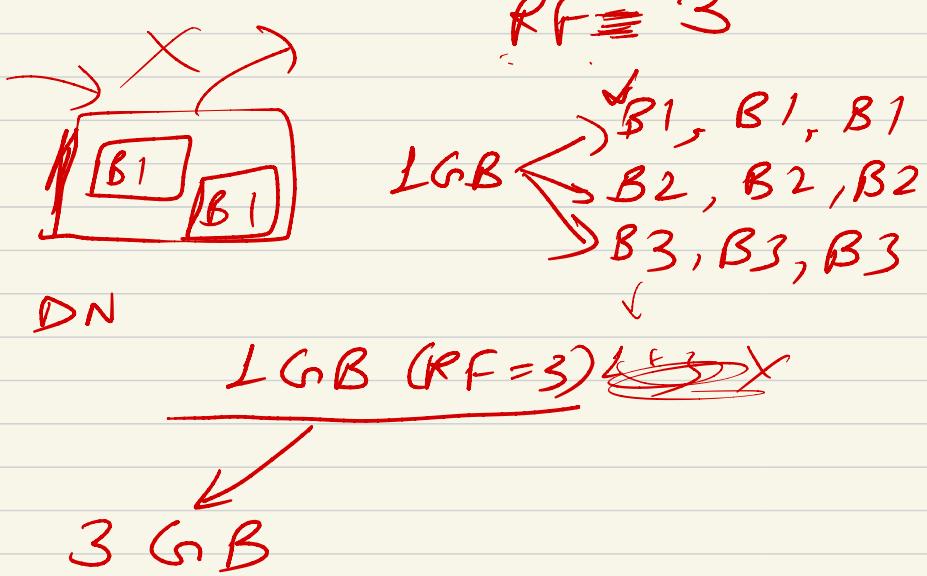
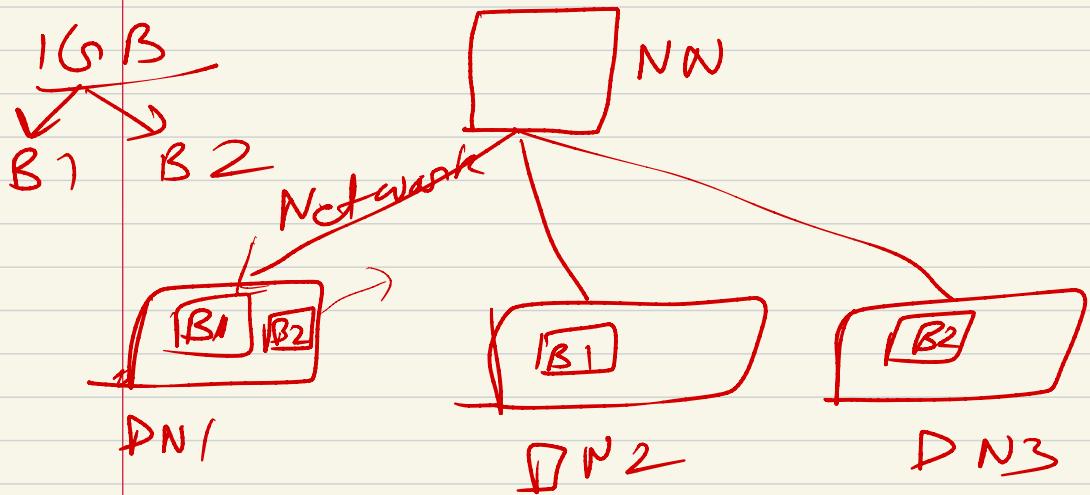
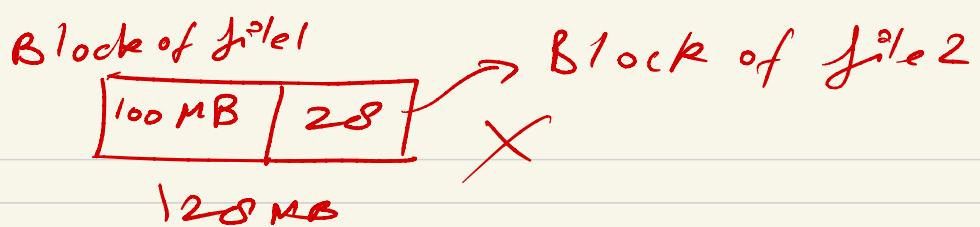
$$1\text{GB} = 1024 \text{ MB}$$

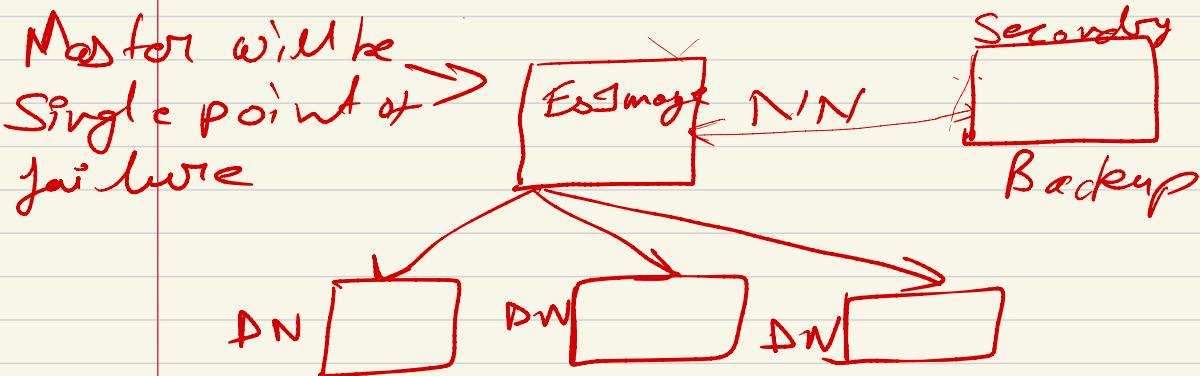
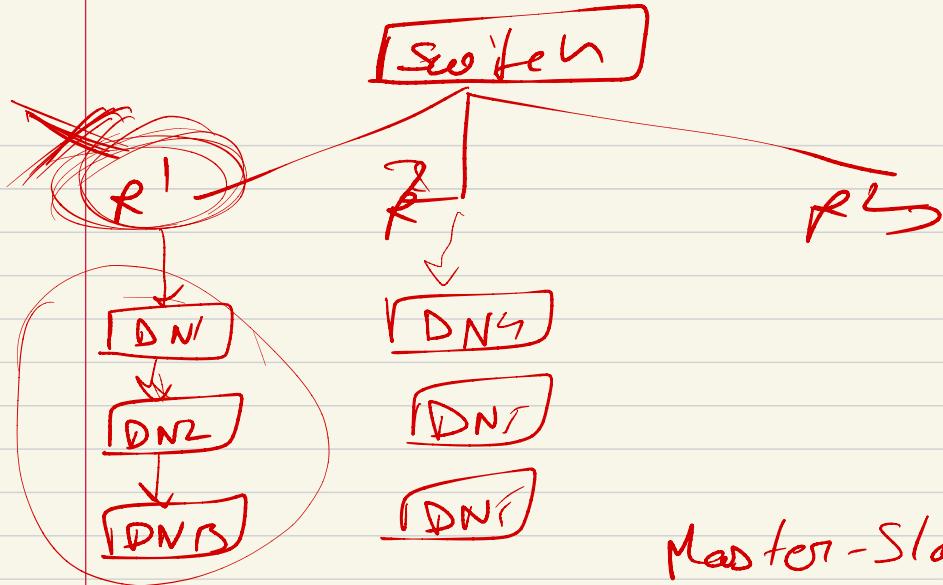
$$128 \text{ MB} = 1 \text{ B}$$

$$1 \text{ MB} = \frac{1}{128} \text{ B}$$

$$1024 \text{ MB} = \frac{1024}{128}$$

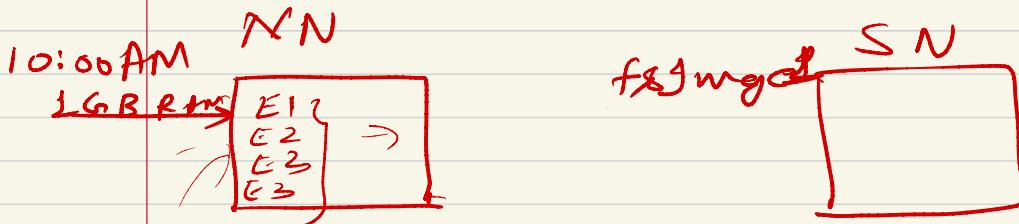
= 8 Blocks



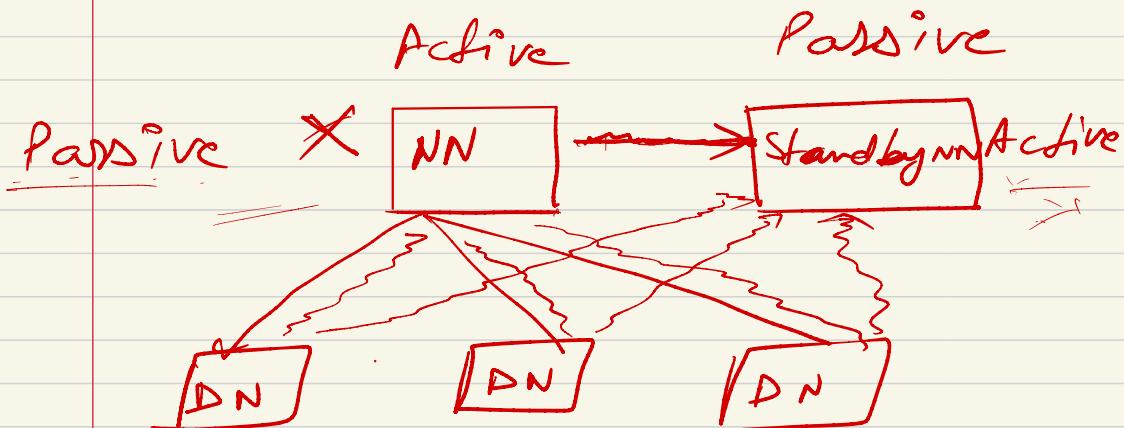
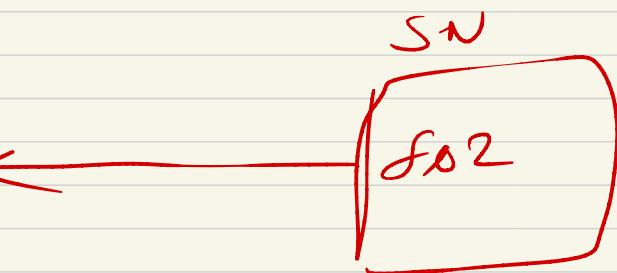
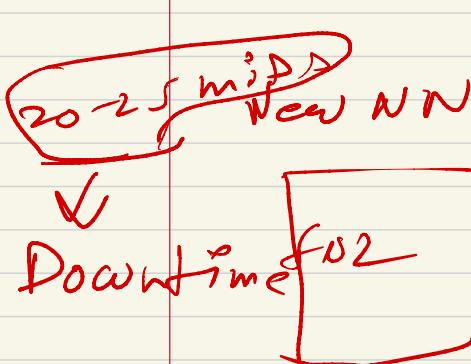
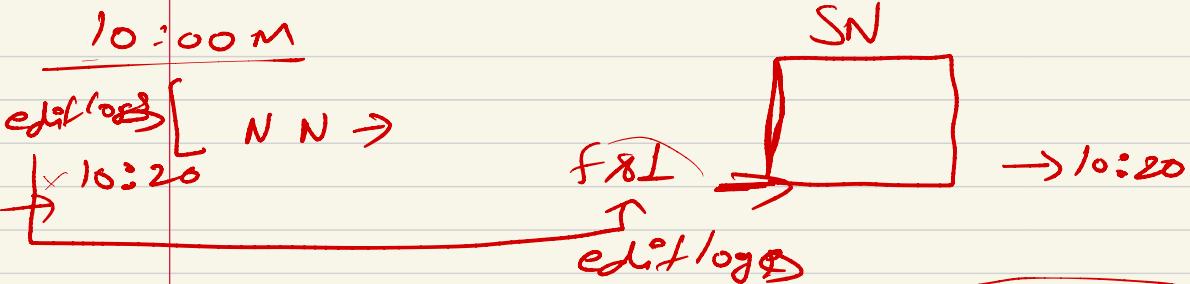


### Secondary NameNode

↳ Checkpointing of fsImage



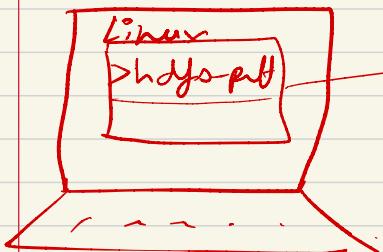
Checkpoint Window  $\Rightarrow$  20 sec



hdjs - put src dest

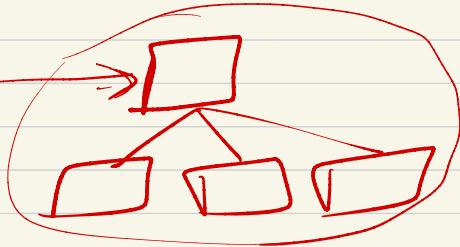
> cat filename

> hdfs -cat filename



client

connect



cluster