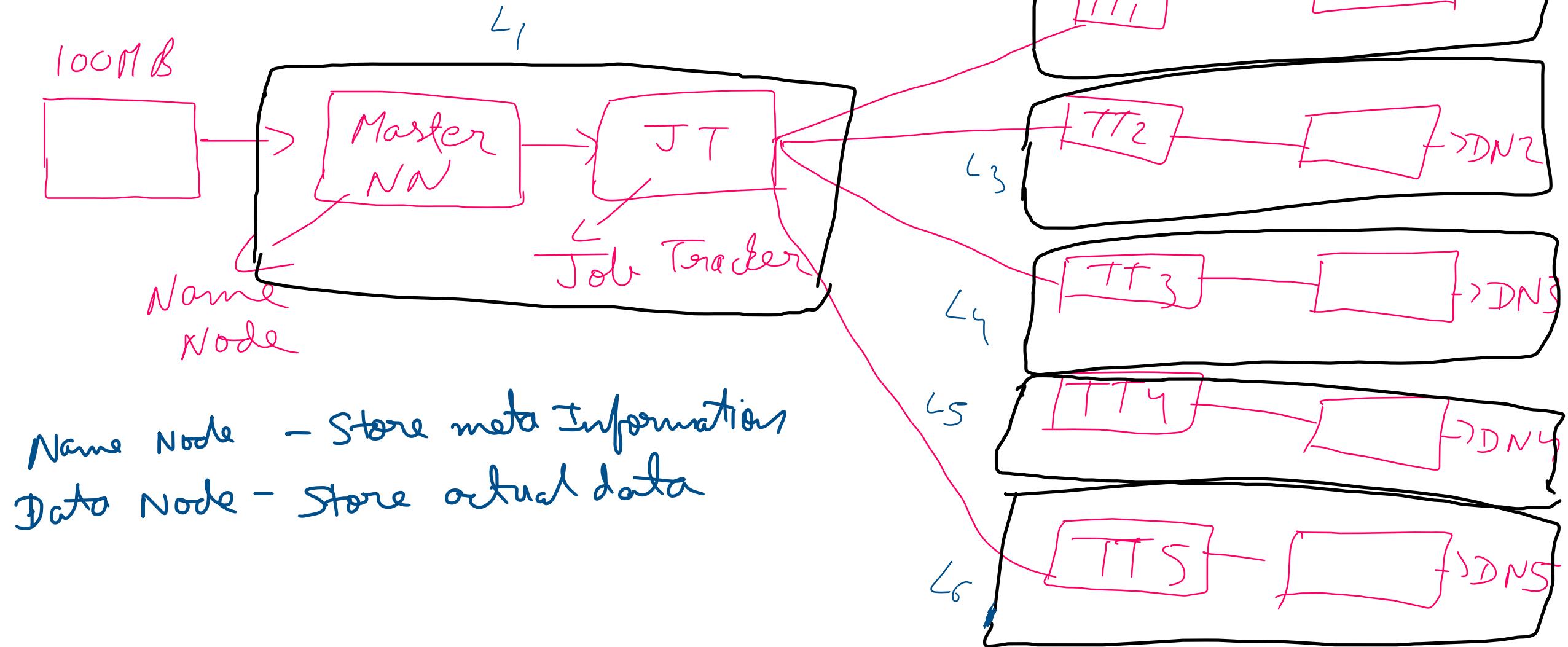


# Hadoop 1.X Architecture



Job Tracker - work allocation done by Job Tracker and track all data node what is going there.

Task Tracker - only track one Data node

1000MB

(divide entire data into small blocks)

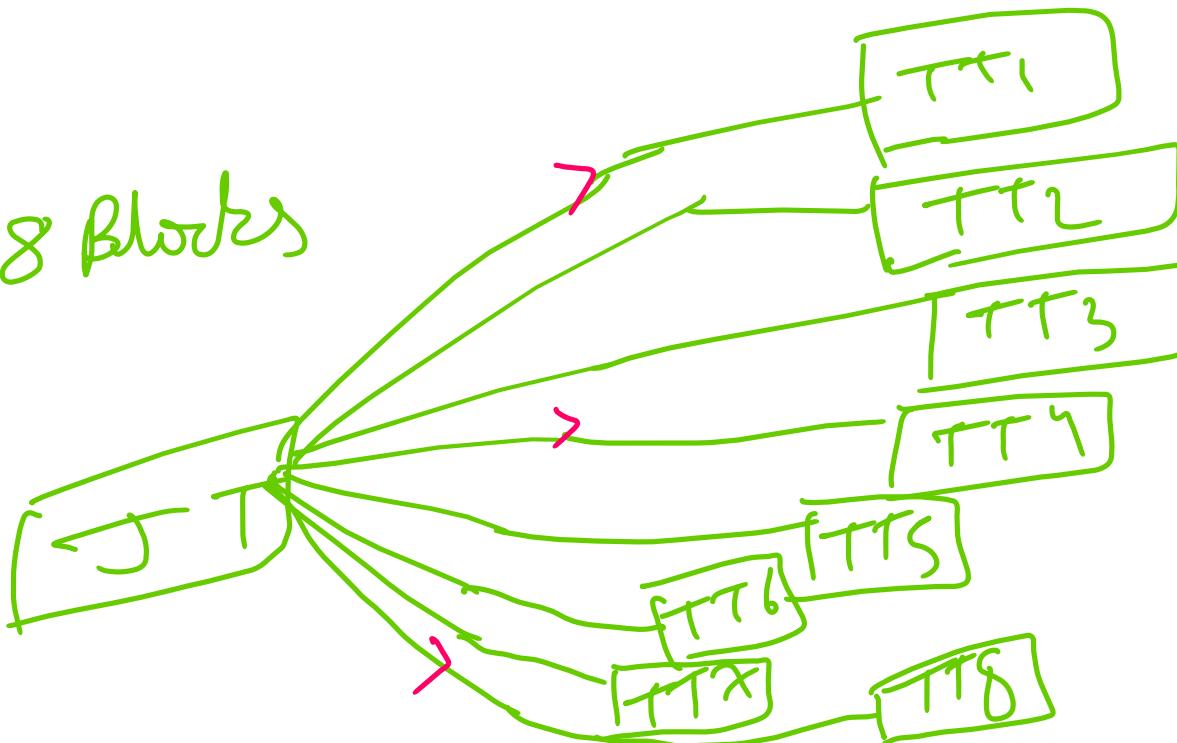
128

128

128

128

⇒ 8 Blocks

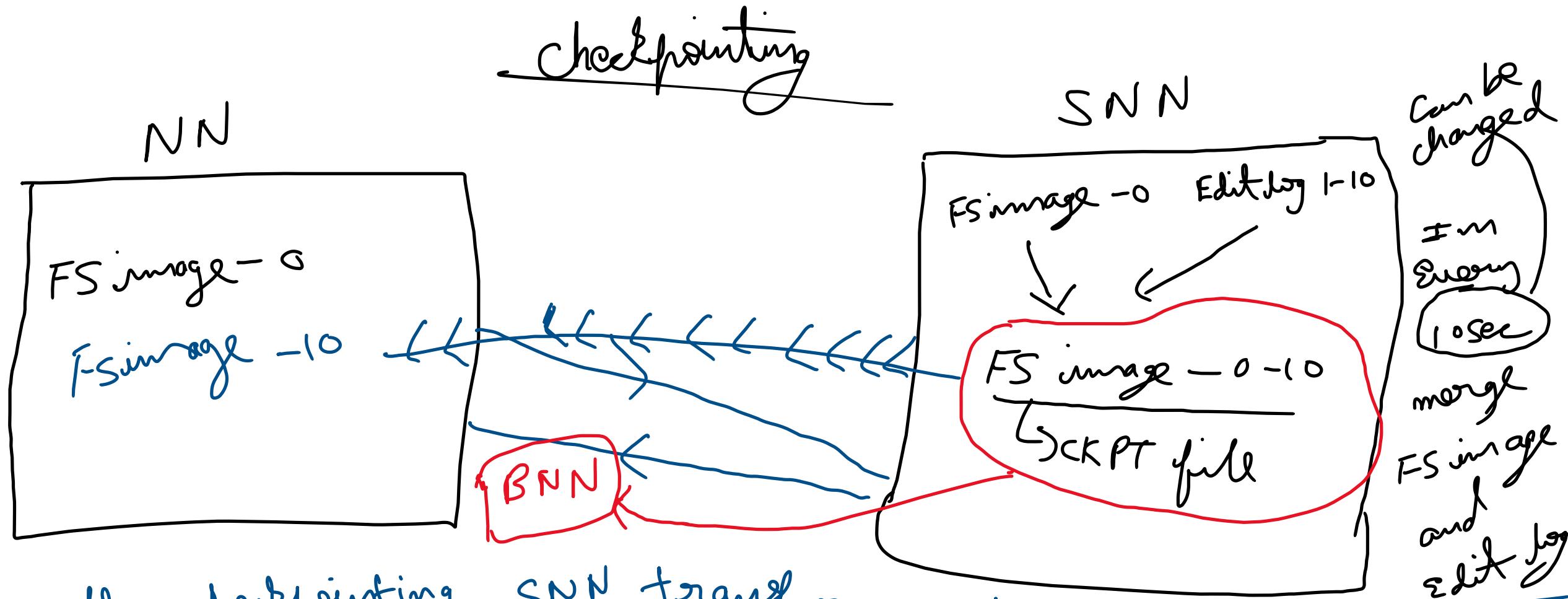


knows which DN is available or not and assigns task accordingly

→ available DN

# Secondary Name Node

FS = File System

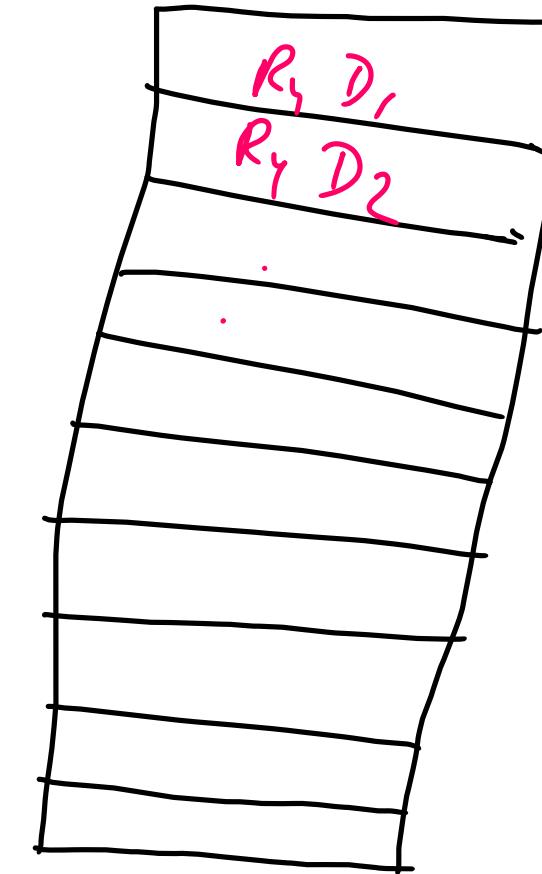
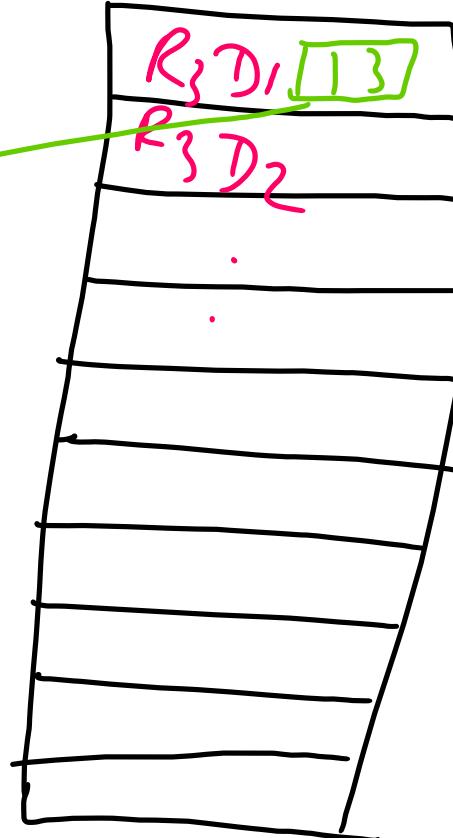
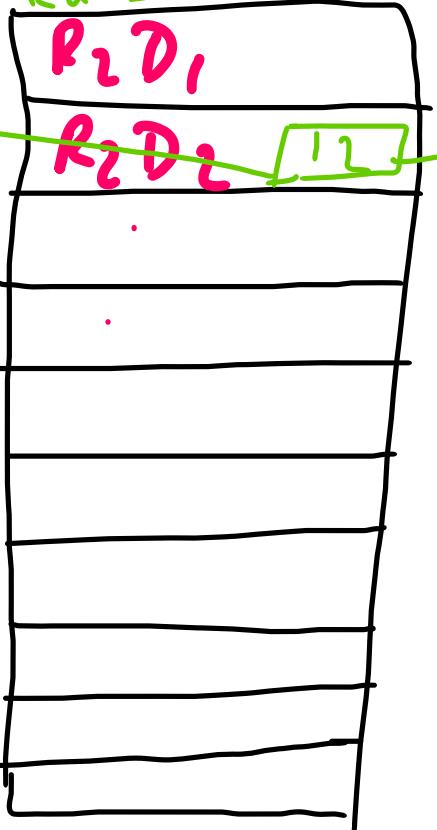
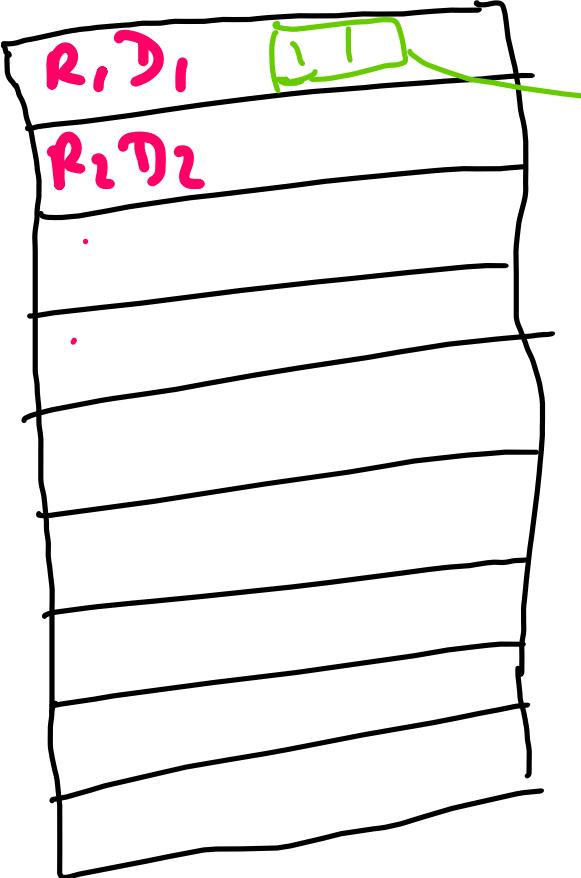


after checkpointing, SNN transfer FS image to NN in equal interval of time. SNN first check whether NN is free or not.

SNN : helps NN To create checkpoint CKpt file by merging FS image and edit log which are stored and buffer for Name Node.

How to manage System failure how replica replacement happens?

→ Block Number 1, replica 1



$R_1$

$R_2$

$R_3$

$R_4$

## Problems in Hadoop 1.X

Job tracker was responsible for multiple things

- 1) Resource Management
  - 2) Job Scheduling
  - 3) Job Monitoring
- 
- ```
graph TD; 1[1] --> RM[Resource Management]; 2[2] --> JS[Job Scheduling]; 3[3] --> JM[Job Monitoring]
```

Hadoop 2.x → YARN

YARN - Yet Another Resource Negotiator

