

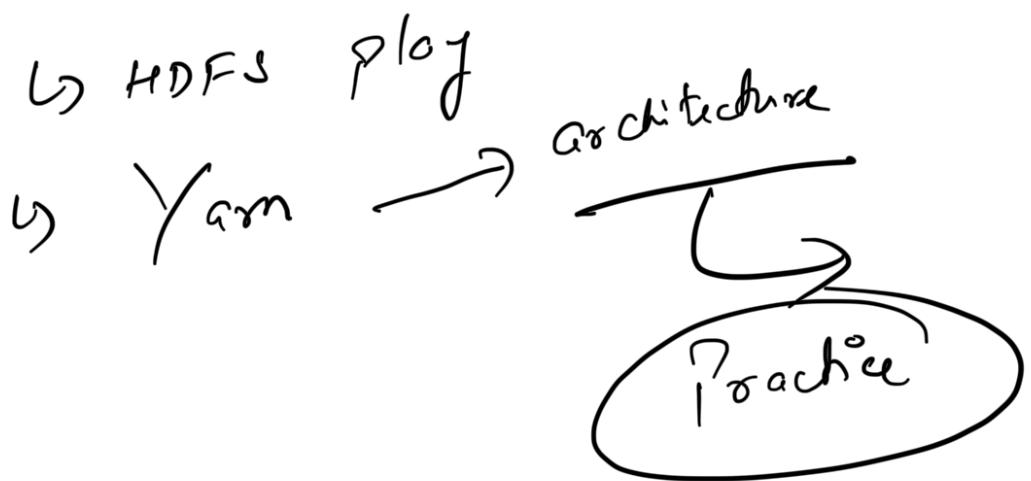
Agenda

HDFS demo

Yarn

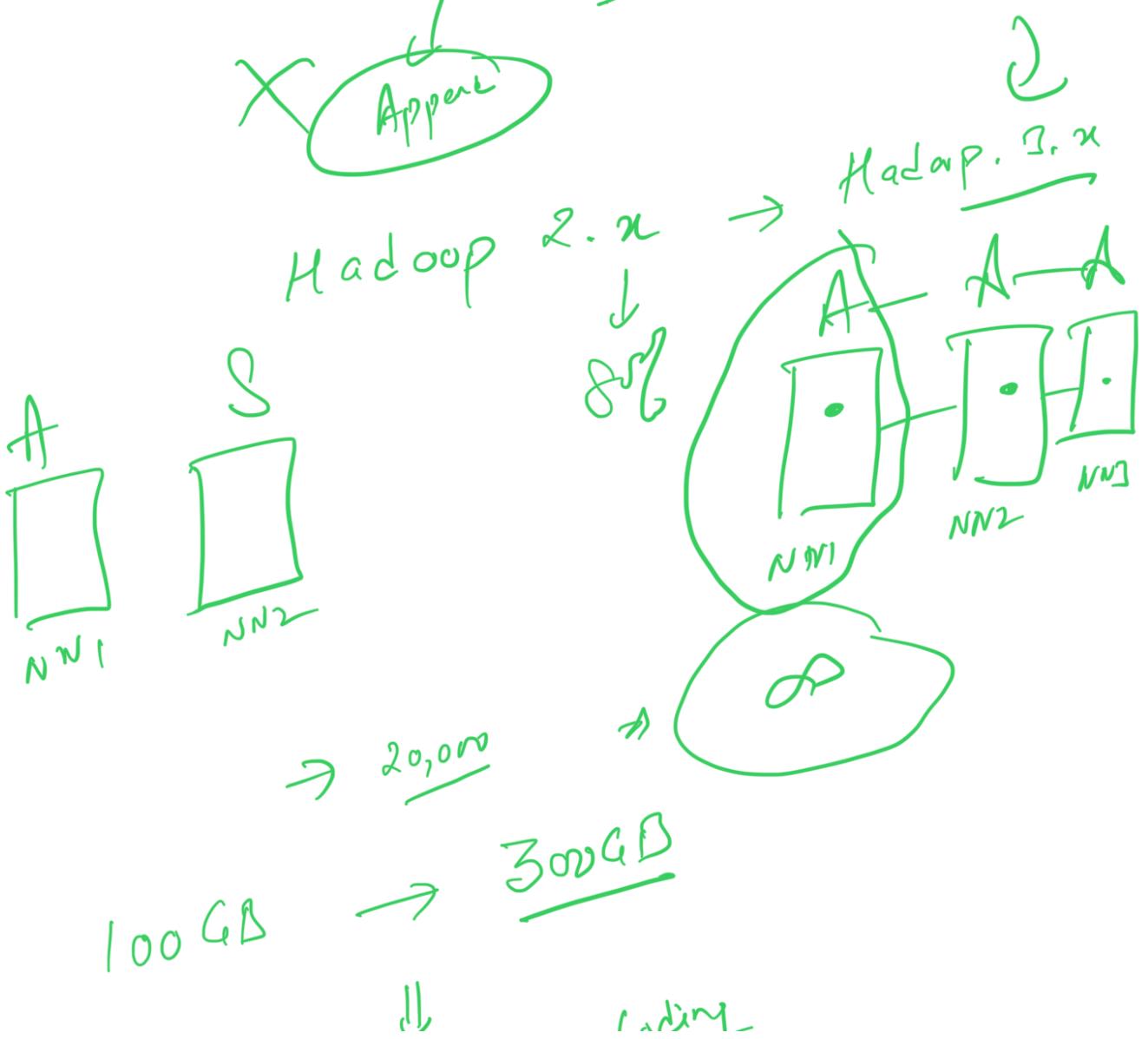


Credit: IMDB

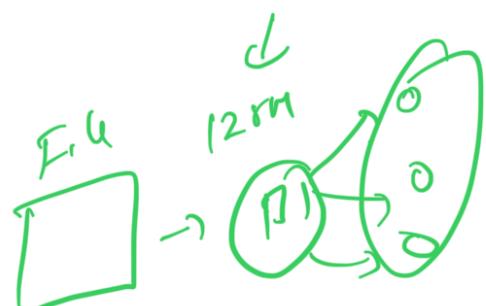


HDFS = WORM

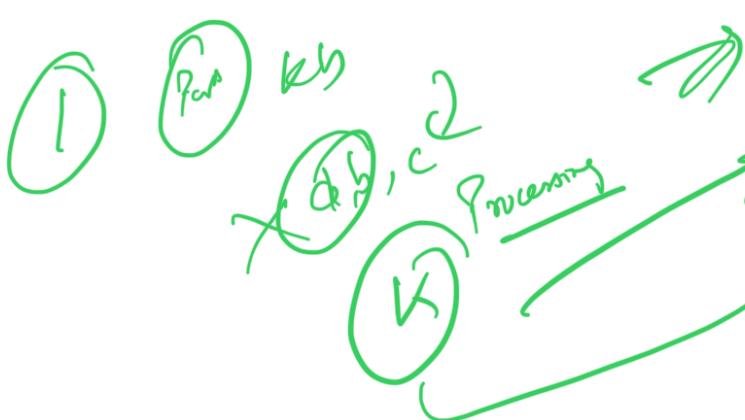
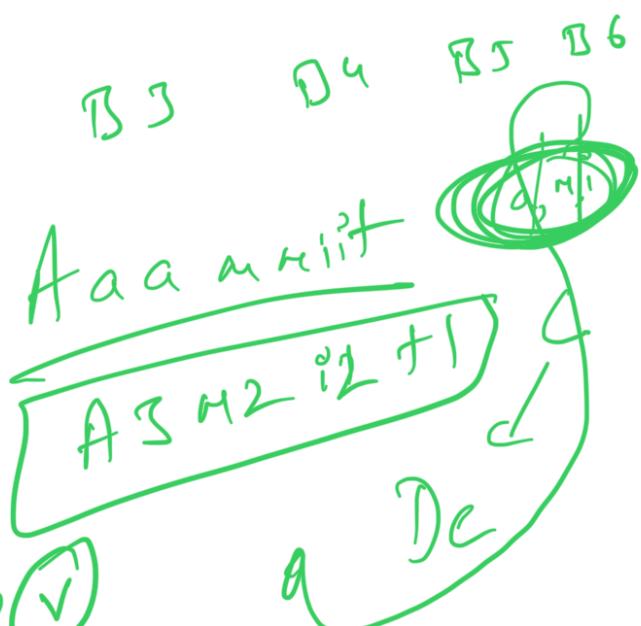
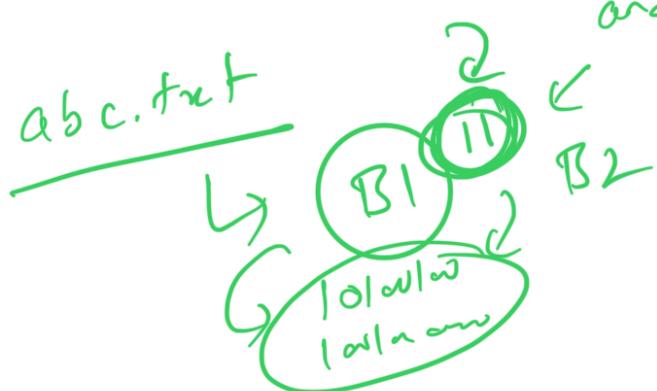
↳ Write Once Read Many



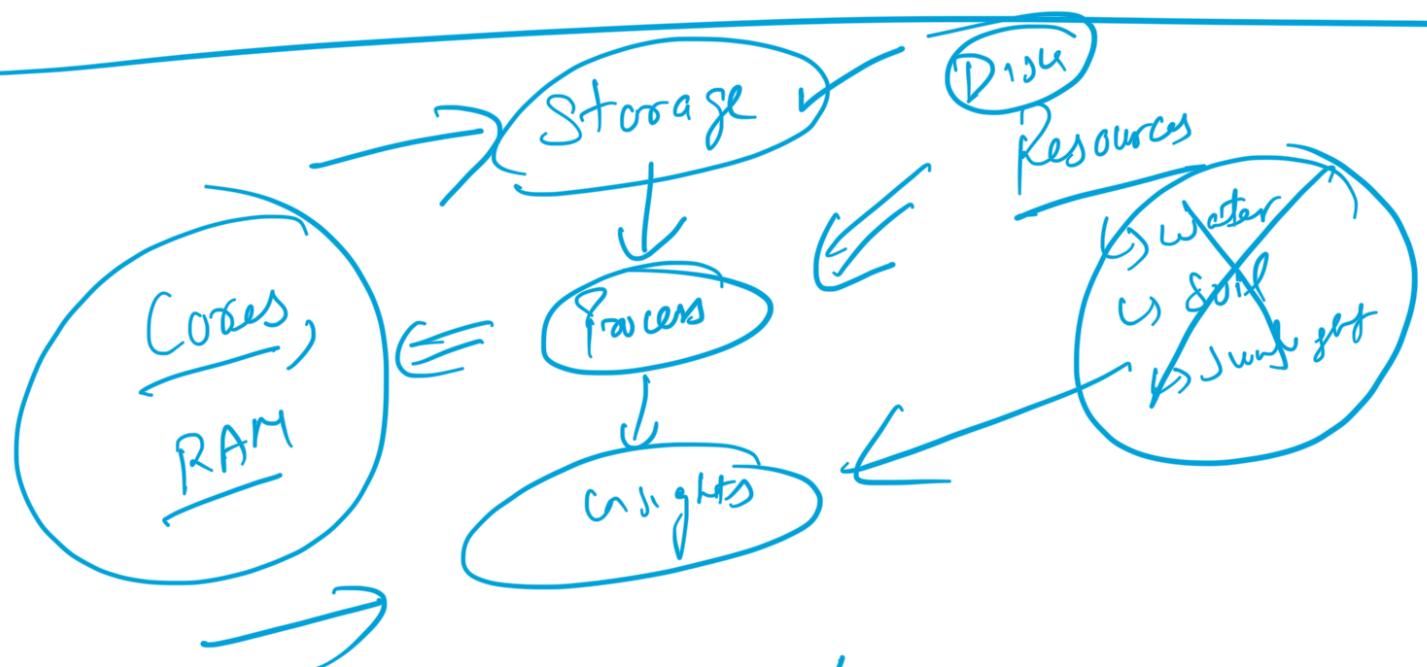
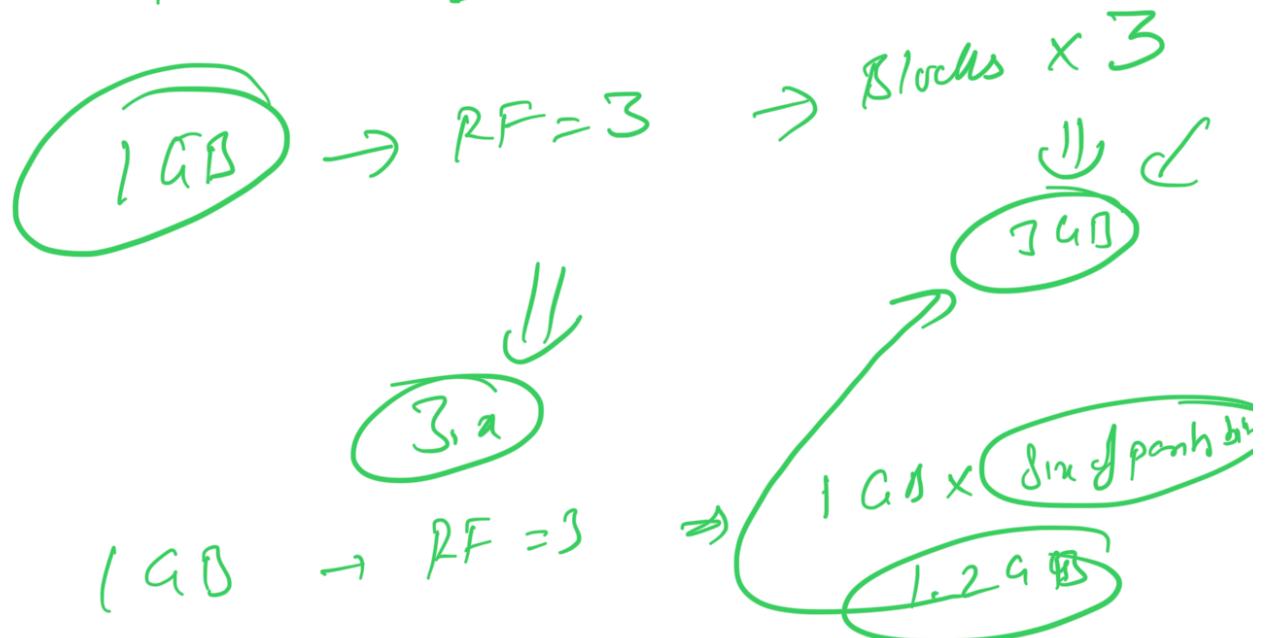
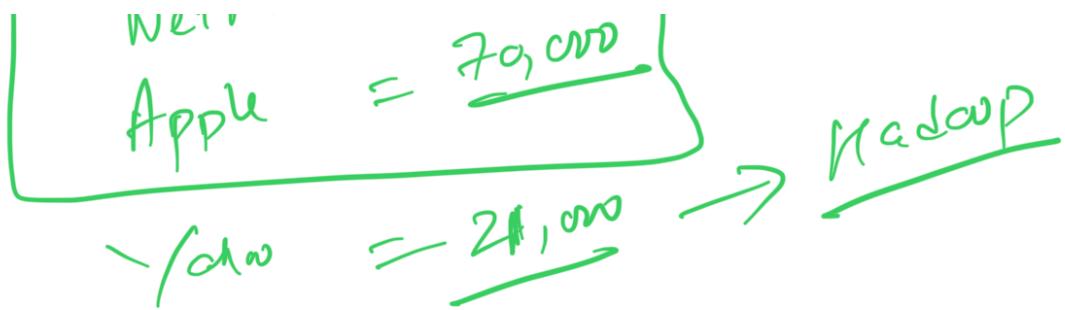
Erasure ^{concern} of
 → is a data protection method that
 divides the data into fragments
 expands it with redundant
 data pieces and stores it
 across diff locations.



Coding Policy | RS-G3-1024K
 → splits data into 6 data blocks
 and 3 parity blocks



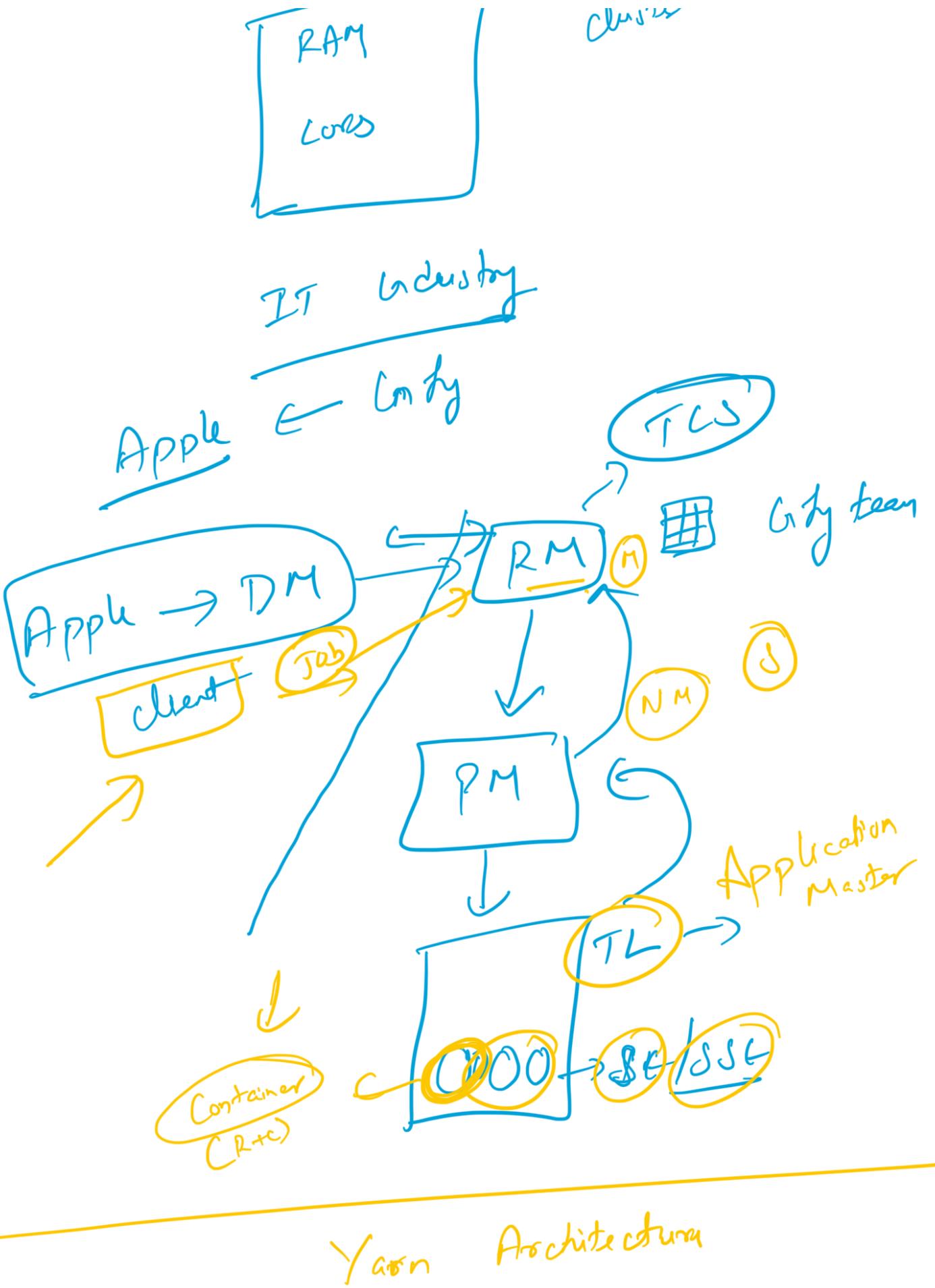
$$n \cdot t \cdot f \cdot h = 40,000$$



YARN

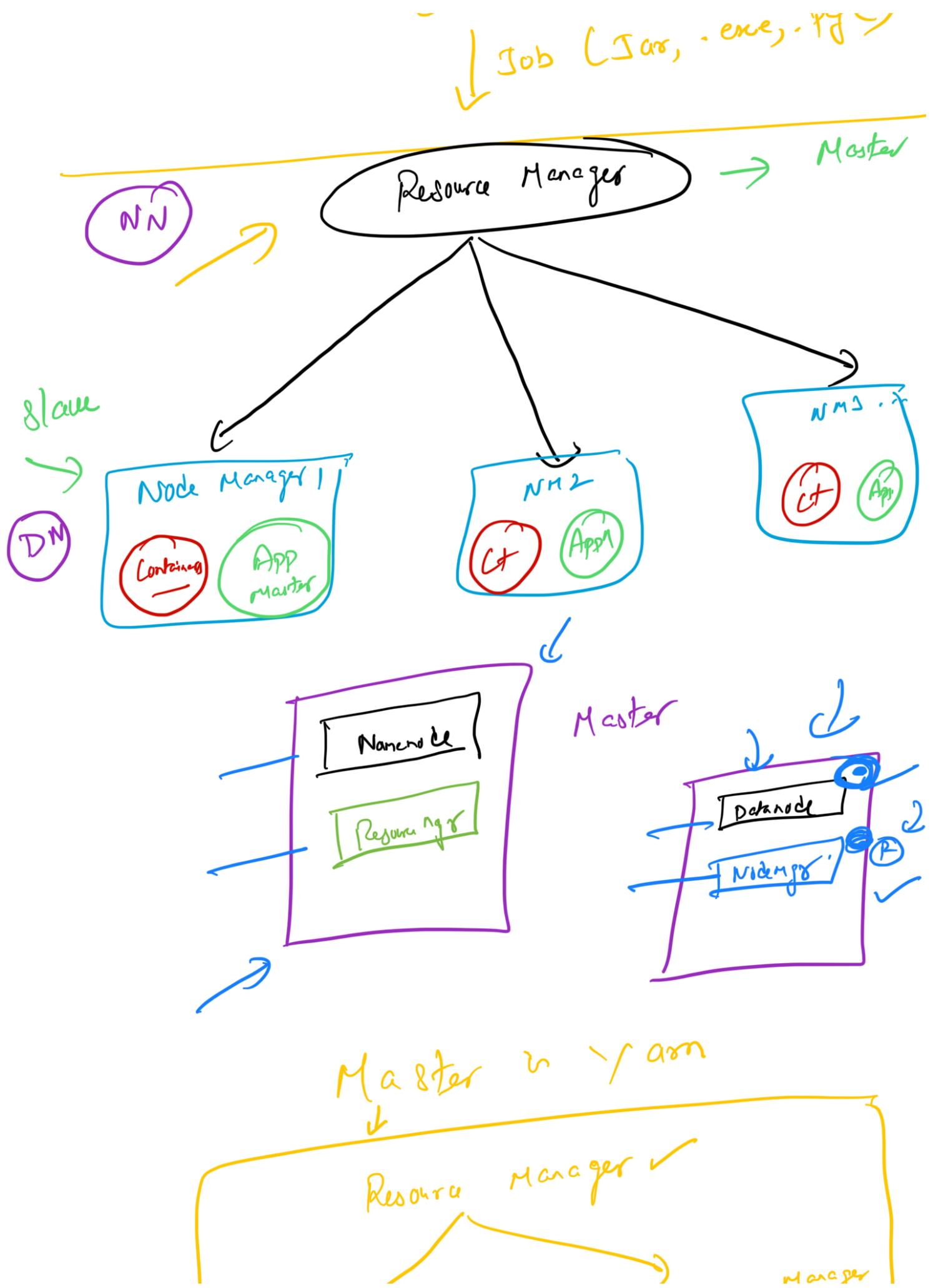
↳ Yet another Resource Negotiator

... +



client

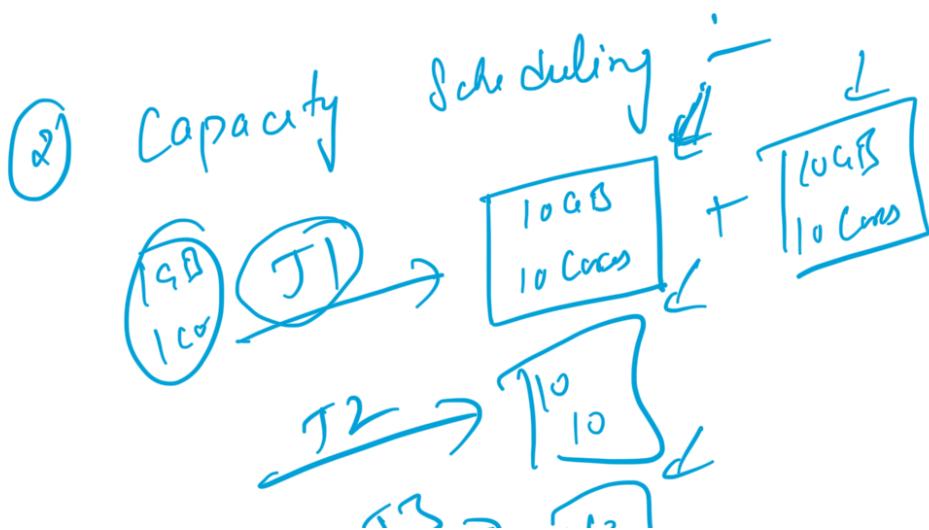
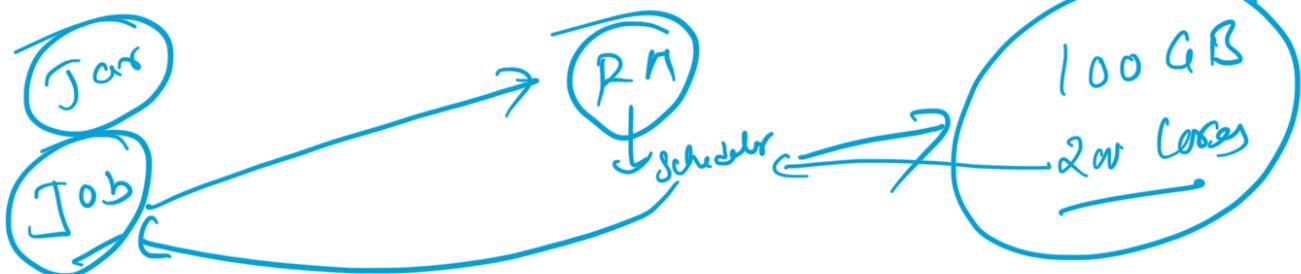
and



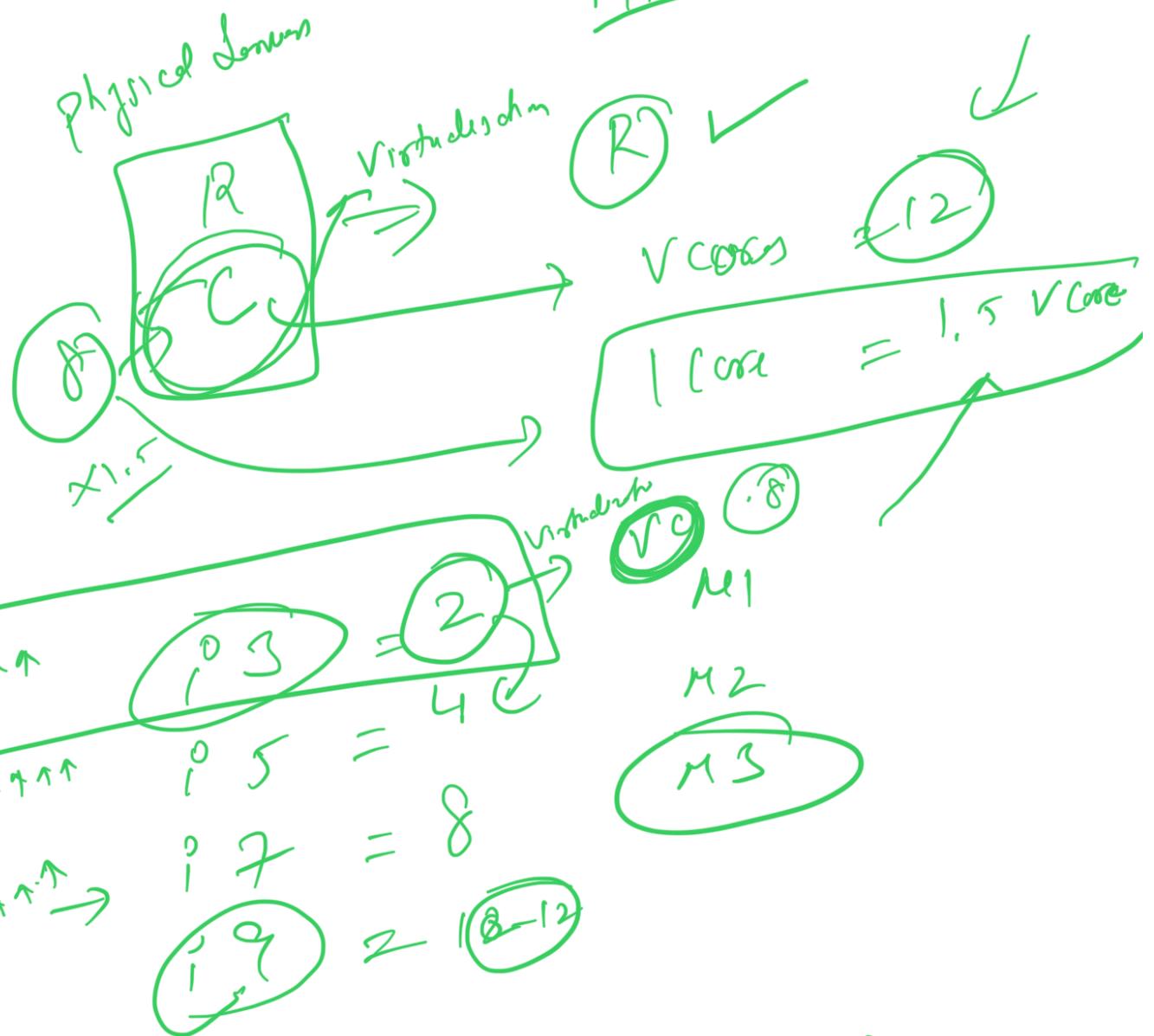


→ Scheduler :-

- ↳ does not monitor or track the status of application
- ↳ NOT Scheduler does not restart failed tasks.
- ↳ has a policy plugin to partition cluster resources
- ↳ FIFO Scheduler
Capacity Scheduler
Fair Scheduler



③ Fair scheduling : 



① RM Scheduler

② Application

$\xrightarrow{\text{accept}}$

↳ accept jobs

↳ get the resources

↳ restart the APPMaster

