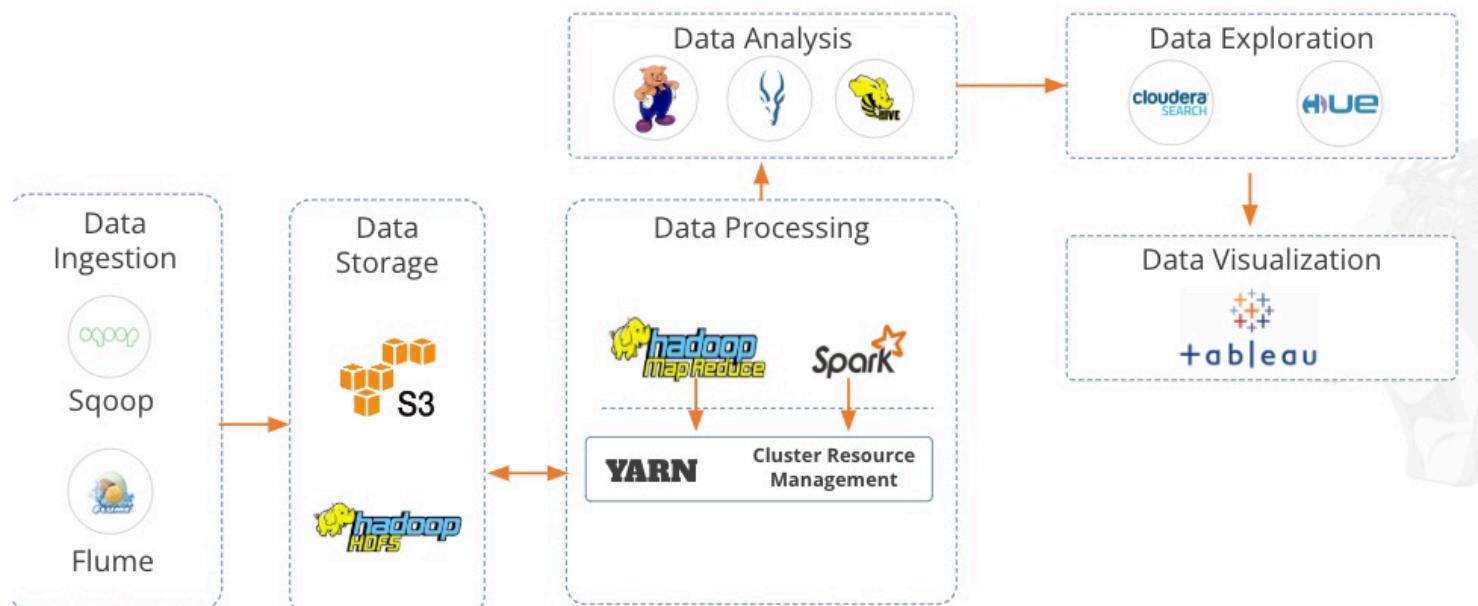
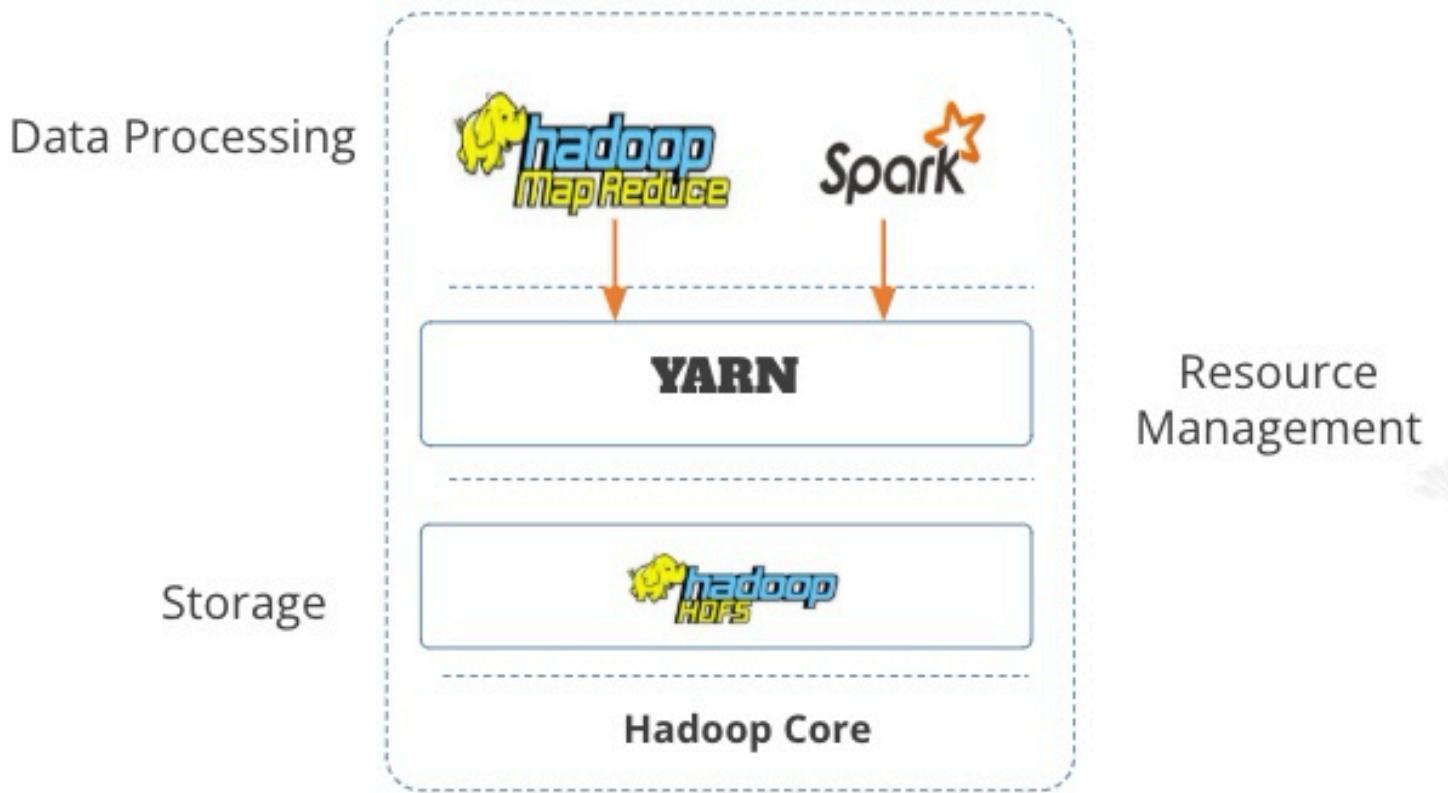
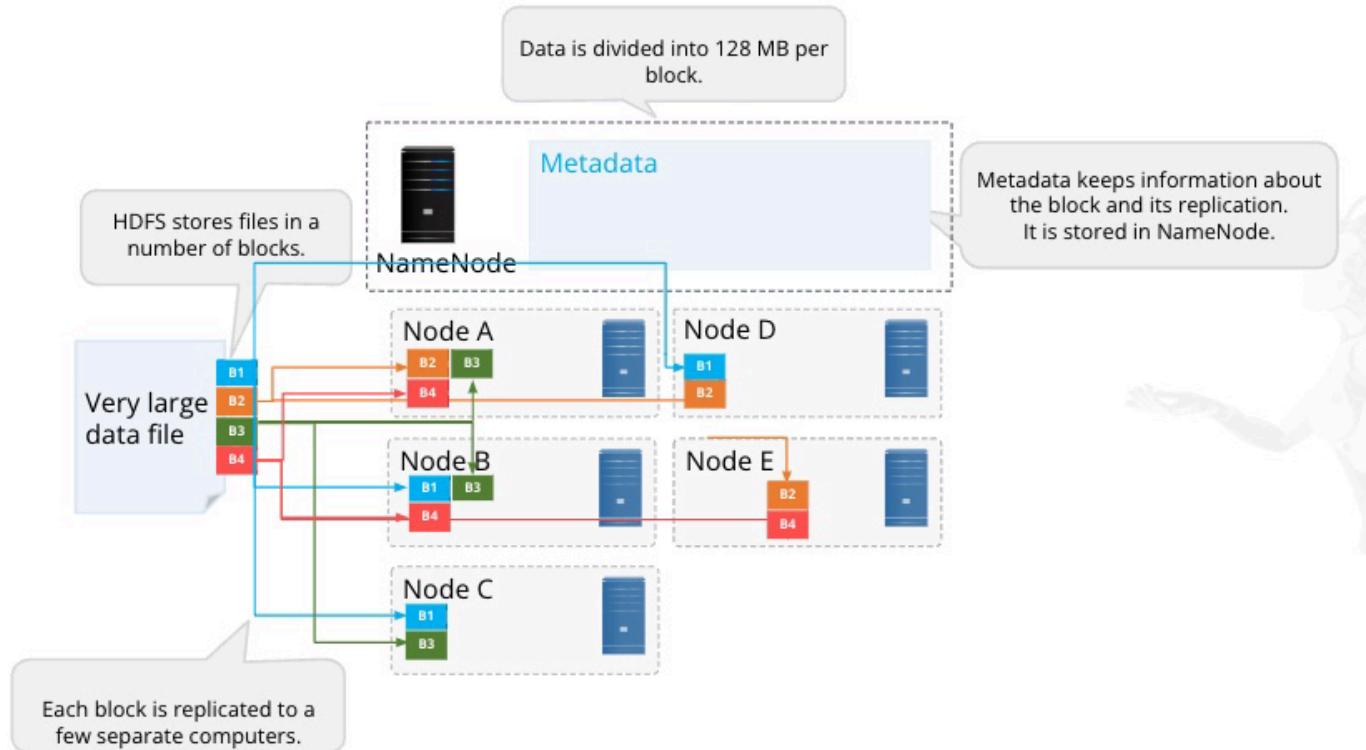
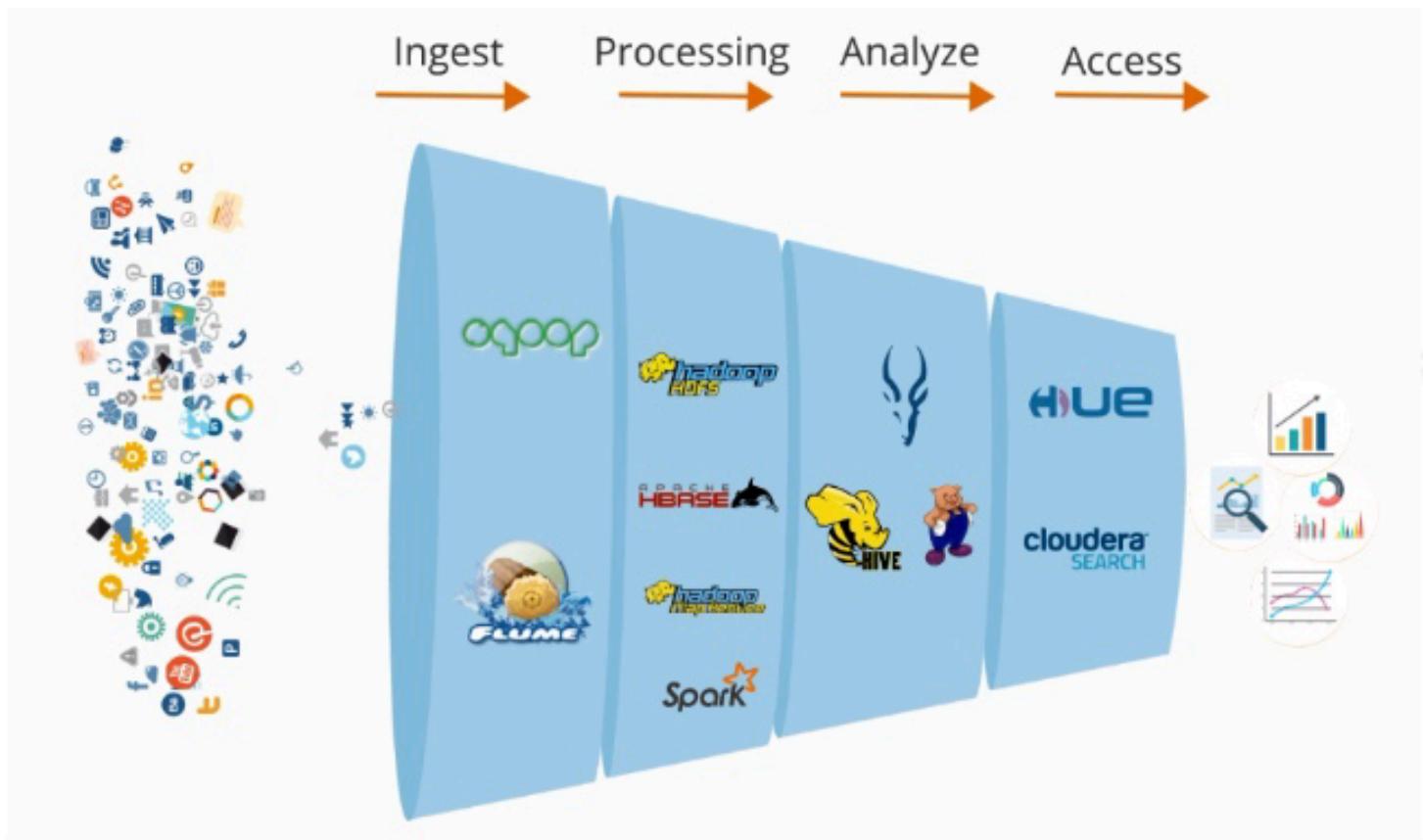


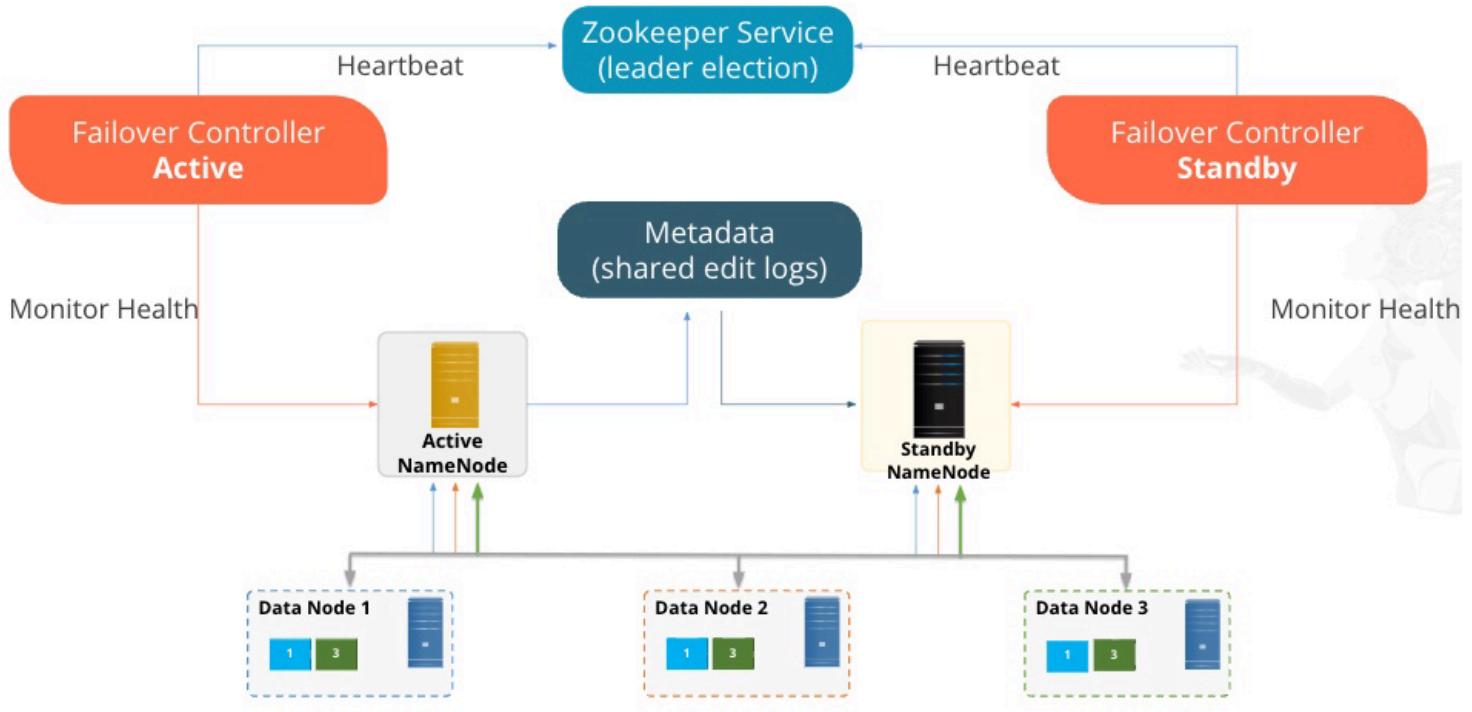
Hadoop Ecosystem Fundamentals Of Distributed Systems

- Intro to Hadoop
- Traditional system vs Hadoop
- Hadoop Core System
- Components of Hadoop system
- Big data processing
- HDFS
- YARN

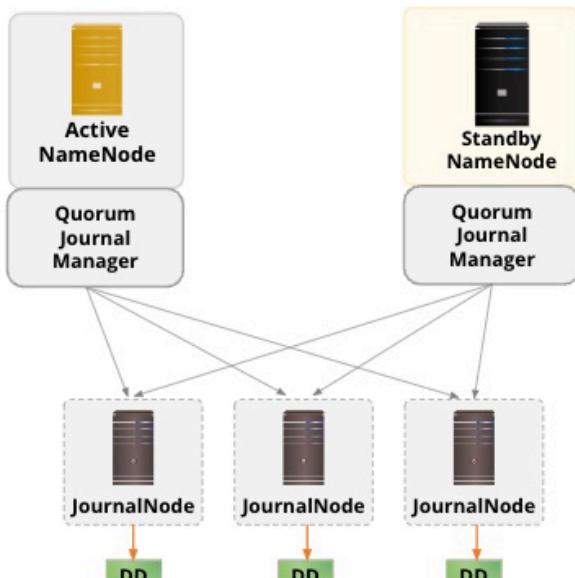




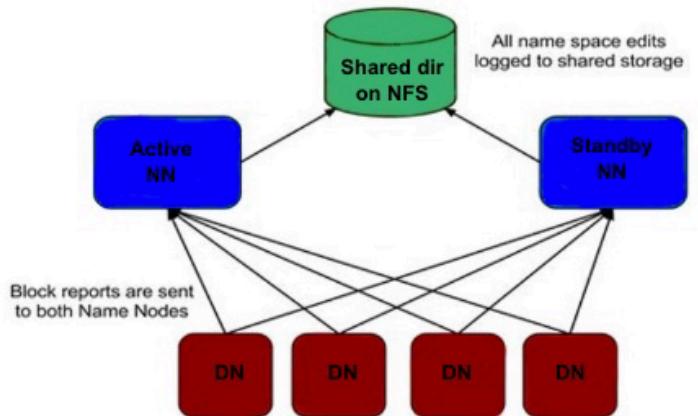




Quorum-based Storage

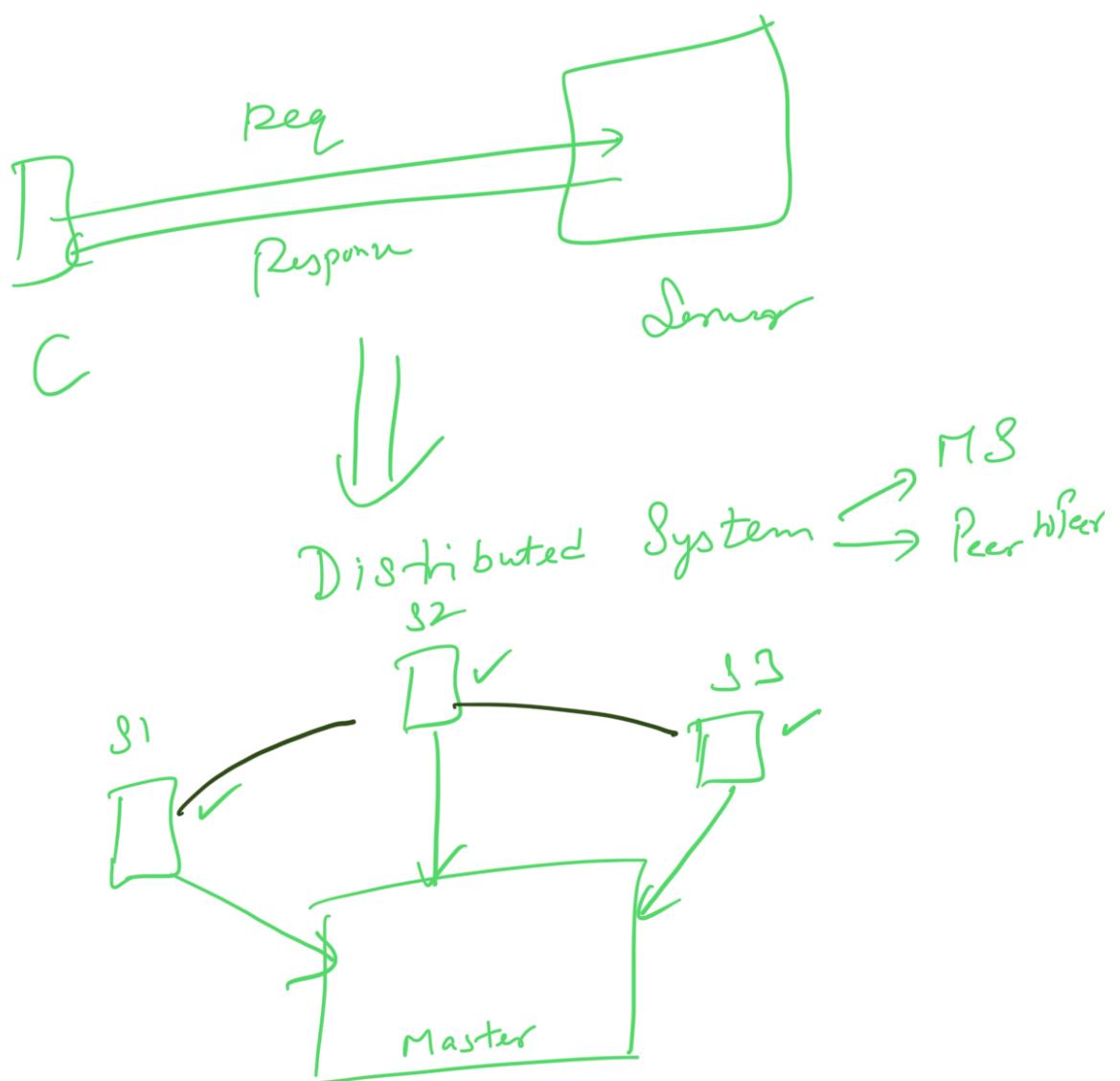


Shared storage using NFS



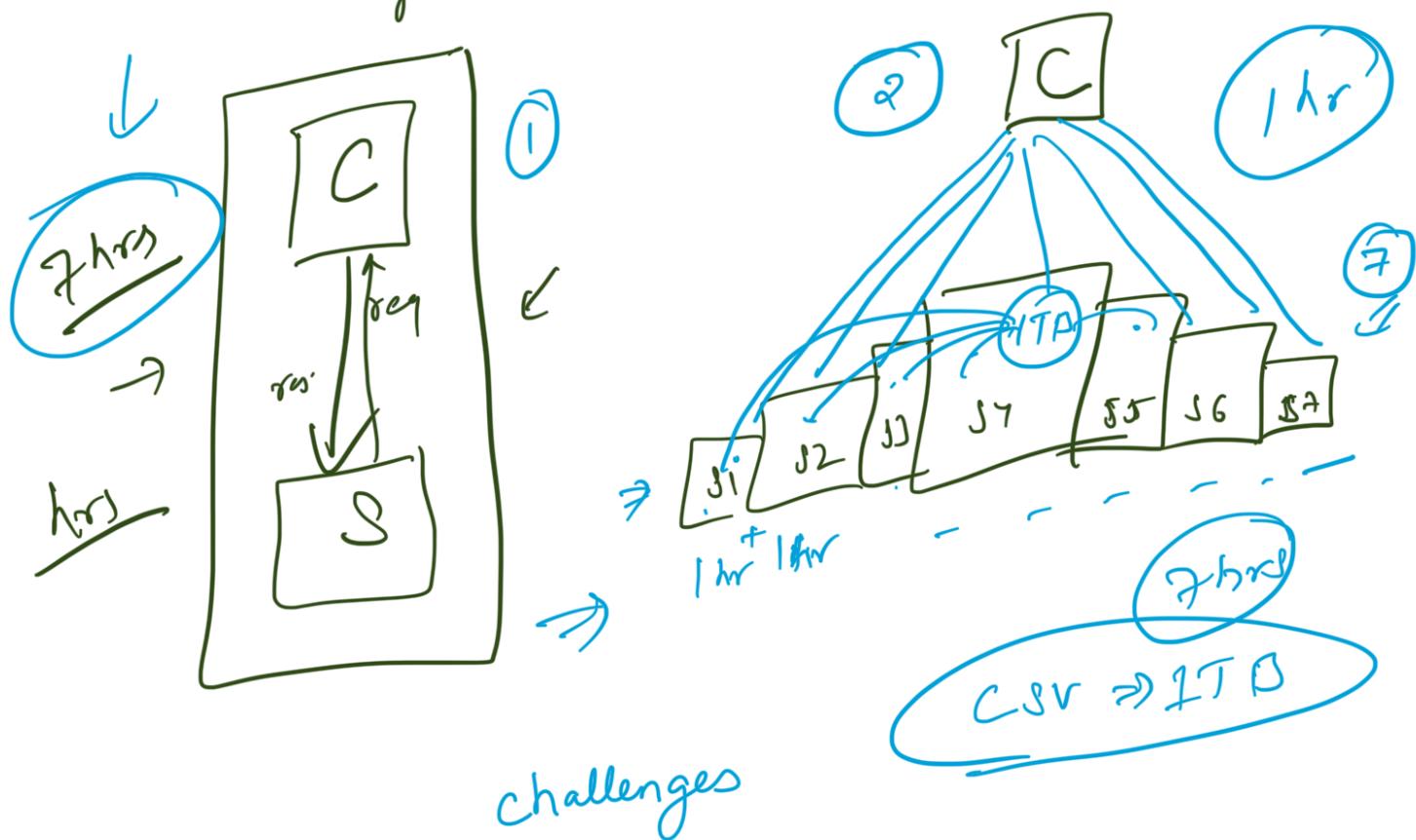
DE

Distributed System
↳ Cluster
↳ Nodes
↗ Services



- ⇒ It is a model in which components located
↳ ... communicate &

on a networked computers -
Coordinate their actions by passing
msgs.



① System failure

② Limited Bandwidth

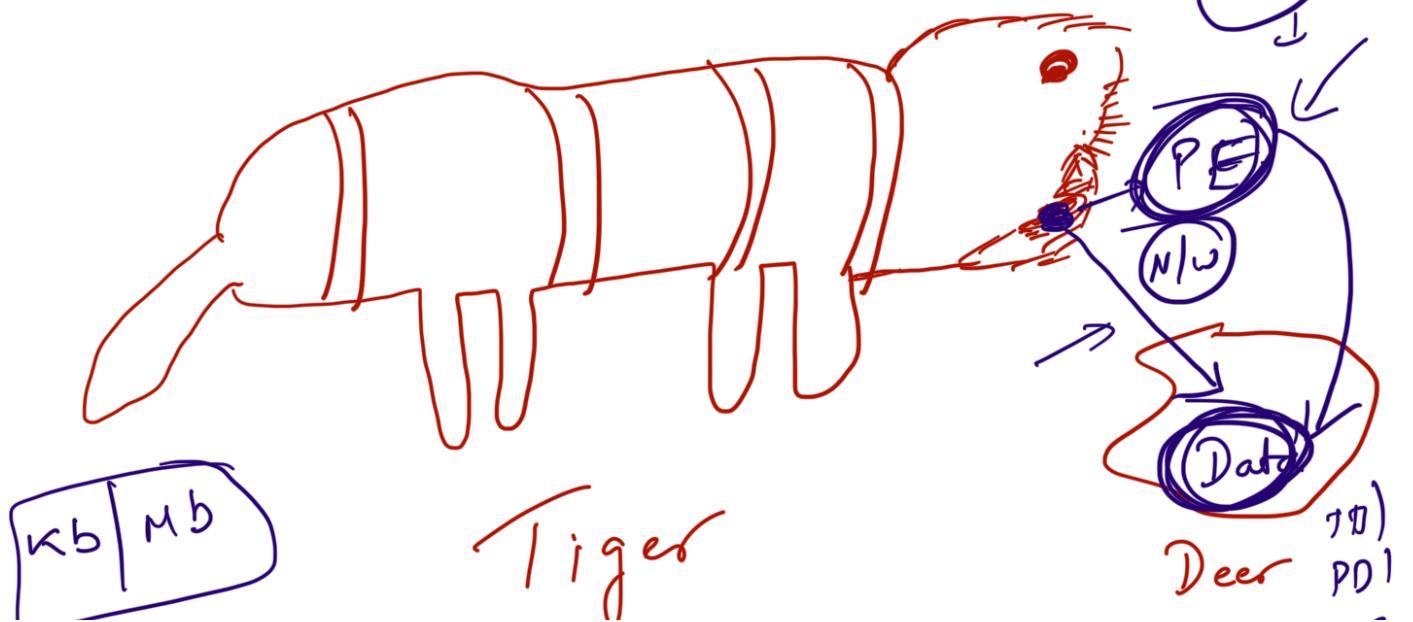
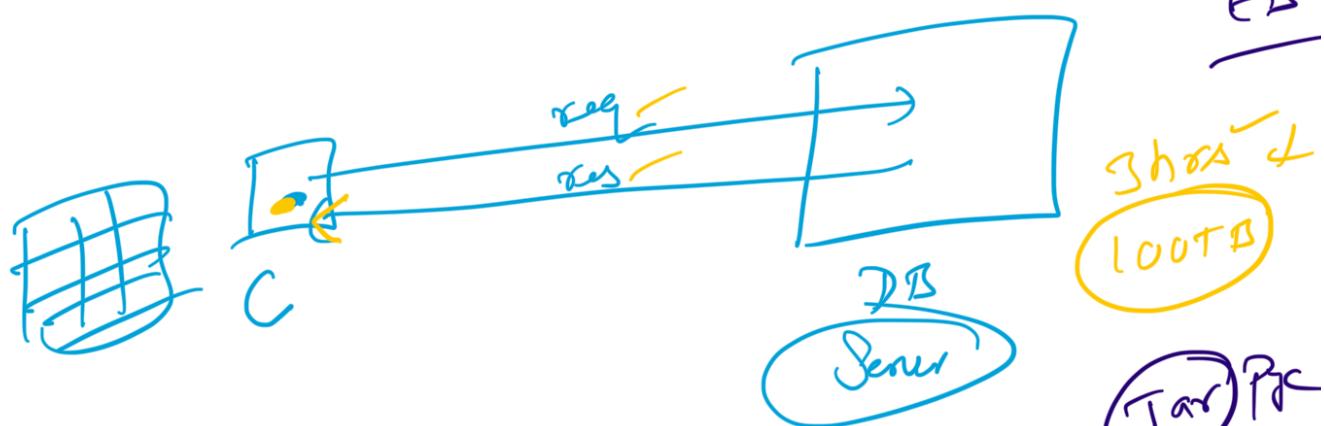
③ High Programming Complexity

Ph - 3 \Rightarrow Hadoop

Hadoop

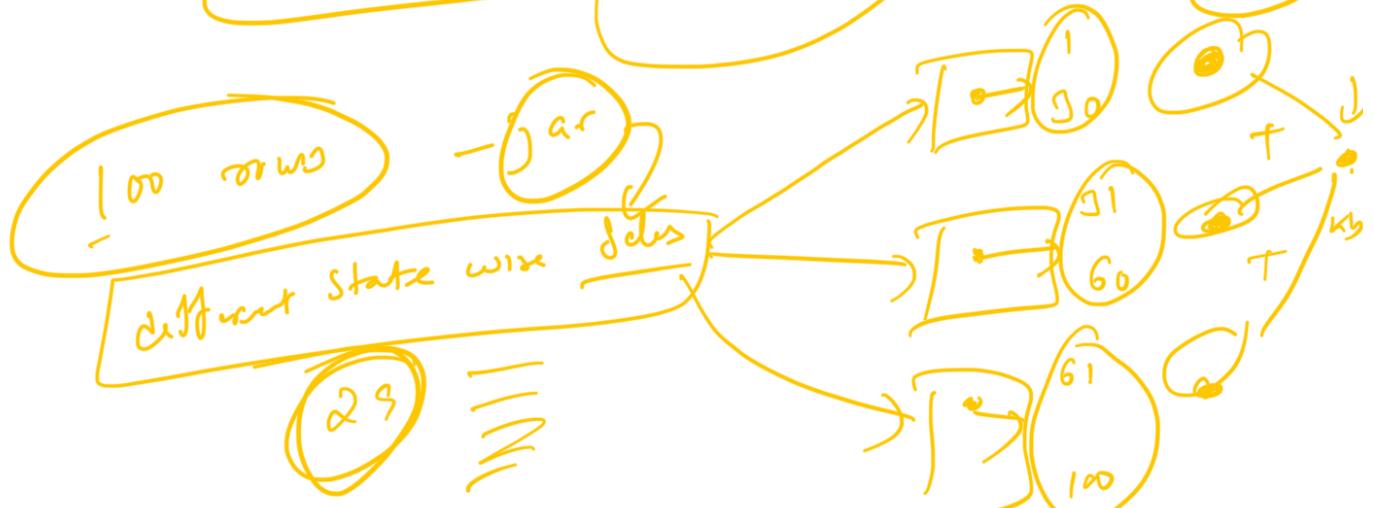
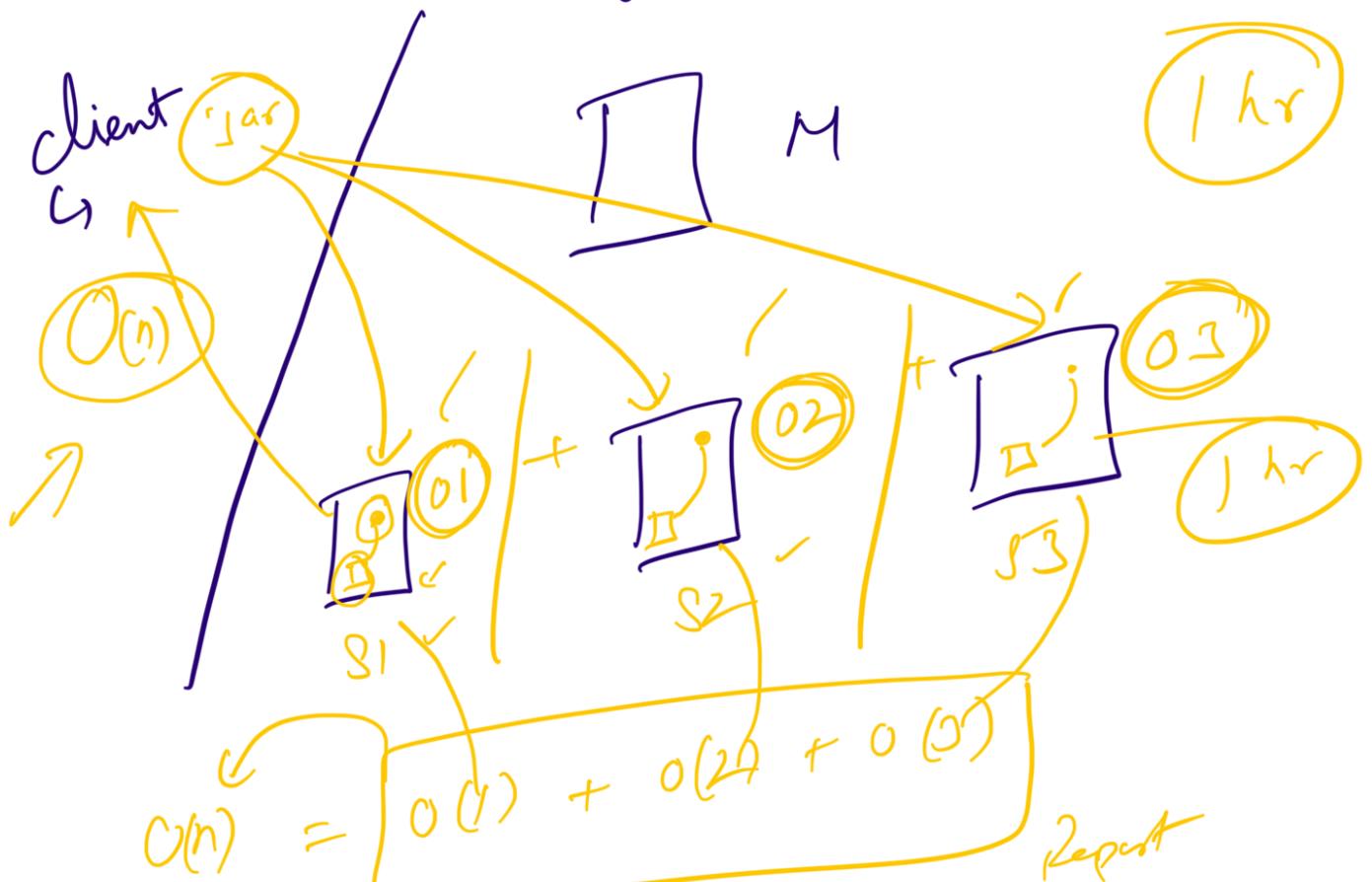
→ It is a framework that allows distributed processing of large data set across clusters using Simple

of Community Computers) -- d
Programming models.



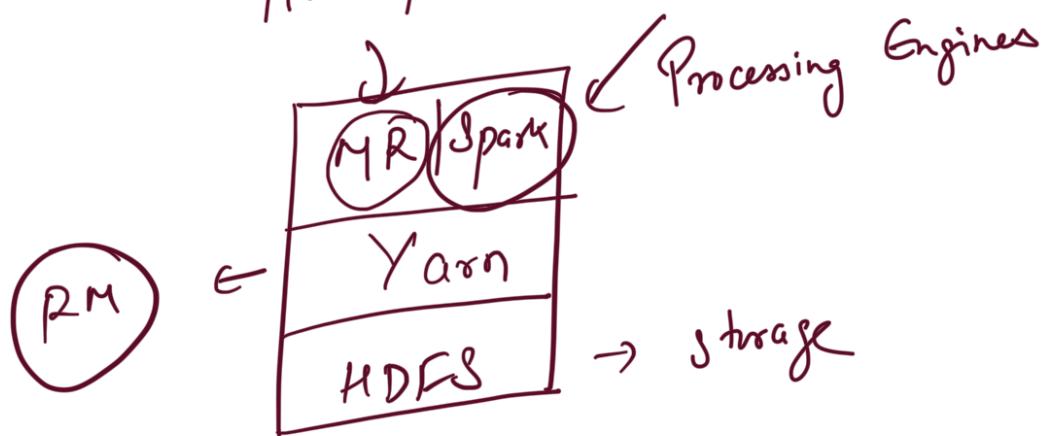
eB

→ PE which is of HB
comes close to Data and
Processing happens locally.

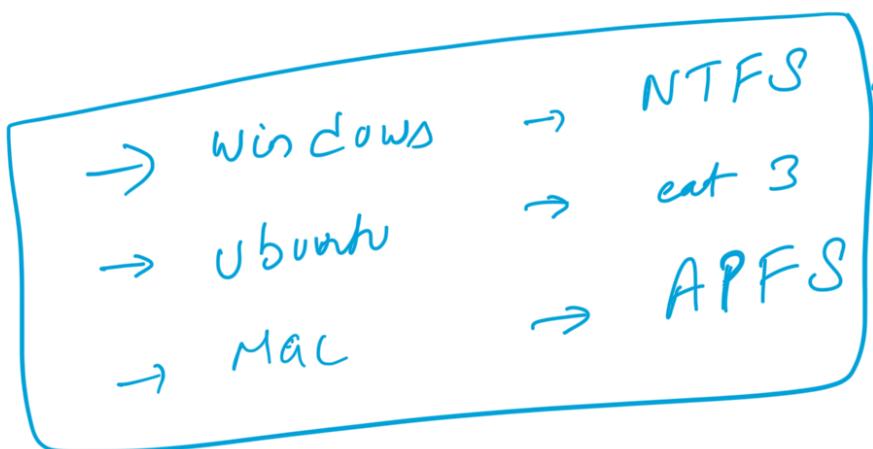


J	Doe	III
John	Doe	III

Hadoop Core Components



HDFS (Hadoop Distribute File System)

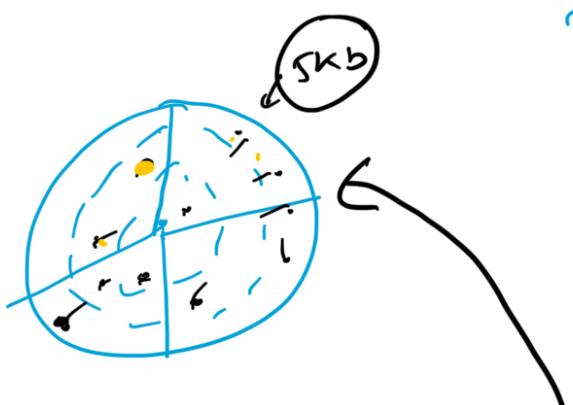


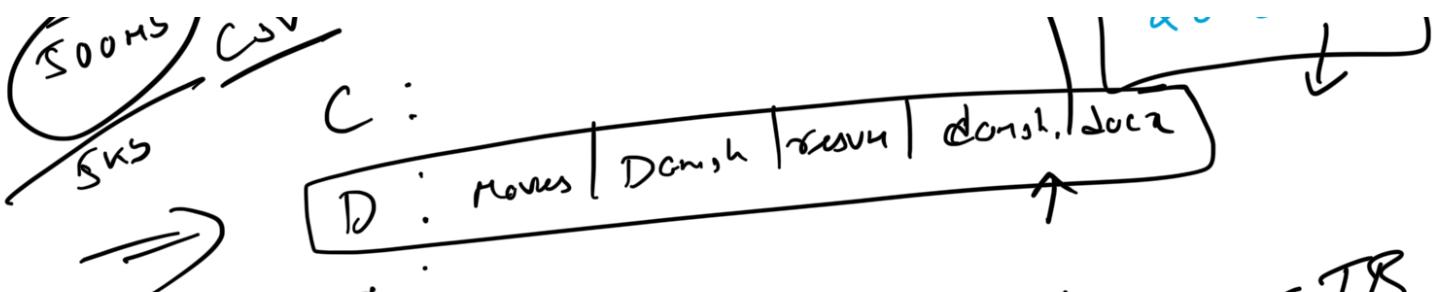
Danish. Lock

↳ 100 KB

5 kb
block

↳ 50 blocks

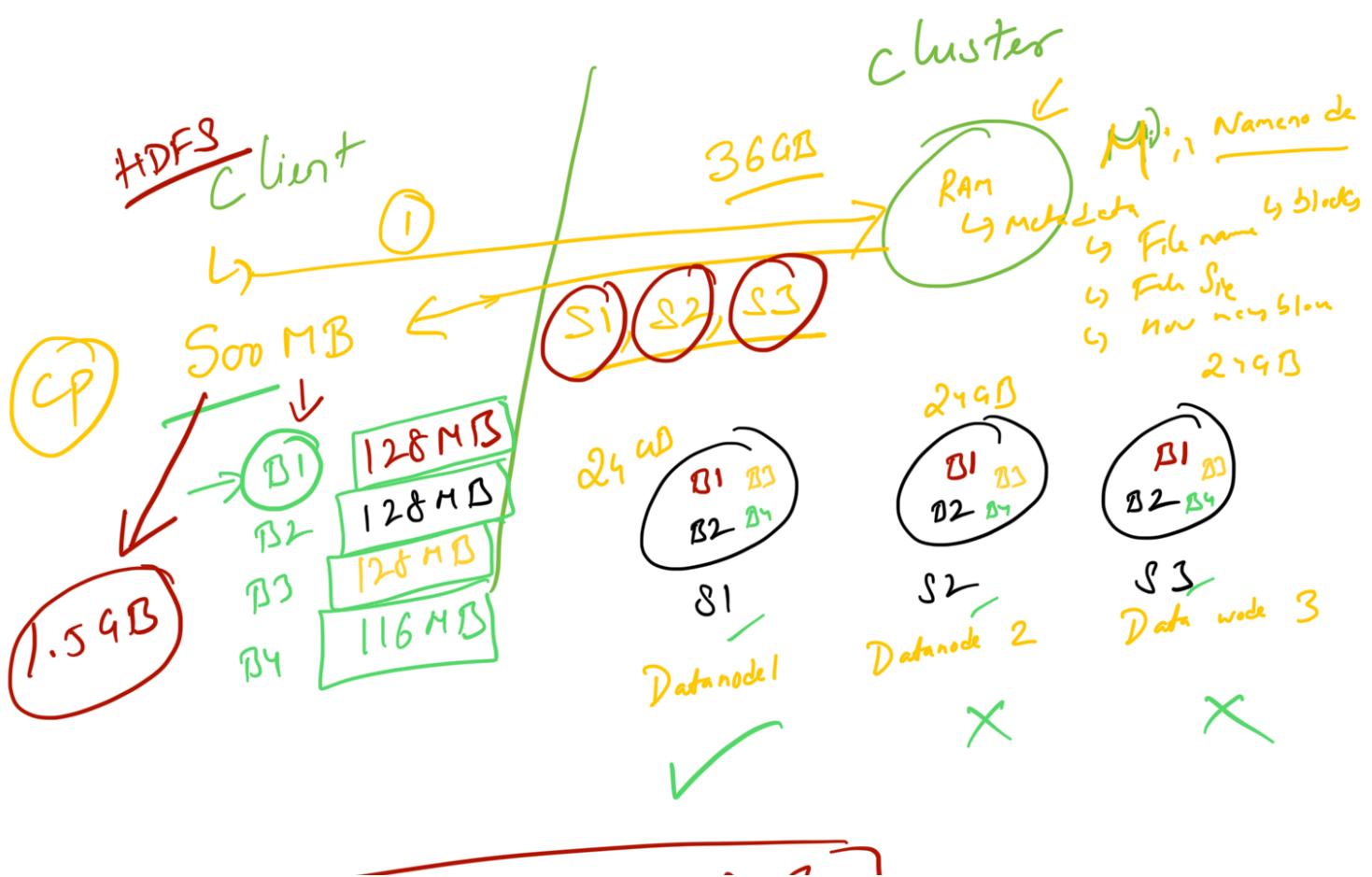




→ Cost $\Rightarrow \$10000 \text{ to } \14000 per TB
 → Speed \Rightarrow
 → Reliability \Rightarrow

HDFS

- ↳ Focus (1 Terabyte / dc)
- Speed
- Reliability (copies)



Block size = 128 MB
Replication factor = 3

File size = 477 MB
Block = 128 MB
RF = 3
Blocks = $\frac{477}{128} + 1$ = 12