

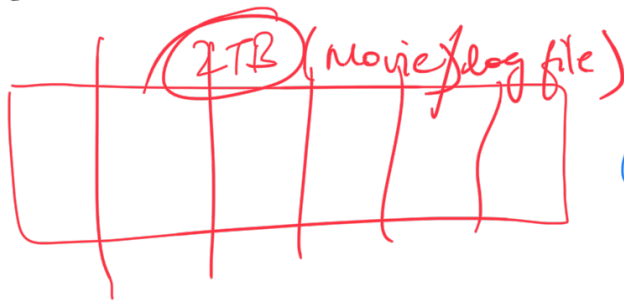
Quora core study

- ① Large file systems
- ② Elastic Sca
- ③ Microservices vs Monolith

- ① Losing data
- ② Large files
- ③ Hinders parallel processing

## SYSTEM

- ① Store files
- ② Read files
- ③ Pattern finding.



- ④ Large files
- ⑤ Guarantees not lose file



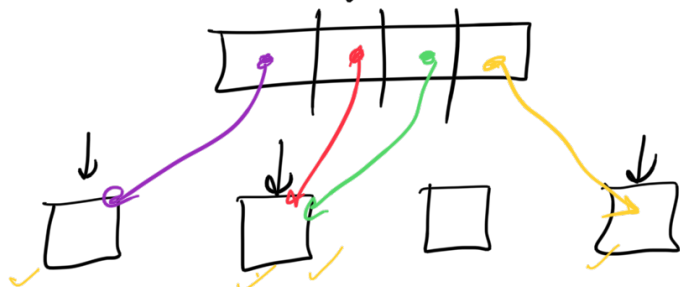
128MB

## ① Namenode Server

F1:B1 → m1  
F1:B2 → m2  
F1:B3 → m2  
F1:B4 → m4

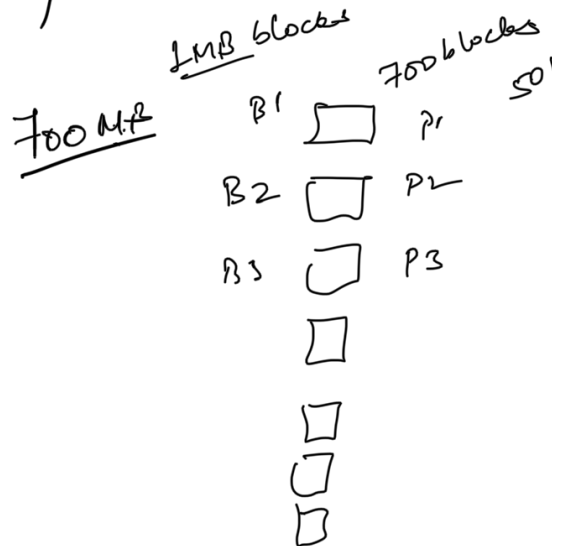
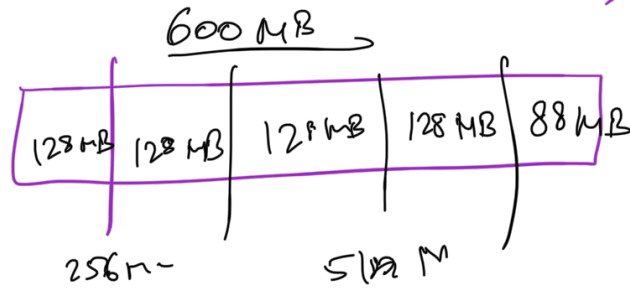
128MB

Break the file into blocks



{ File + Block } → machine

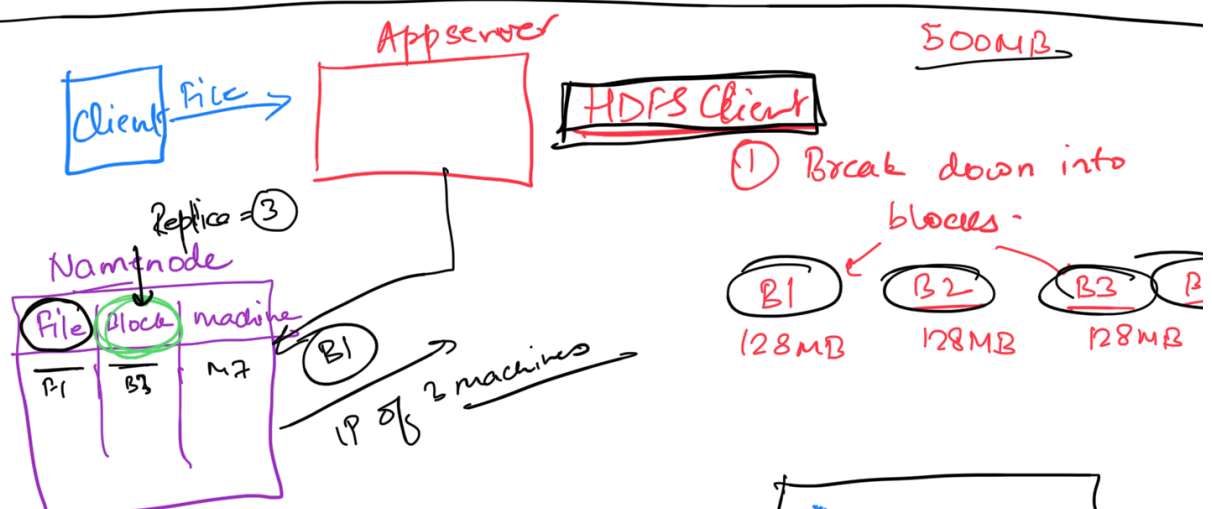
$$\frac{100 \text{ TB}}{128 \text{ MB}} = \frac{100 * 1000 * 1000 \text{ MB}}{128 \text{ MB}} \approx 1 \text{ million}$$



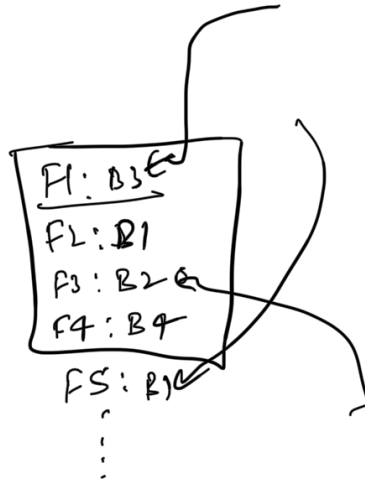
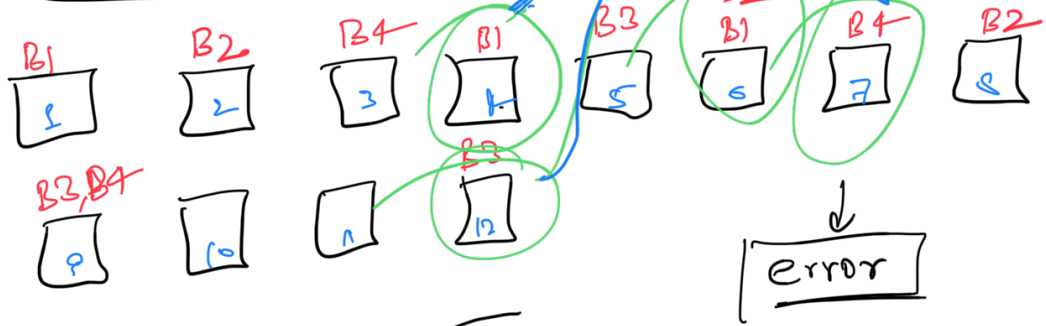
① Split file into blocks (HDFS 2.0 → 128 MB)  
 40 MB → 1 block of size = 40 MB

② Namenode  
 B1 → machine

③ Replication → every block is stored on R # of unique machine

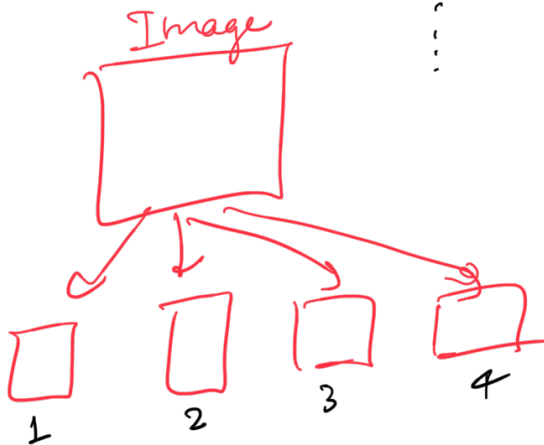
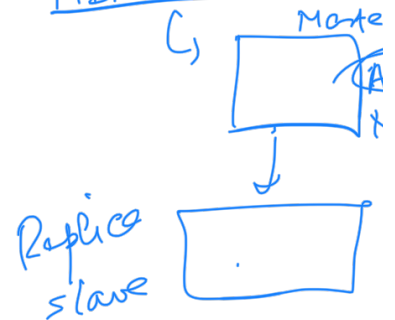


Master  
Slave format  
Data Nodes



HDFS v1 ← Multi-NN

HDFS v2

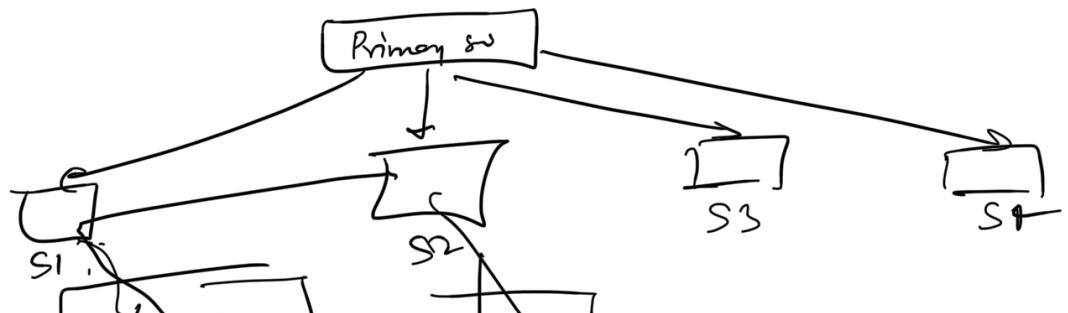


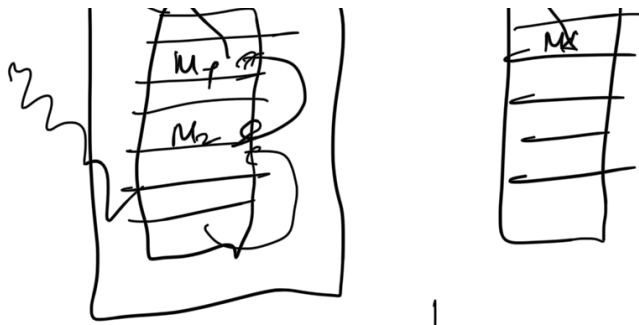
adavet.jpeg    B1    M3  
...    B2    M2

HIVE

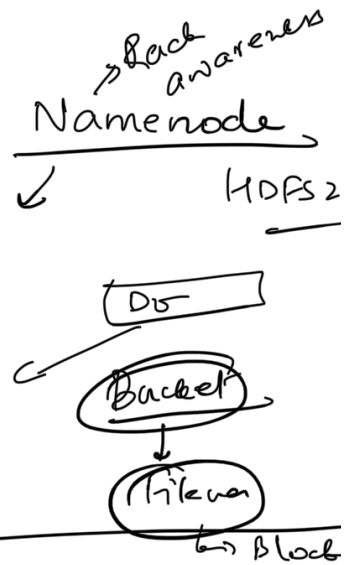
↳ SQL query

↓  
(Map Reduce script)

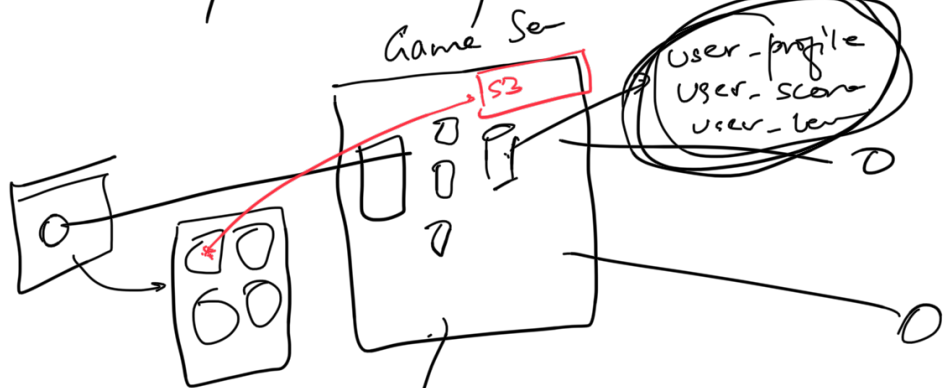




(M1)	(R1)	Free space	Load
M2	R1	—	
M3	R1	—	
⋮	R2	—	
⋮	R2	—	
M12	R3	—	



(S3)

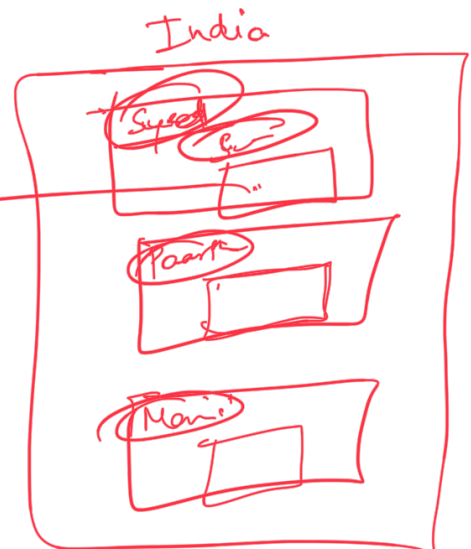


latency

CDN cached

S3/HDFS

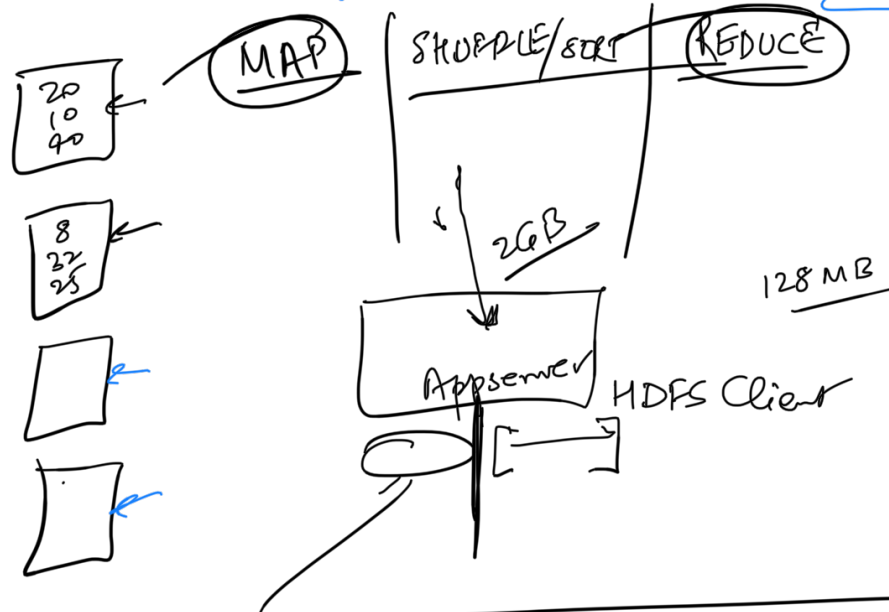
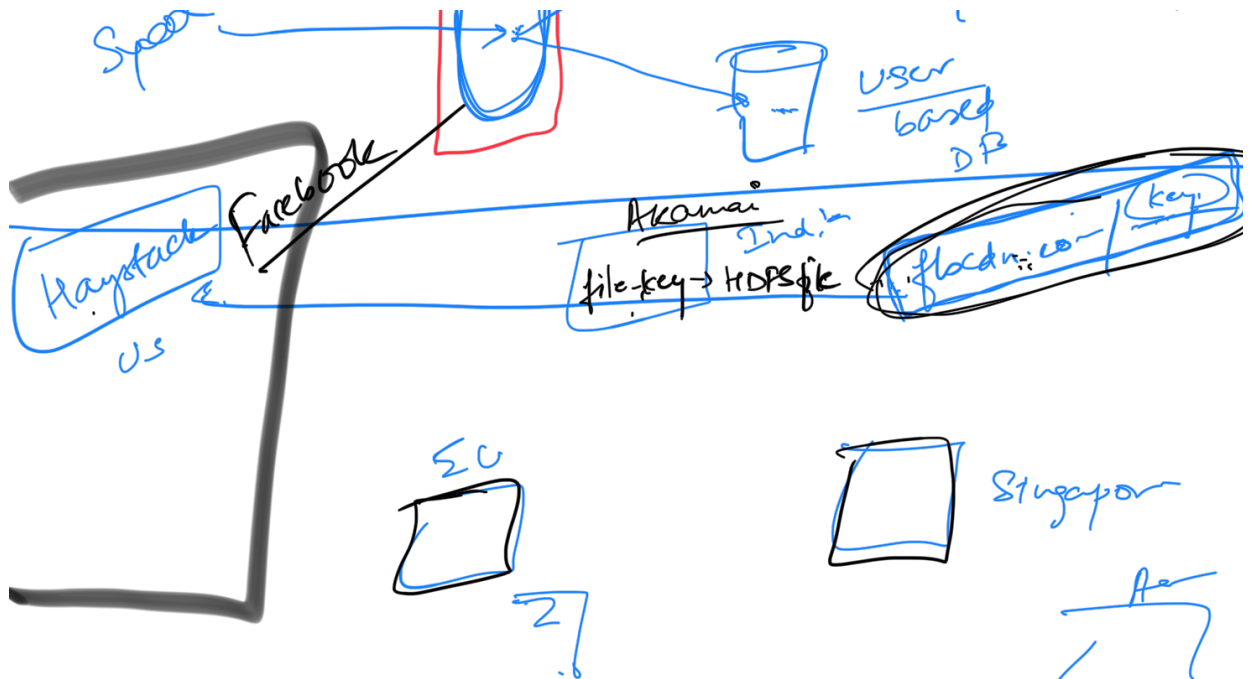
US East



Images | Videos

S3/HDFS

L3/S2 → I/O



## QUORA

- ① Post a question
- ② Ranking answers for a question
- ③ Answer questions
- ④ Q/A news feed
- ⑤ Topics (Questions can belong to multiple topics)

Question page	
Q. What is scale?	of Core
At: <u>---</u>	
AL: <u>---</u>	
As: <u>---</u>	

Q&A have creators as owners

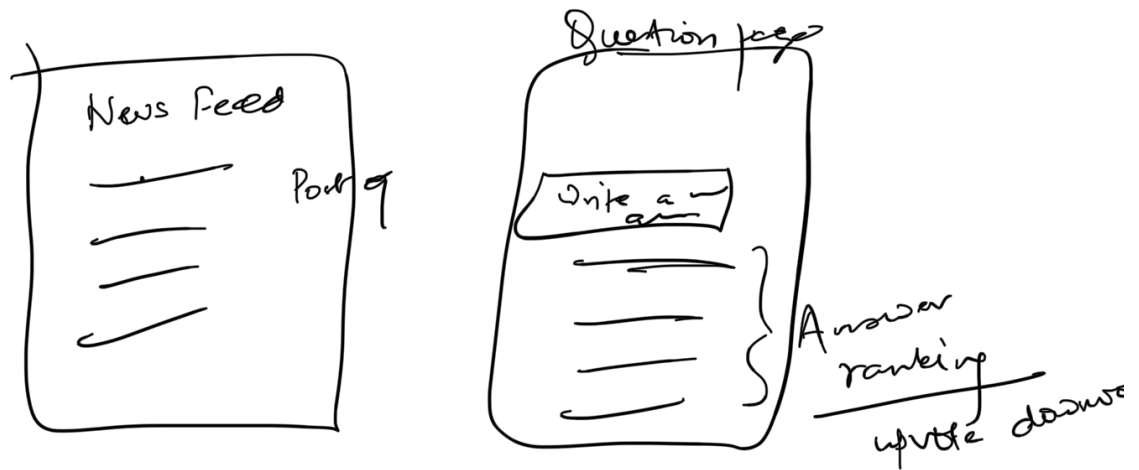
① Requirement MVP

② Estimation Scale

① Shortlist

② Write/

- users can follow multiple topics)
- ⑥ Answers → text (restriction on size)  
→ Media
- ⑦ upvote / downvote answer
- ⑧ Search Questions.
- ⑨ List/ Search Topics
- ⑩ Notification → new question/answer on a topic.
- ② Design goal <sup>read vs</sup>
- ④ API design  
→ actual an



### Estimation of scale

- ① Add question }  
→ text
- ② Add answer }  
→ text
- ③ user → topics
- ④ topics list ←
- ① sharding
- ② write vs read hea
- 10K questions added every day.
- 50M DAD
- 1 Question → 5 Answer

$$10,000 * 365 * 5 * 5 \Rightarrow 9125 * 10,000 = 91.25 \text{ Millions}$$

①

90M Q&A

sharding or not?

50 million registered user  
↳ topics

②

50M DAUs

write heavy  
read heavy  
both