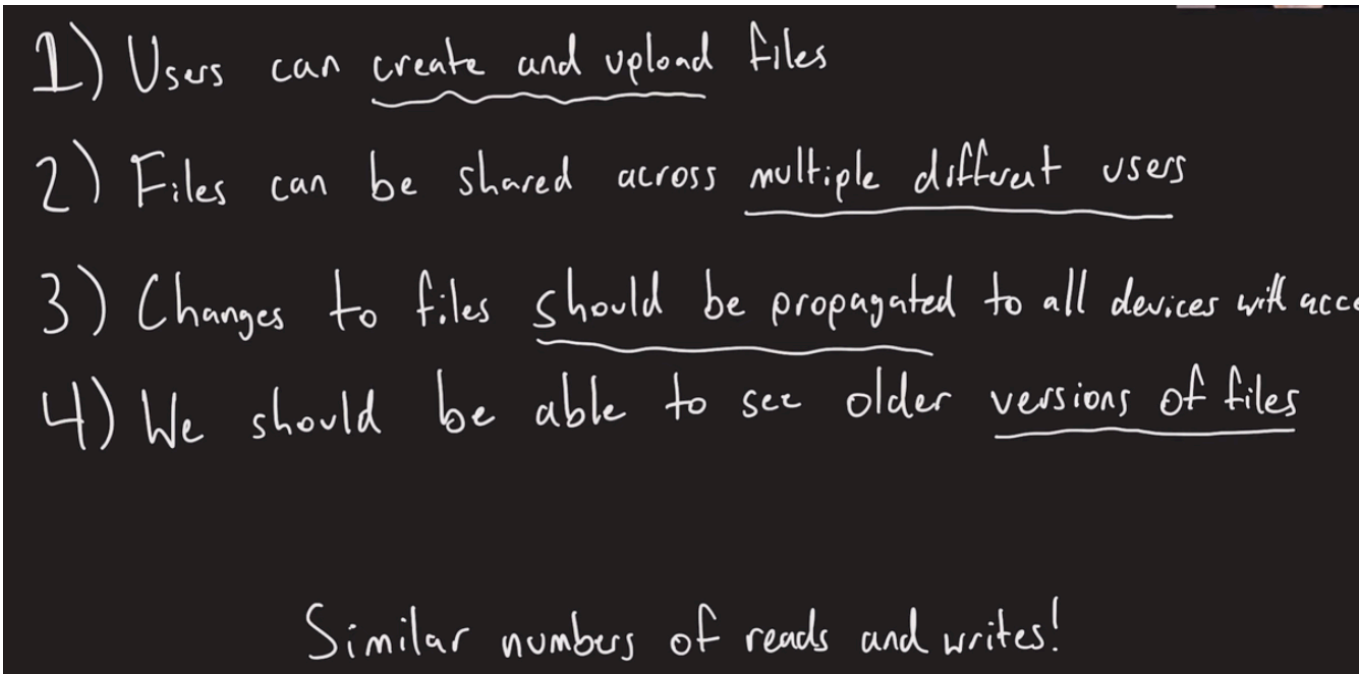
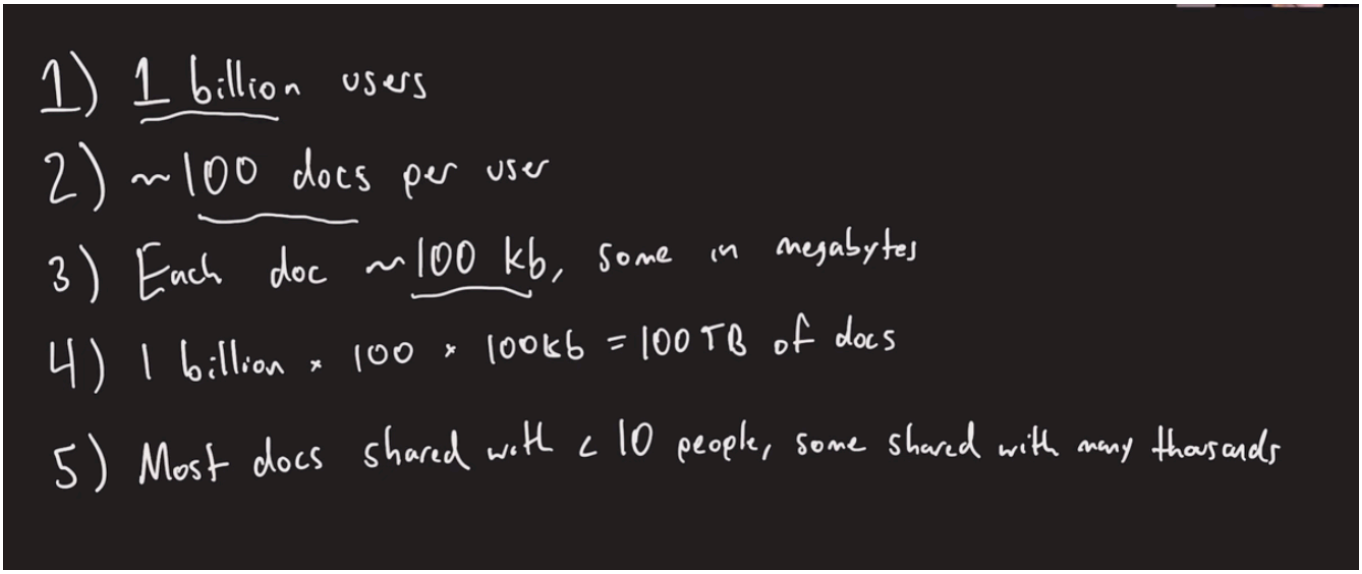
DropBox-GoogleDrive

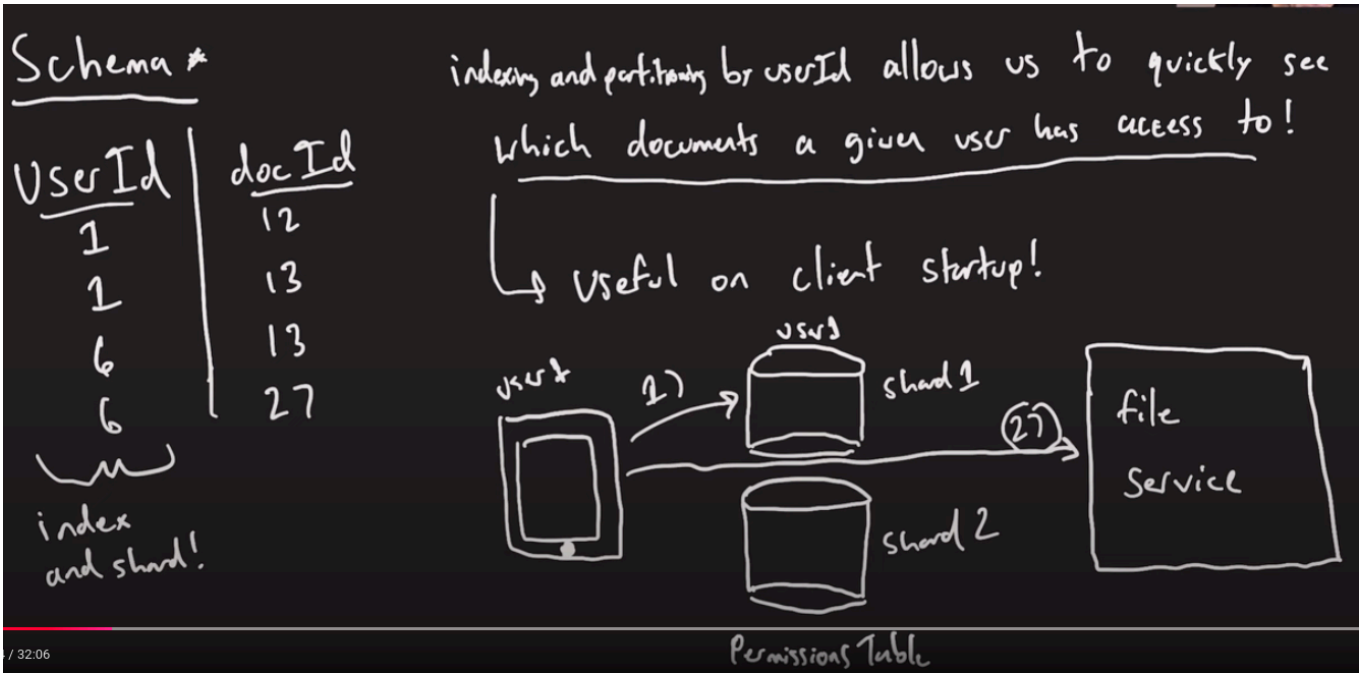
Requirements



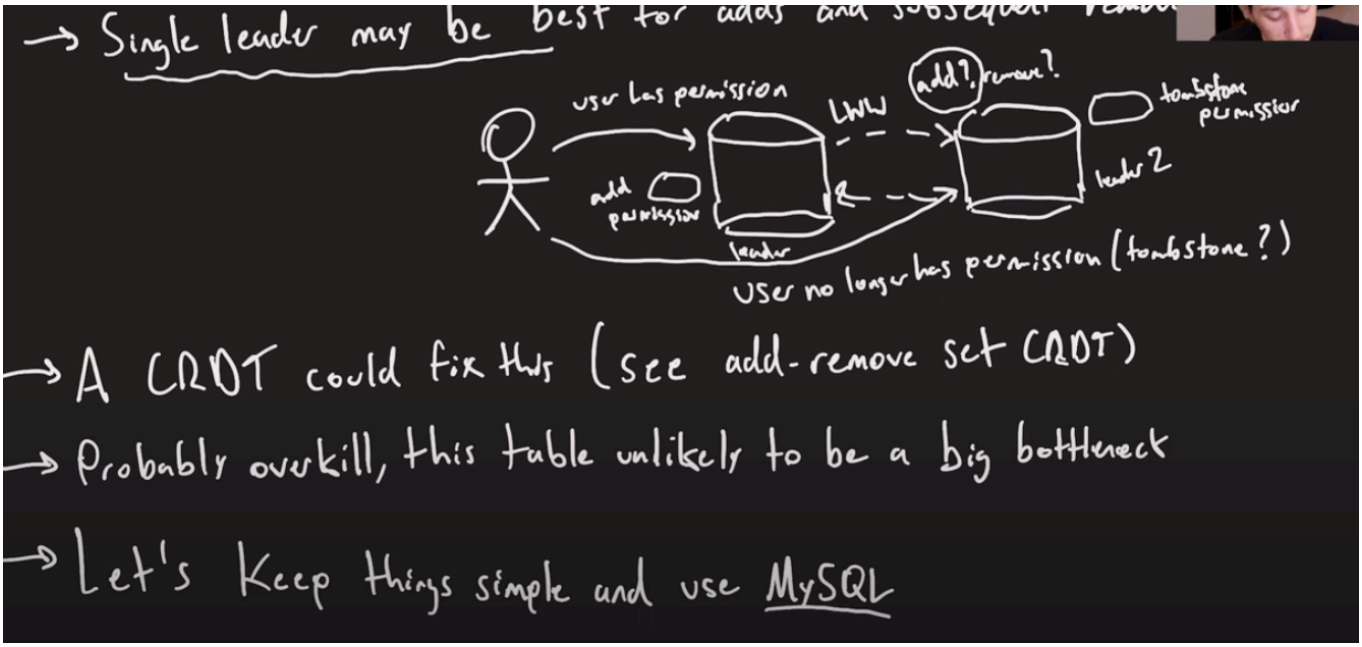
Capacity Estimates



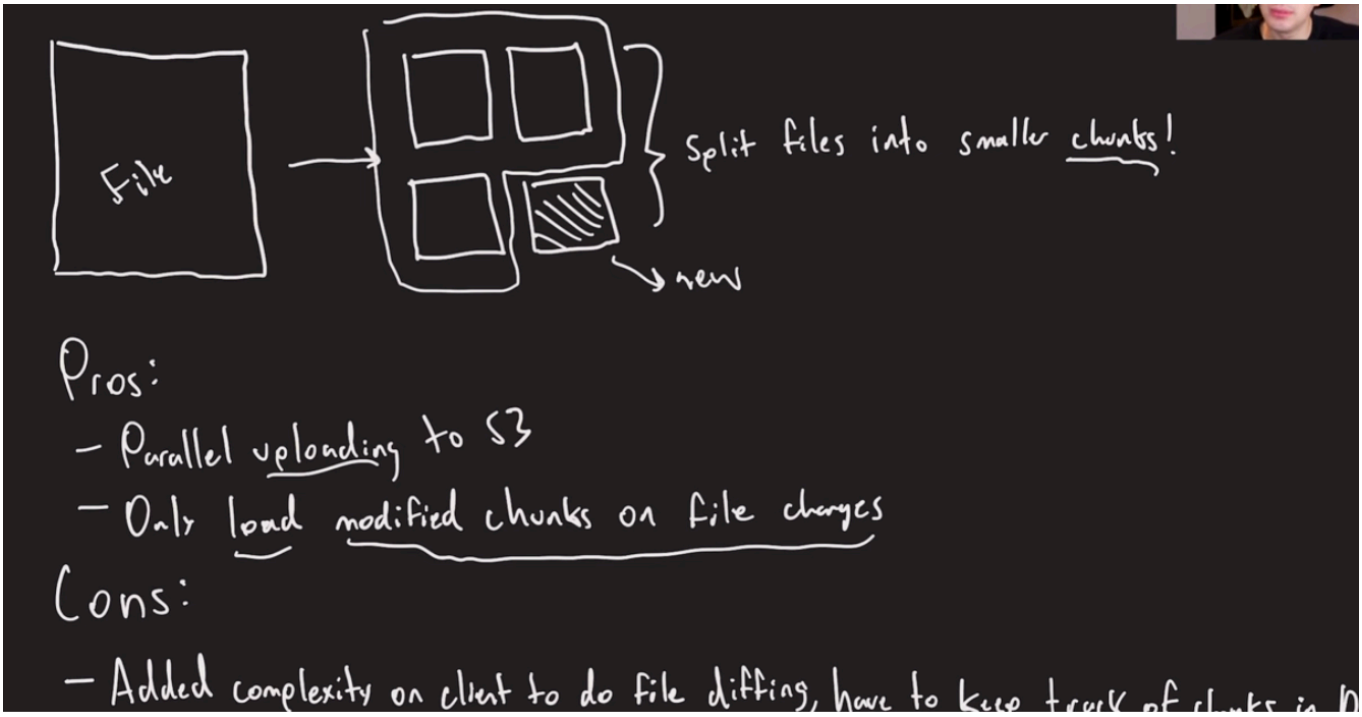
Document Permissions

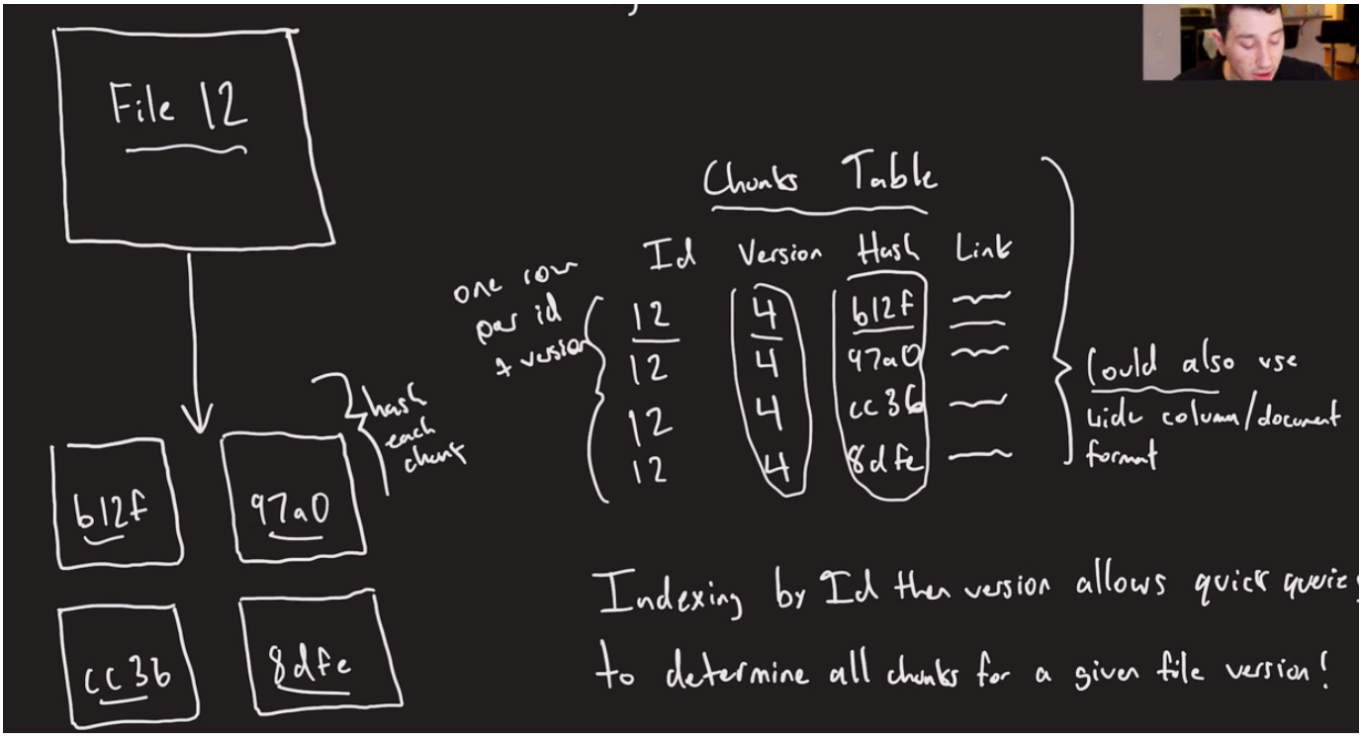


Replication

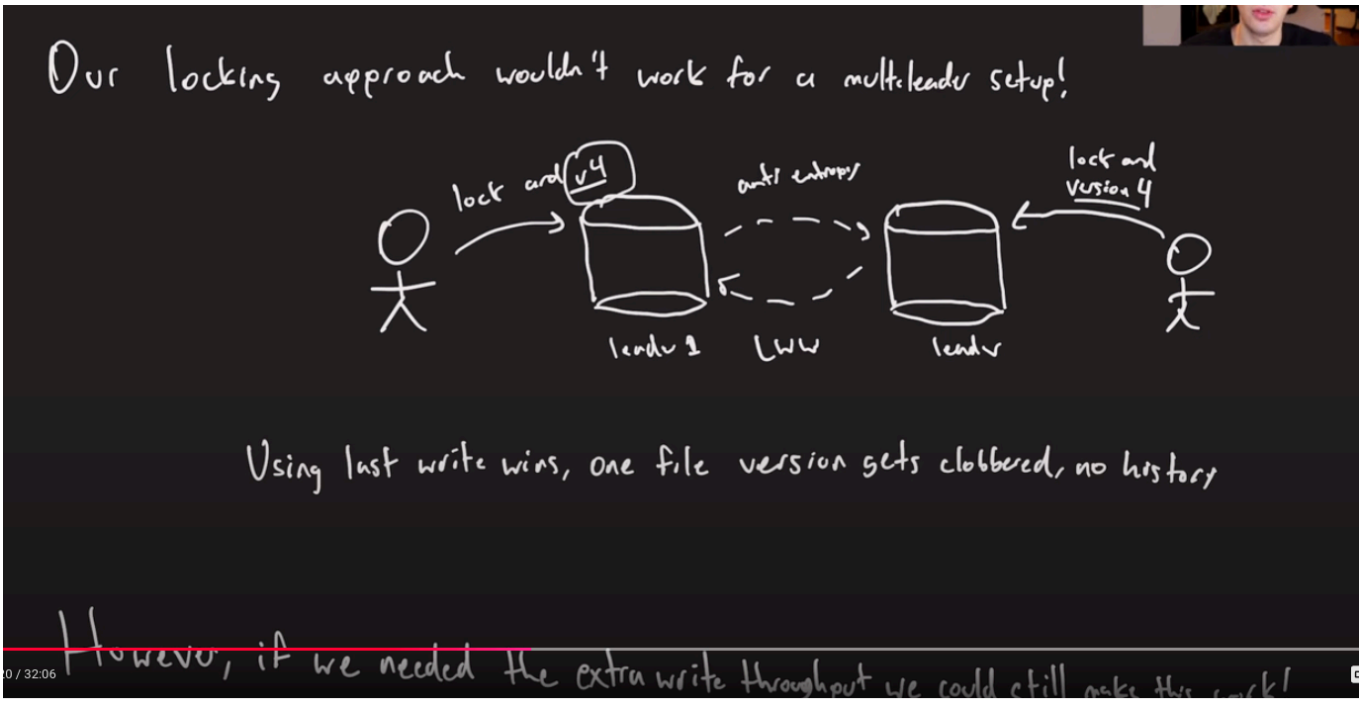


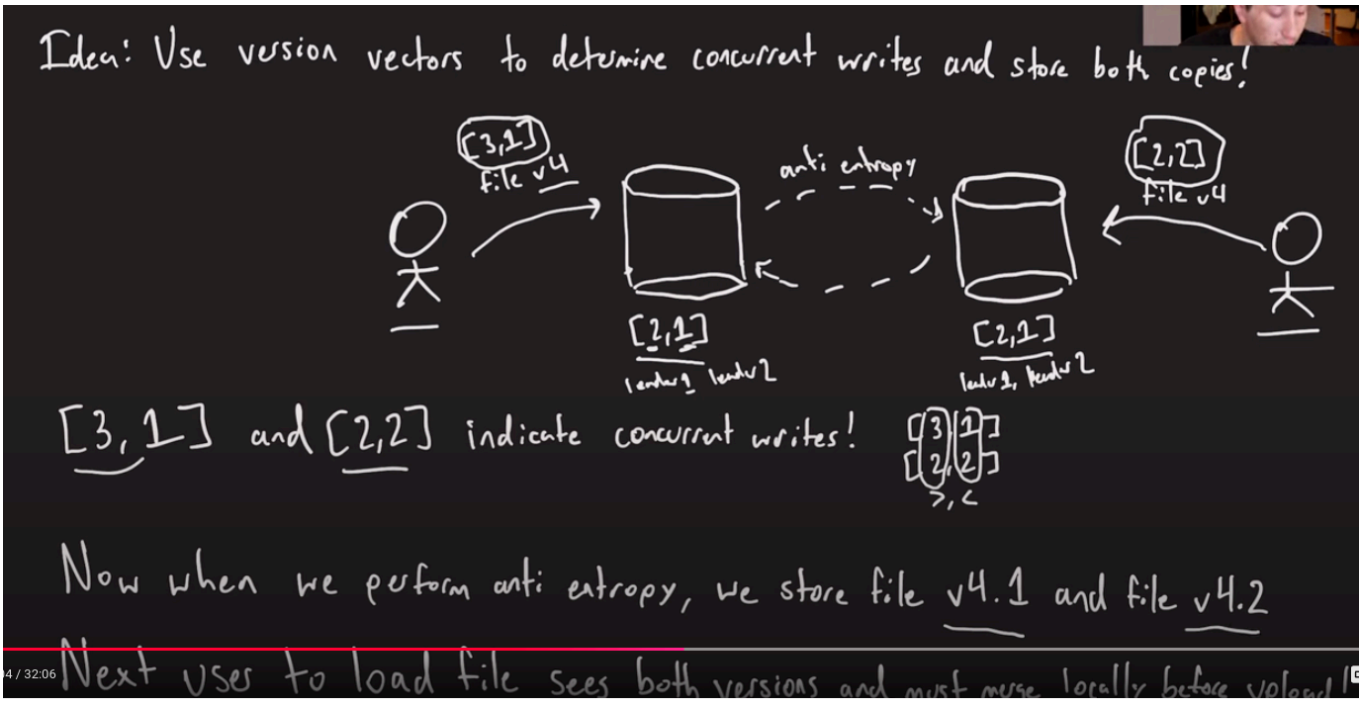
FILE Uploading





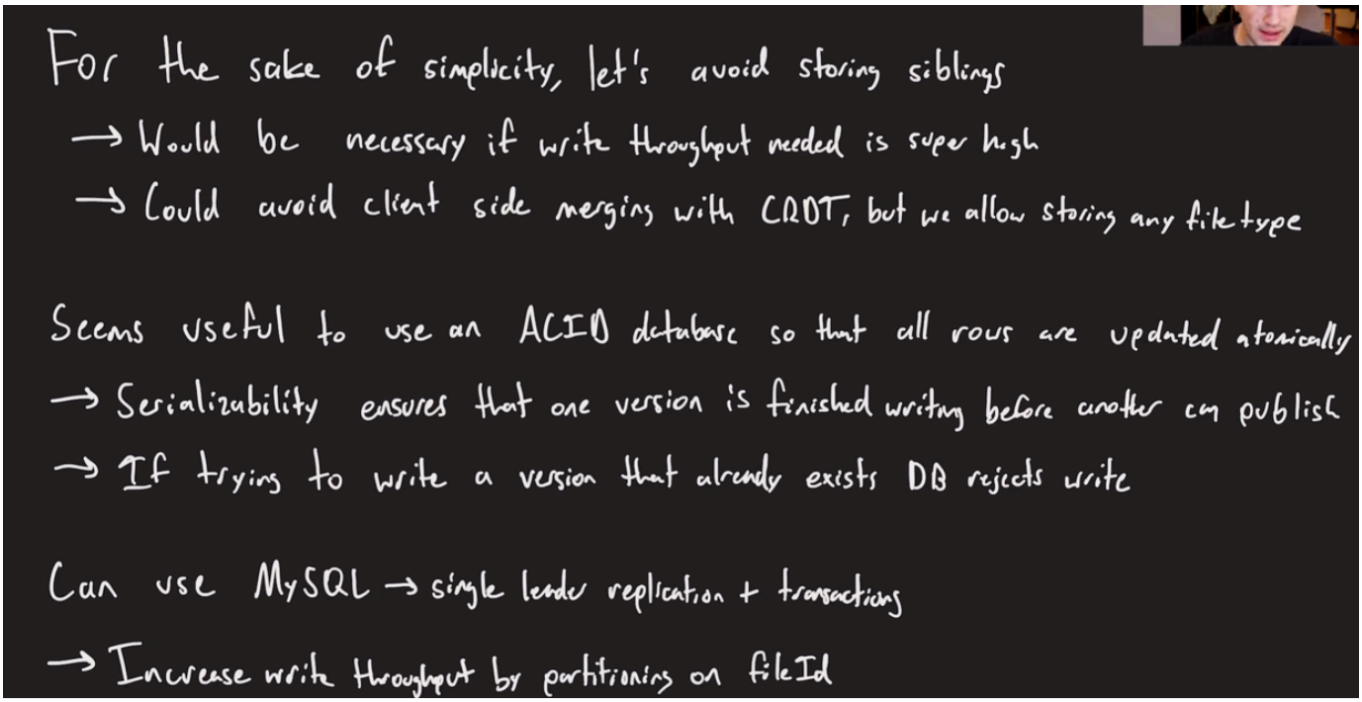
Chunks DB consideration



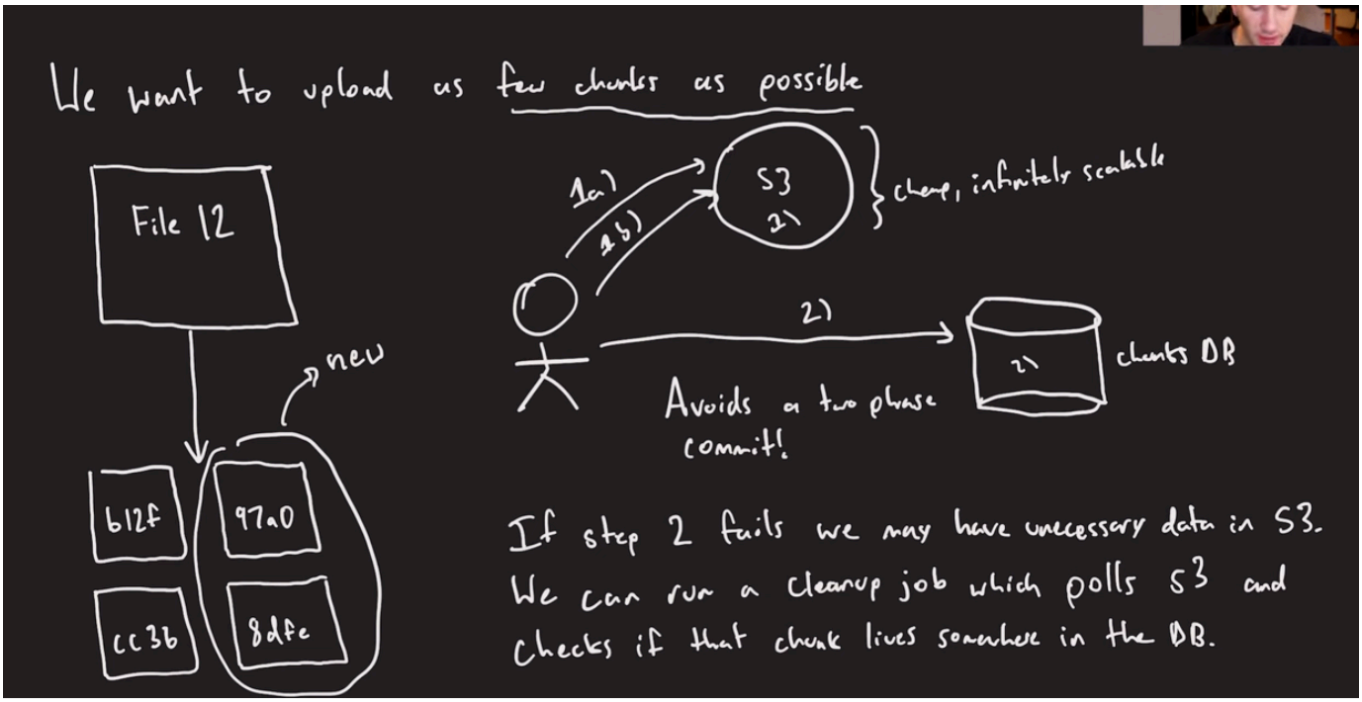


Next user to laod the file sees both versions and must merge locally.

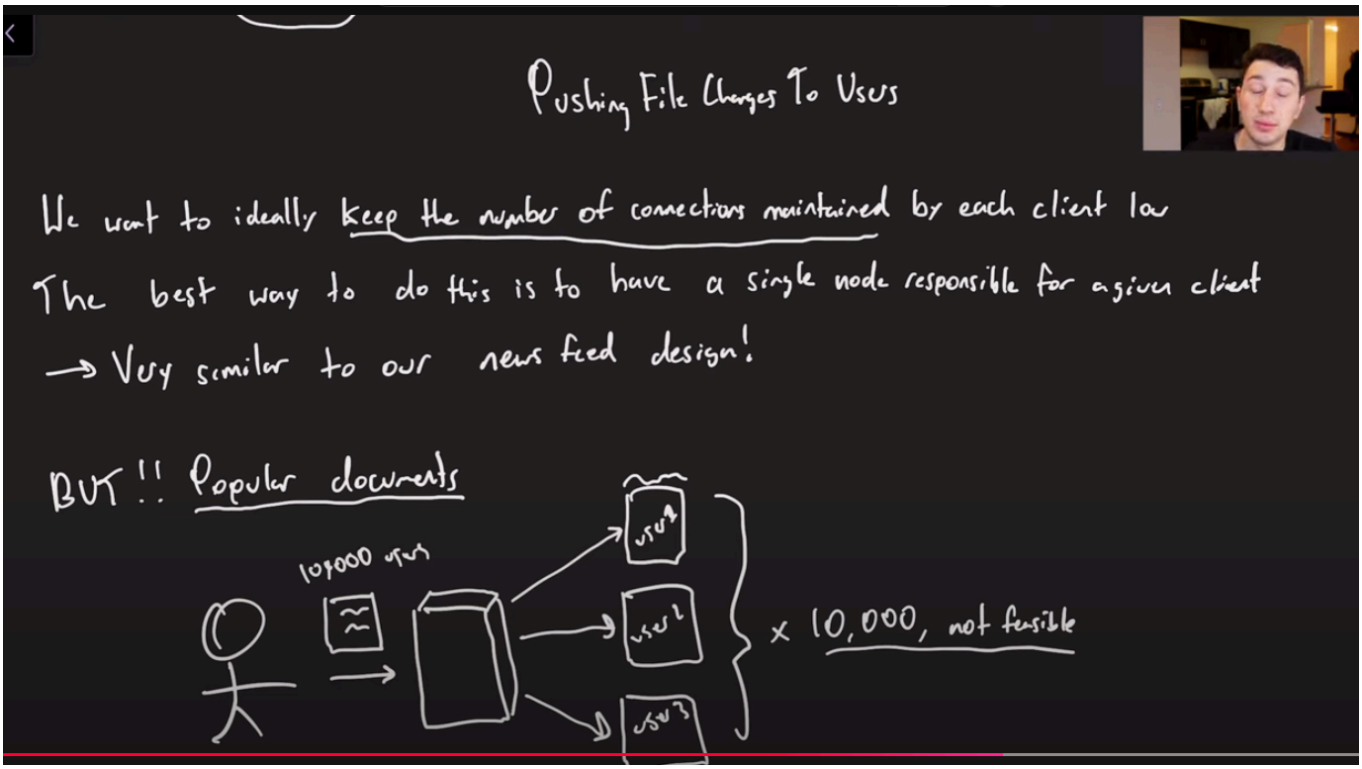
Lets try to upload using siblings

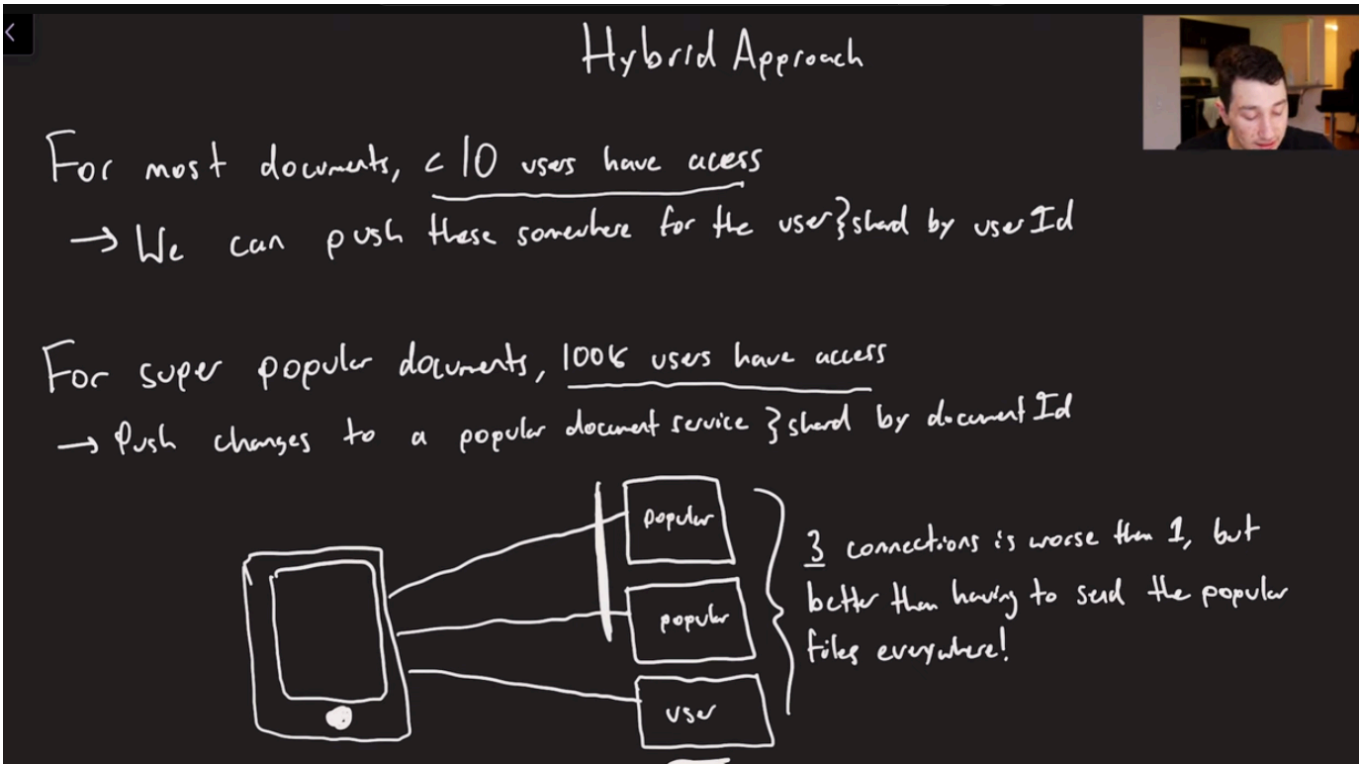


Uploading the actual files

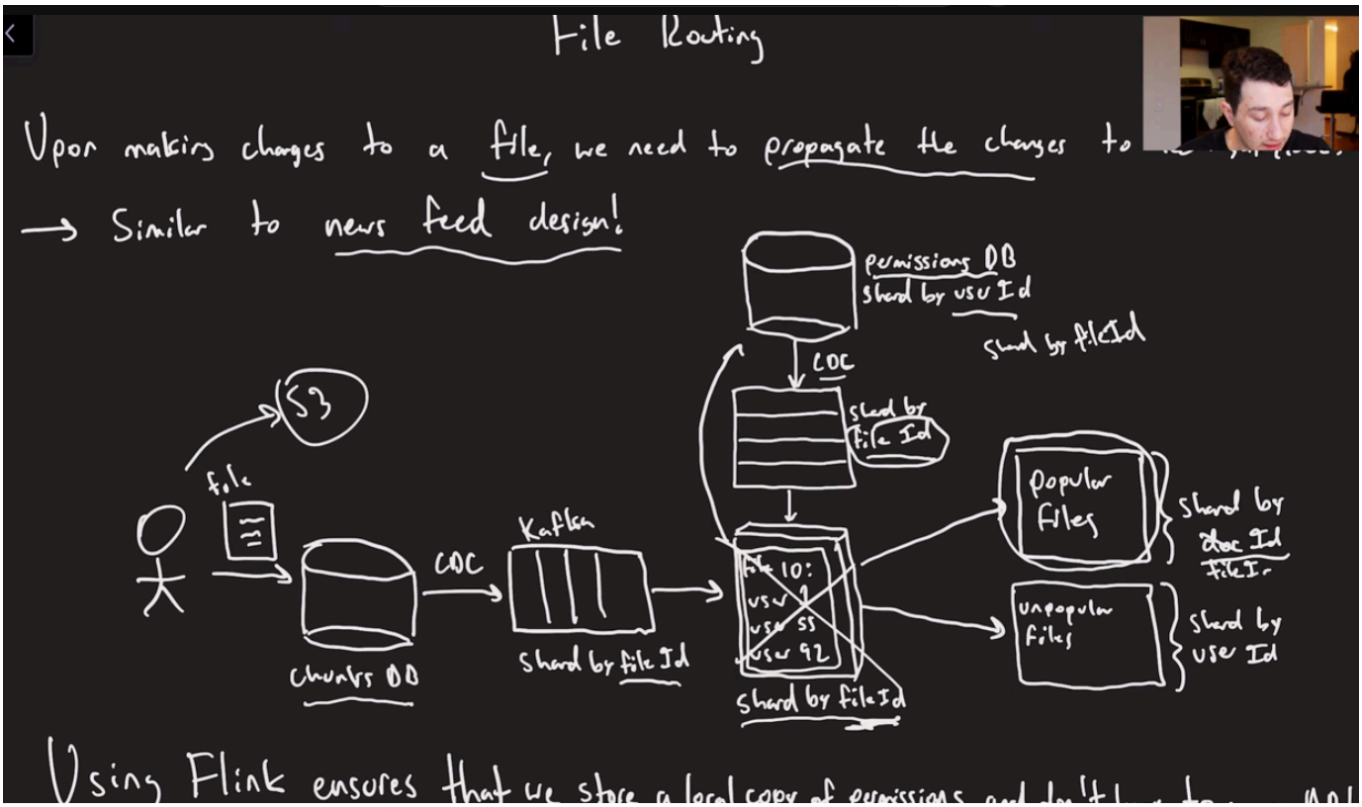


we will upload parallely the chunks to s3 and when they have completed we will upload to the chunks Db.



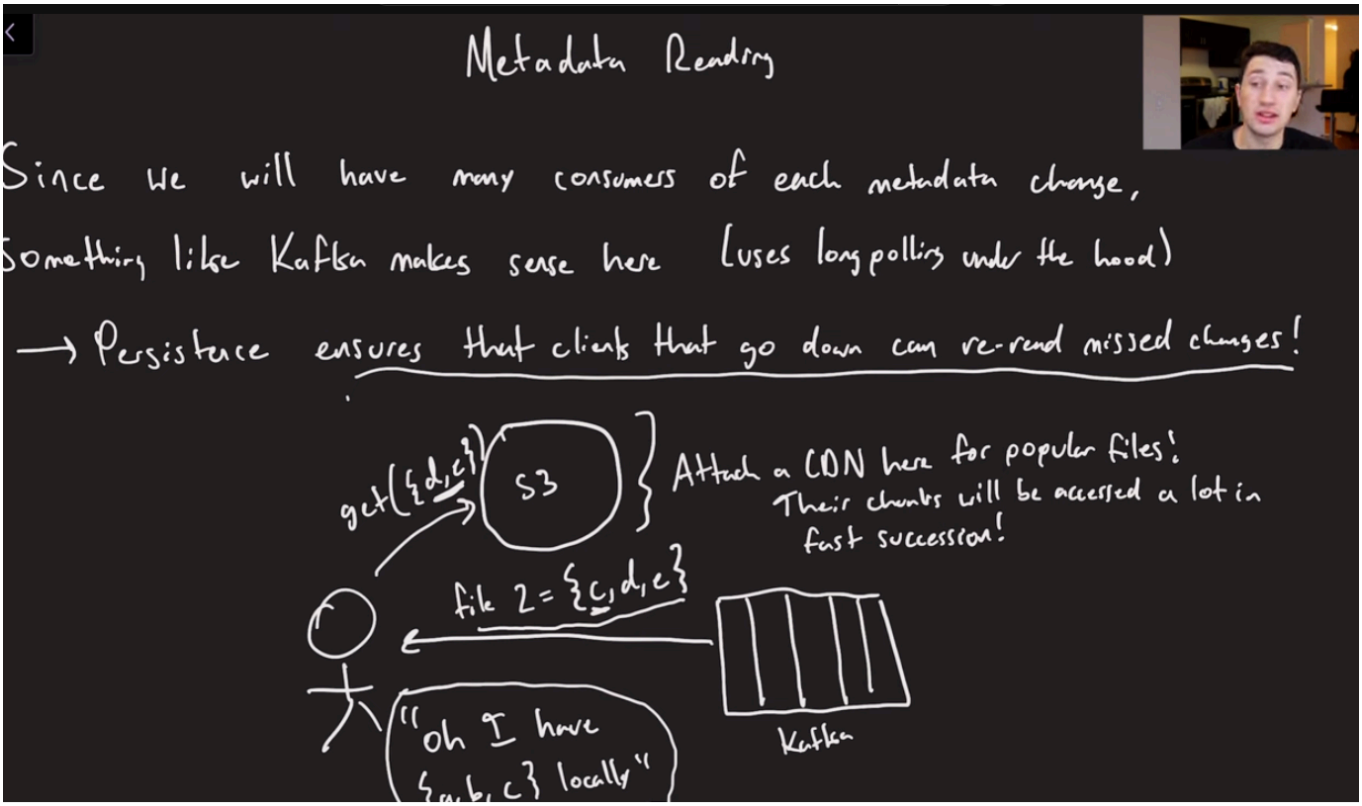


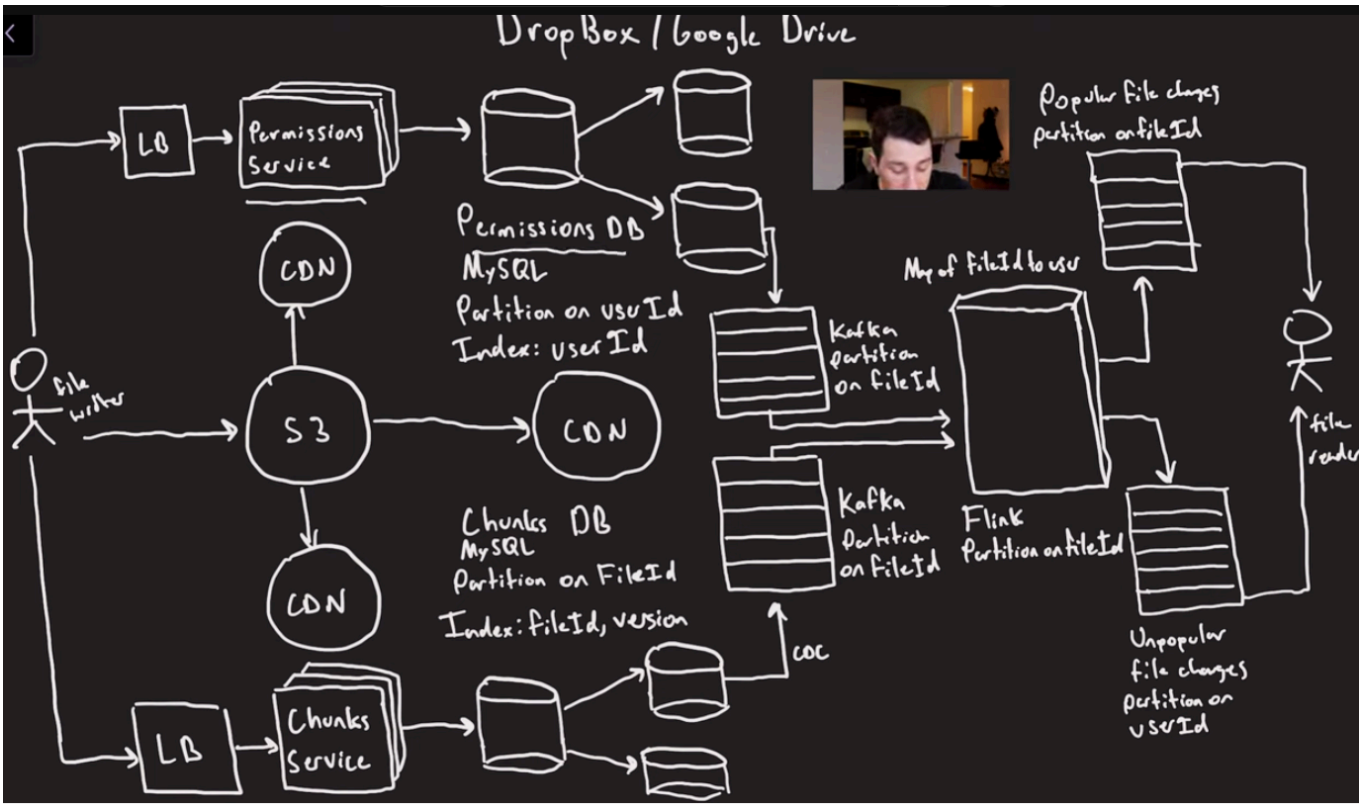
for popular files we will push the new file to a popular document service and the user will open a new connection for that file.
for the unpopular files we will send the file changes to all the users who are using that file.



Using flink ensures that we store a local copy of permissions and dont have to to query Db

Metadata REading




DropBox / Google Drive
LB
Permissions Service
Permissions DB
MySQL
Partition on userId
Index: userId
CDN
file writer
S3
CDN
CDN
Chunks DB
MySQL
Partition on FileId
Index: fileId, version
LB
Chunks Service
CDC
Kafka
Partition on FileId
Kafka
Partition on FileId
Map of fileId to usr
Flink
Partition on fileId
Popular file changes
partition on fileId
Unpopular file changes
partition on userId
file reader