

Upload videos - ingestion
 Feed / timeline of videos - distribution
 Storage of videos
 videos of 10 seconds allowed
 convert into different formats and resolutions

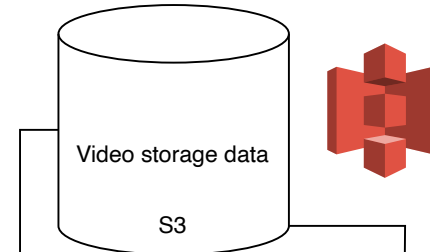
- highly available
- eventually consistent
- once i click on upload btn, the video should be uploaded as soon as possible
- There can a little delay between videos shown to followers after the video gets uploaded
- fault tolerant
- highly scalable

1 billion users in total
 10% DAU - 10 million
 100000 creators
 write to read ratio = 1 : 100
 200 files for 4 format each = 800 files * 3
 resolutions = 2400 files
 200 K videos per day * 100 MB = 200 GB
 200 GB * 3 formats = 600 GB
 600 GB * n resolutions = 1.2 TB a day

- 3 workflow steps in pipeline
1. categorizing the video, if it is longer than 60 sec, just reject it
 2. filtering the video
 3. convert into multiple formats
- parallel processing to create a distributed approach



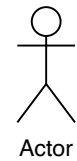
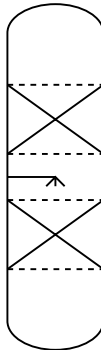
different workers



reliable because its AWS infrastructure
 save time and implementation of creating our own object storage
 immutable form of storage



u



Actor

User data
 video meta data
 video file

YOUTUBE upload SERVICE
 Ingestion

Acknowledgement
 that it might take a
 few seconds

Divide the file into
 different chunks

Chunk 1

Chunk 2

Chunk n

Parallel processing

Format 1

Format 2

Format n

Res 1

Res 2

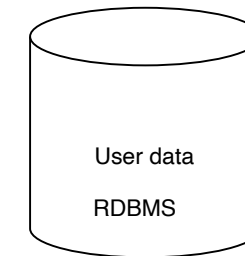
Res 1

Res 2

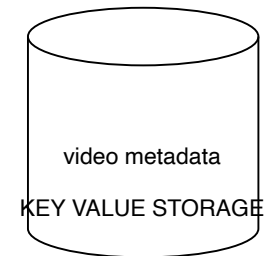
Res 1

Res 2

Chunk aggregator service



ACID properties
 relational database



video metadata is not going to be organised as user data .
 Columns are not fixed.
 no of reads >>>> no. of writes
 low latency
 no need of joins

1 user requests video
 of another user

video serving SERVICE
 to see videos

3

user clicks on a
 thumbnail video

2

multiple files of the same video
 in different formats and
 resolutions

fetch user2 data from user
 table and relevant thumbnail
 from video metadata and
 send to requesting user