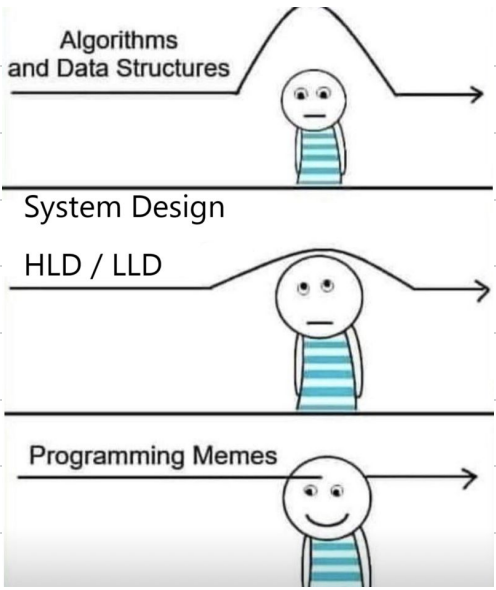
Session-10

ML-System Design-2

Netflix Recommendation Engine

June 25, 2025



My brain when I am sitting in a system design interview

Agenda

- Overview of Netflix Recommendation system-design

Where do we start ??

SWE->

1. Scalability
2. Fault tolerance.
3. Availability

MLOPS – System design

1. Data Versioning / Flow
2. Model Structuring / Components
3. Training – Inference. Skew
4. Concept drift.

System design

1. LLD → Low Level design
   - Implementation details
2. HLD → High Level design
   - architecture

Design a Video Streaming Platform like Netflix

☆ Clarifying Questions

1. 1 Billion Users with roughly 200 million DAU
2. 1 million videos → on platform
3. 1000 Vidos uploade any given day
4. User access 5 times in a day

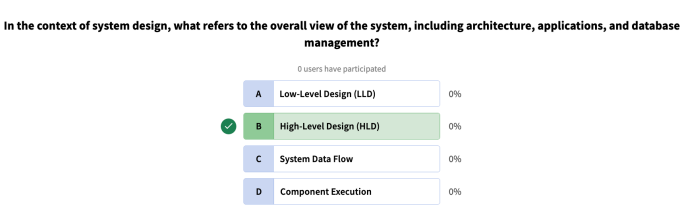
$$\frac{200 \times 10^{6} \times 5}{24 \times 60 \times 60} = 12K \text{ req/second.}$$

1 Functional Req

a. Users should be able to stream the video on platform (pause/resume video playback).
b. People should able to search video on the platform.
c. Personalized Recommendation for users.
d. Producers, Directors, and Teams should be able to upload content on the website.
e. It should be available on iOS/ Android/ Web.
f. Users should be able to like/dislike/rate/comment the video.

2 Non - Functional Req.

1. U- Low latency & High availability
2. Scalable & Efficient.



What is an essential aspect of system design in regards to understanding its scale?

0 users have participated

A	Estimating the color palette for the UI	0%
B	Programming language to be used	0%
✓ C	Making reasonable assumptions about system metrics like requests per day or database calls	0%
D	Selecting a font style for documentation	0%

End Quiz Now

Leaderboard

Based on all quizzes from the session

2	RS	Renu Singh — 2/2 — 187.03
1	SK	Santhosh K — 2/2 — 188.23
3	A	Akshara Shukla — 2/2 — 184.03
4	VS	Venkata Sai Pramod Kumar — 2/2 — 180.47
5	PW	Popatsing Waghela — 2/2 — 178.69
6		Arindam Maji — 2/2 — 178.20
7	L	Lokesh Sonawane — 2/2 — 171.80
8		Gajanan Todeti — 2/2 — 171.39
9	k	Phani Kishore R PMP — 2/2 — 170.76
10	CS	Chirantan Sharma — 2/2 — 169.42

Different components we need to design for Netflix (MVP)

1. Data Sources
2. Data Ingestion Pipeline
3. Data Storage Systems
4. Data Processing and Feature Engineering
5. Recommendation Engine (building ML algorithm)
6. API Layer
7. Caching
8. Monitoring and Logging
9. Security And Compliance

1 Data Source

a Users / User-item interaction.

(i) Like / dislikes / reviews / ratings / comment
(ii) Watch history
(iii) Search history
(iv) User interaction with platform
(v) wish list / watch list
(vi) Daily time spent
(vii) Age, gender, demographics / Language.

b Content / Content-metadata.

(i) Movies / Shows - (Genre, director, title, cast, duration, description, Rating) (Kids, 18+)

(ii) Release year

(iii) Content popularity metrics
↳ IMDB
↳ Rotten tomatoes

(iv) Technical specifications → 1080 / UK / 480p

(c) External data

1. Social media trend / Sentiment
2. Third party reviewers

2. Data Ingestion Pipeline

Client

1. Information regarding your device / its usage
2. Wakeup / Shutdown
3. Click stream data

Server

1. Which results were shown / served.
2. What was clicked on.

```
"userId": "user123",
"sessionId": "session456",
"events": [
{
"eventType": "view",
"page": "/home",
"timeSpentSeconds": 10,
"timestamp": "2025-01-27T10:00:00Z"
},
{
"eventType": "scroll",
"scrollPosition": 300,
"timestamp": "2025-01-27T10:00:05Z",
"page": "/home"
},
{
"eventType": "click",
"elementId": "signup-button",
"timestamp": "2025-01-27T10:00:10Z",
"page": "/home"
},
{
"eventType": "view",
"page": "/product/123",
"timeSpentSeconds": 30,
"timestamp": "2025-01-27T10:01:00Z"
},
{
"eventType": "form_submission",
"formId": "contact-form",
"fields": {
"name": "John Doe",
"email": "john.doe@example.com"
},
"timestamp": "2025-01-27T10:02:00Z",
"page": "/contact"
}
```

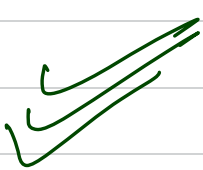

3 DATA STORAGE

Structur SQL

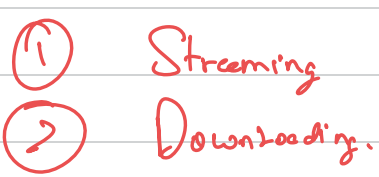
SQL

Non-Structur

NC SQL



```
// Electronics Product
{
"product_id": 1,
"name": "Smartphone X",
"category": "Electronics",
"price": 999,
"attributes": {
"batteryLife": "24h",
"warrantyPeriod": "2 years",
"brand": "TechBrand"
},
"reviews": [
{
"user_id": 101,
"rating": 5,
"comment": "Excellent phone with great battery life!"
},
{
"user_id": 102,
"rating": 4,
"comment": "Good performance but a bit pricey."
}
]
}
// Clothing Product
{
"product_id": 2,
"name": "T-Shirt Y",
"category": "Clothing",
"price": 29,
"attributes": {
"size": "L",
"material": "Cotton",
"gender": "Unisex"
},
"reviews": [
{
"user_id": 103,
"rating": 4,
"comment": "Comfortable and fits well."
}
]
}
// Books Product
{
"product_id": 3,
"name": "Novel Z",
"category": "Books",
"price": 15,
"attributes": {
"author": "Author A",
"ISBN": "123-4567890",
"genre": "Fiction"
},
"reviews": []
}
```



Content Delivery Network

Local data Storage

Mumbai CDN

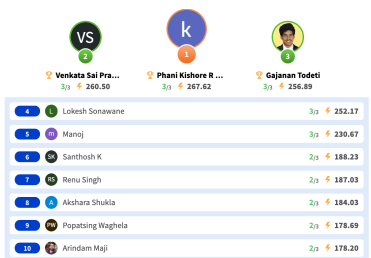
Myntra → CDN → SCR day/month

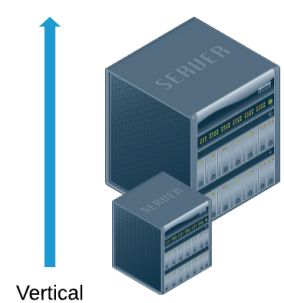
What is the main difference between horizontal scaling and vertical scaling?

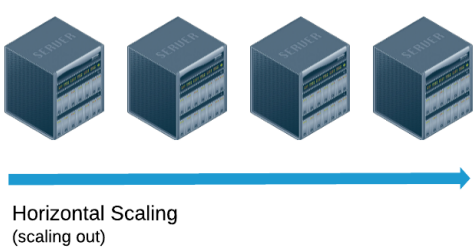
0 users have participated

A	Horizontal scaling adds more machines, while vertical scaling adds new software.	0%
B	Horizontal scaling adds physical locations, while vertical scaling focuses on security enhancements.	0%
C	Horizontal scaling adds resources to a single machine, while vertical scaling adds more machines.	0%
✓ D	Horizontal scaling adds more machines, while vertical scaling adds resources to a single machine.	0%

End Quiz Now







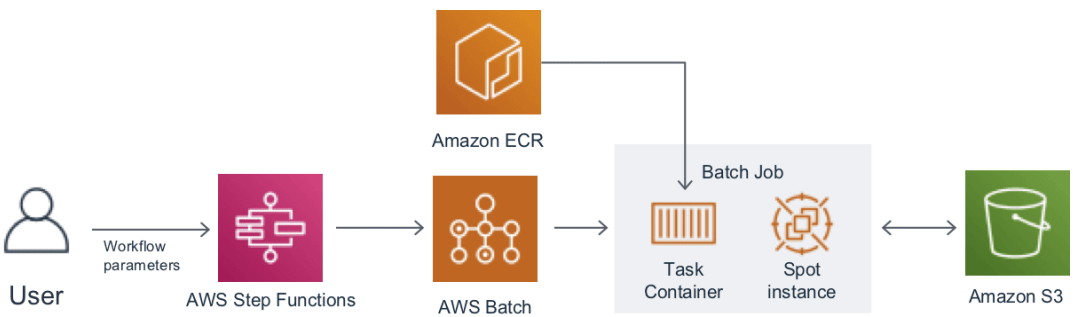
Serverless

not serverless

EC2 instance

ECS instance

serverless



workflow

automation service

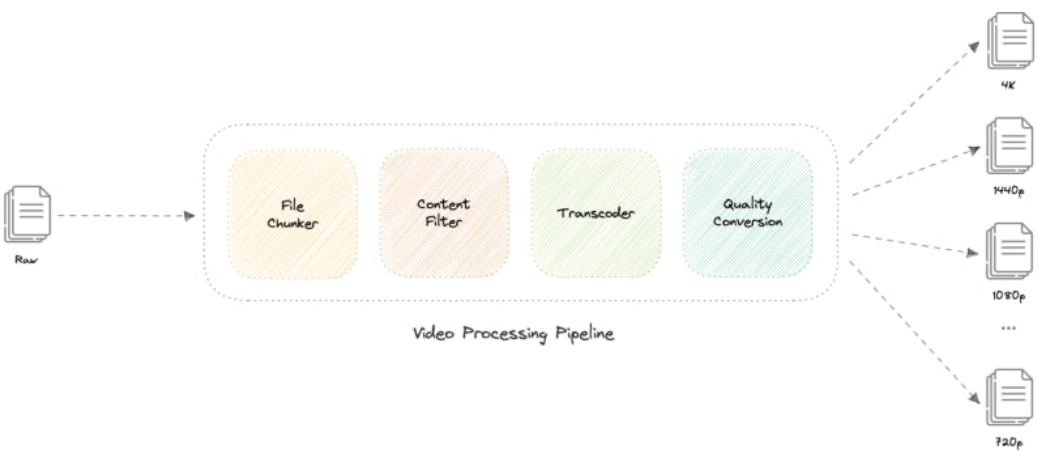
Transcoding

Video → Divide into chunks — 4K
- 1080
- 720
- 4K

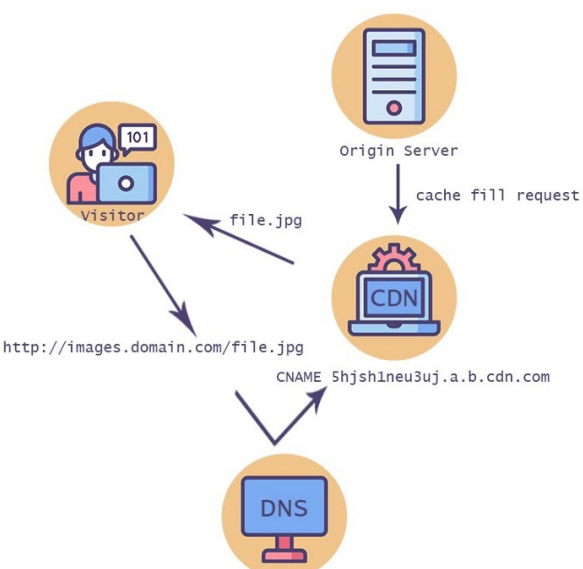
1

C1, C2, C3 — — C10

2



Video Resolution	Video Bitrate
240p	300 kbps
360p	500 kbps
HD 480p	1000 kbps
HD 720p	1500 kbps
HD 720p	2250 kbps
Full HD 1080p	3000 kbps
Full HD 1080p	4500 kbps
Quad HD 1440p	6000 kbps
Quad HD 1440p	9000 kbps
4K UHD 2160p	13,000 kbps
4K UHD 2160p	20,000 kbps



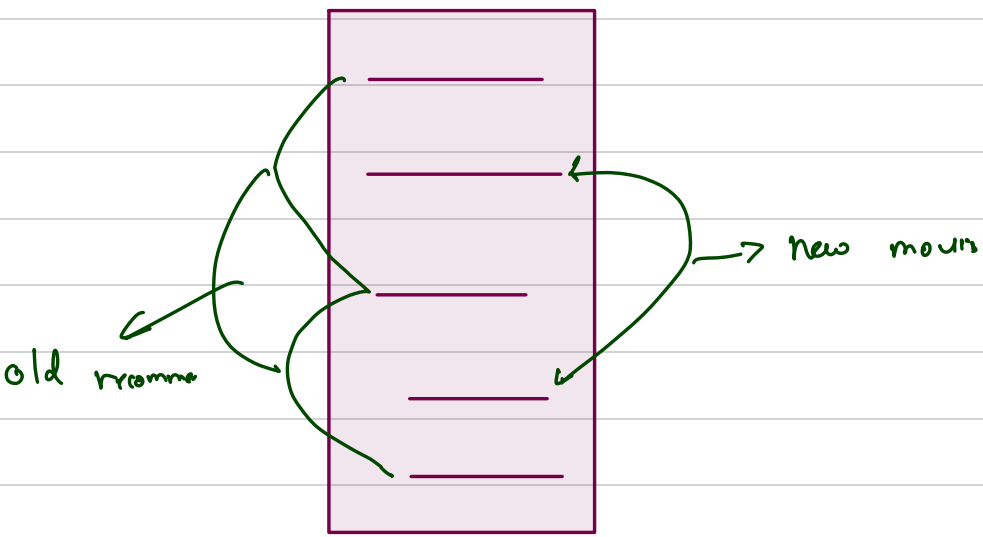
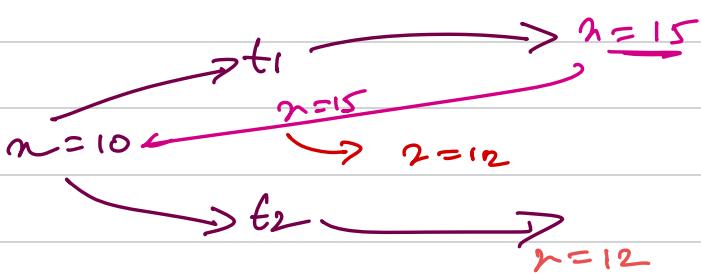
What is the role of a Content Delivery Network (CDN) in video streaming?

0 users have participated

A	To slow down the delivery of content to manage server load	0%
B	To provide a single centralized server for content storage	0%
C ✓	To offer geographically distributed servers for fast content delivery	0%
D	To increase internet connection speeds universally	0%

End Quiz Now

based on all quizzes from the session

2	Venkata Sai Pra...	4/4 352.36
1	Phani Kishore R ...	4/4 360.79
3	Lokesh Sonawane	4/4 344.50

4	m	Manoj	4/4 304.83
5	SK	Santhosh K	3/4 284.06
6	A	Akshara Shukla	3/4 271.89
7	PW	Popatsing Waghela	3/4 269.49
8		Gajanan Todeti	3/4 256.89
9	SG	Sonali Ghosh	3/4 248.00
10	RS	Renu Singh	2/4 187.03