

Session -7

ML - SYSTEM

DESIGN -1

Jan 19, 2024

WHEN YOU'VE PREPARED JUST DSA FOR INTERVIEW



Agenda

- * What does ML system design includes.

Netflix recommendation System

STEPS → ML SYSTEM DESIGN

Steps:

① Clarify (in terms of what you want)

- What data do we have?
 - Some latency (Latency)
 - minimum movies a user has watched before we serve
 - list of 100 movies recommendation
- ①
- similar. 80:20 → new tgt

② Data:

- ① User: details → Watch history
- demographics
 - average short
 - age
 - gender
 - Review / Rating (Feedback)
 - Watchlist

→ Daily watch time

② Movie meta data → director

→ actor

→ genre

→ length

→ fears of release

→ rating

③

User - content date → Time of the day

when you
watch

→ Device

→ locations

→ Day of week
- Sunday

④

External data → Publicity / Premise

→ News

→ Festival

Client Side data / Server side data



All the logging data → client side

device-id

When you req.
what?

Phone model → what server suggest

Union

device-id

Wake-up/shut-down time

→ what

assigned

Time you watched

was

harder to track on PC

before dial you click

your response

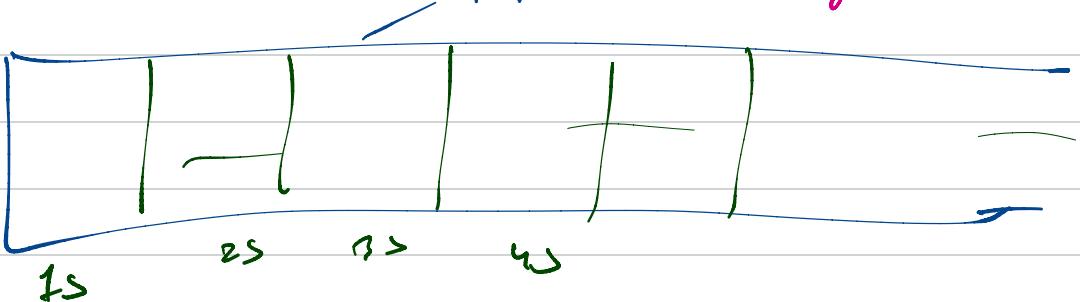
as click-stream
data

PC

Sent to Server

topic

Streaming data



S1

S2

S3

S4

fins

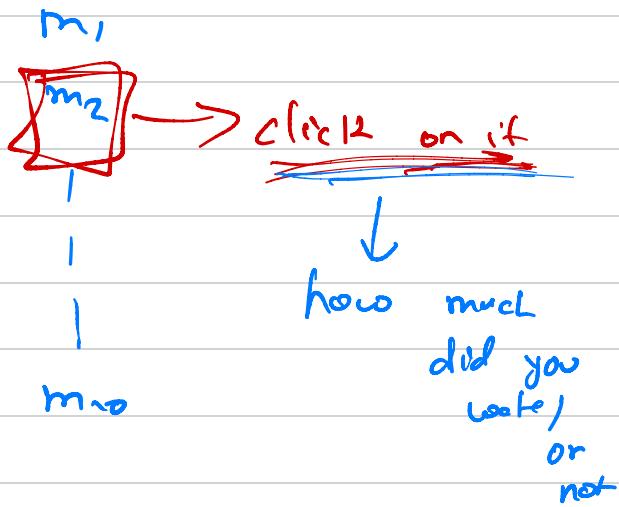
P1

P2

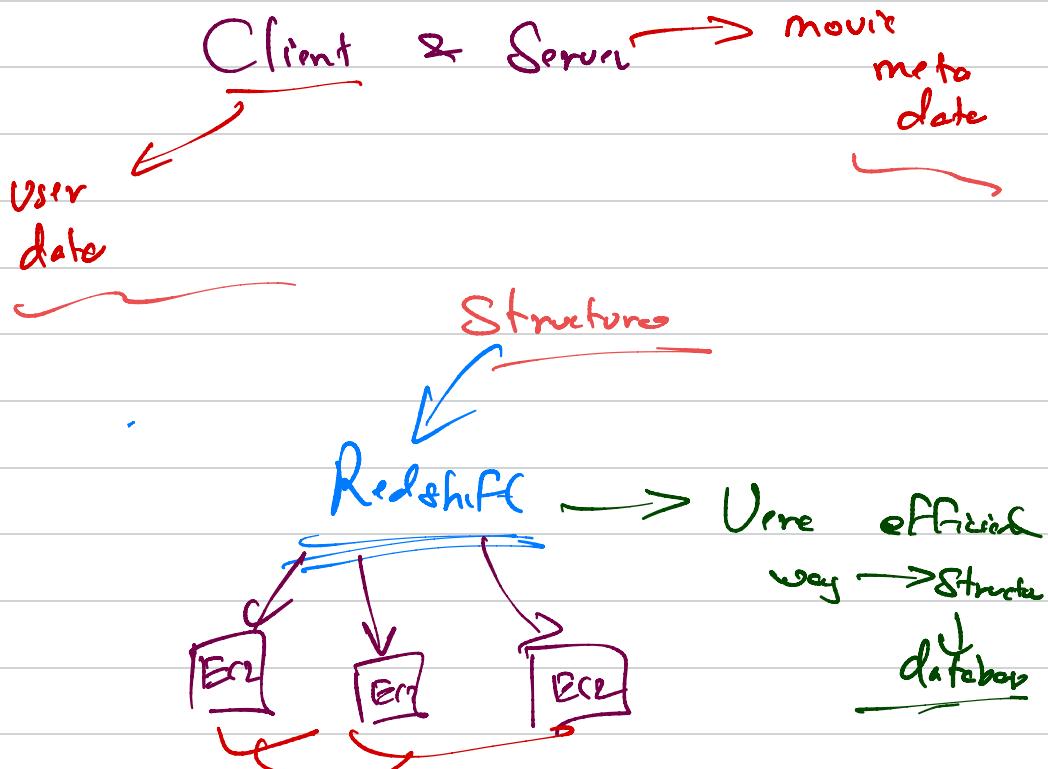
P3

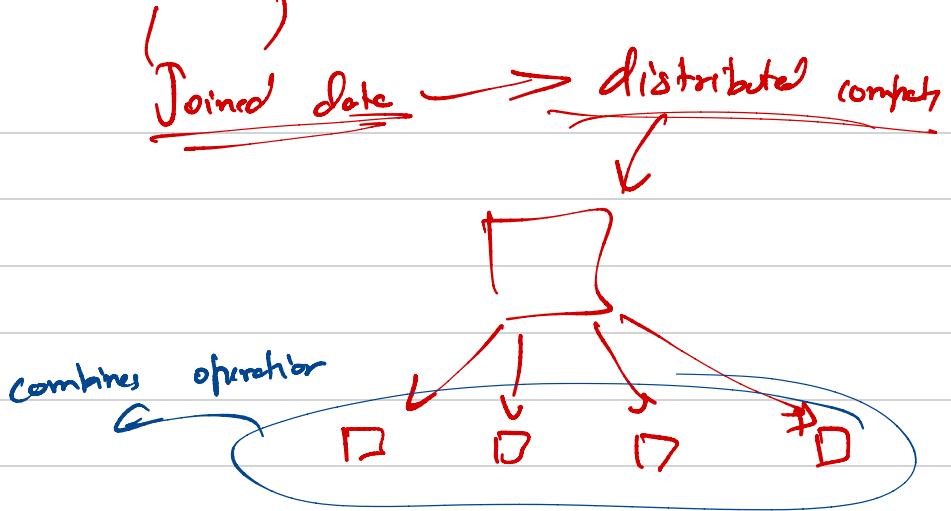
P

Explore →



→ Now you're joined

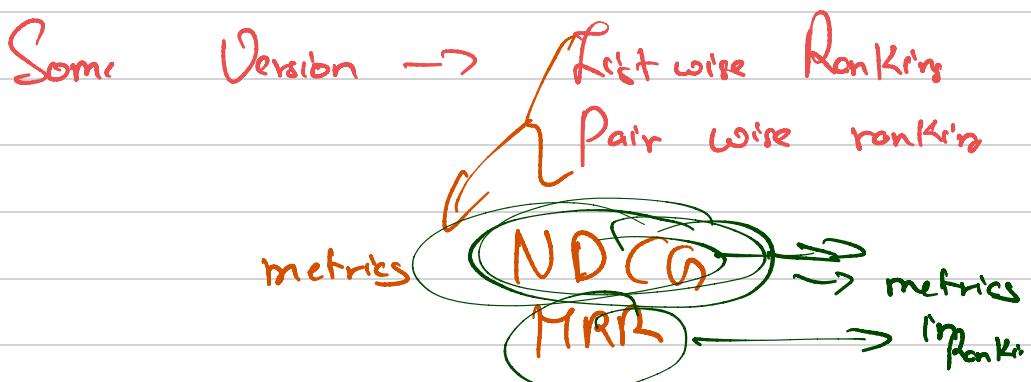




No-SQL \rightarrow Dynamo DB (Provider by AWS)

② Feature Engineering:

③ Model Selection : \rightarrow Collaborative - Filtering
 \rightarrow Content-based Filtering



→ XG BOOST / In 2022
→ LGBM MHT recomm.

Metric would you use →
→ F1-score
→ Precision@12
→ NDCG
→ MRR

OPTUNA

$$\text{max-depth} = [1, 3, 5, 7]$$

$$n_estimator = [100, 200, 300]$$

time → 100

CICD

① Push to docker

② Deploy it to ECS / Push it to ECR

→ A/B Testing → latency thru
→ stability
→ Online metric score ??

→ Drift

Output of model + Custom weights
for component accuracy

