

## Assumptions

INPUTS SOURCES - n/w  
- pipe  
- dump

TYPES OF DATA - plain text, json, xml, csv  
relbms, document, unstructured, logs,

DATA DISCOVERY - Automated, Cyclic, Batch, ~~SPAR~~  
- Online / offline

DATA PROCESSING - Batch, Real Time, Almost Real Time,  
PRE - n - Mini Batch

SERIALISATION - Avro, Kryo, Pickle etc

STORAGE - HDFS, RDBMS, NO-SQL

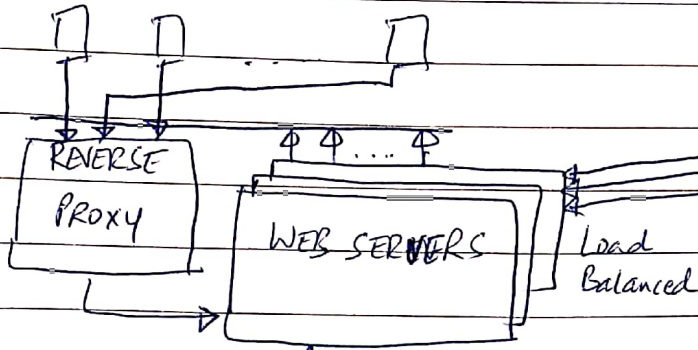
FILE COMPRESSION - PARQUET,

ENCRYPTION

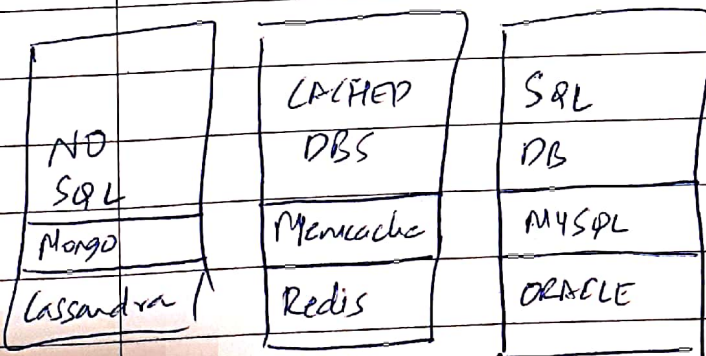
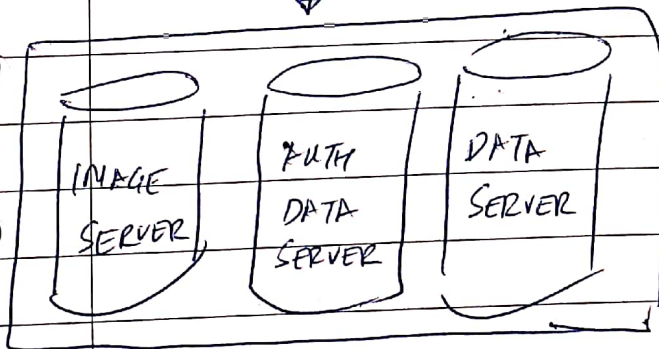
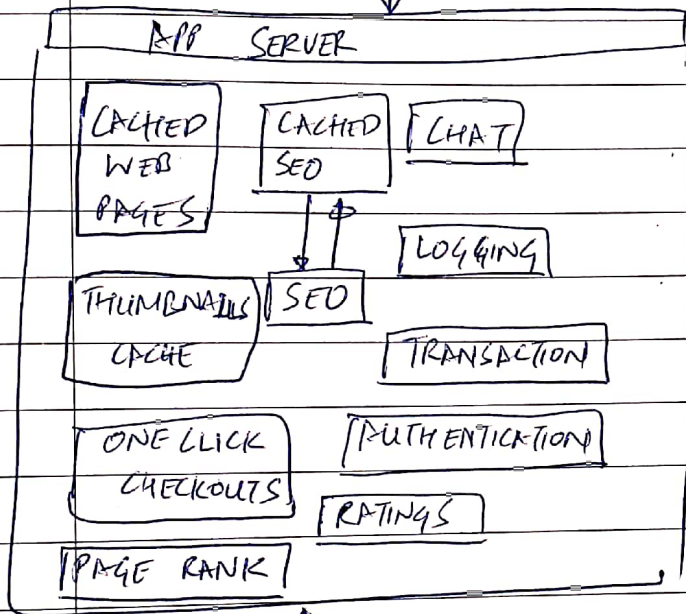
DATA AVAILABILITY TYPES - Back End, Front End, Click data,  
logs,

# HIGH LEVEL

DESKTOP  
CLIENTS

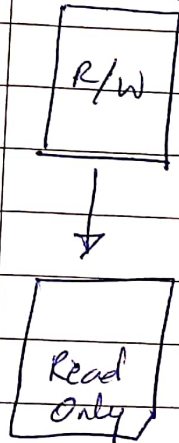


MOBILE  
CLIENTS

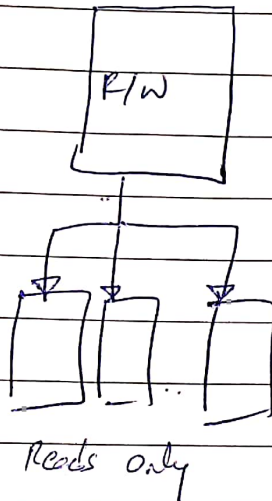


## DATA BASE SCALING (CAP Theorem)

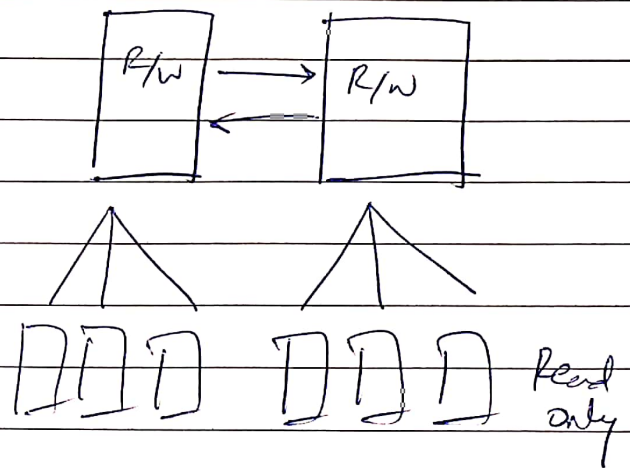
Read only v/s  
Read-write



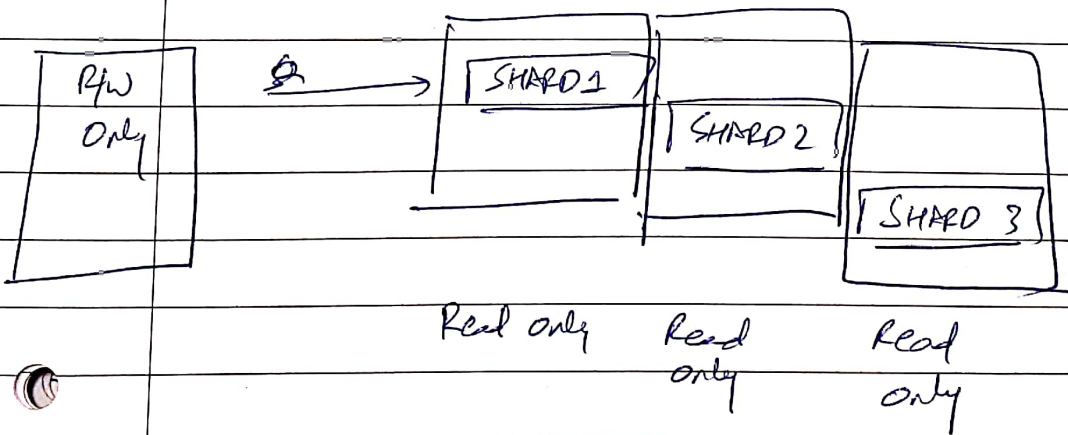
Replication



M/S Replication



SHARDING

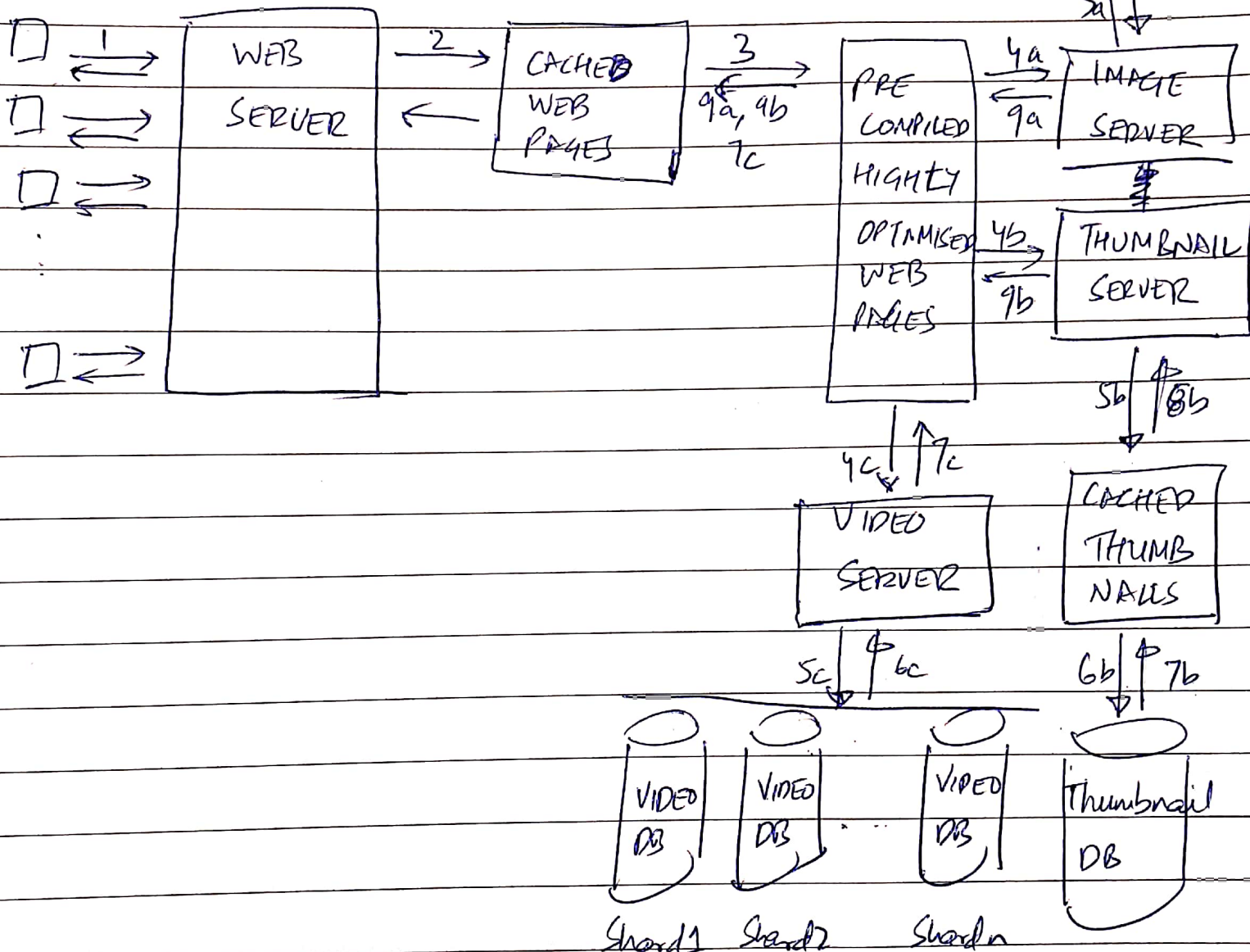


# WEB SERVER REQUEST

Pre-compiled optimised rendering

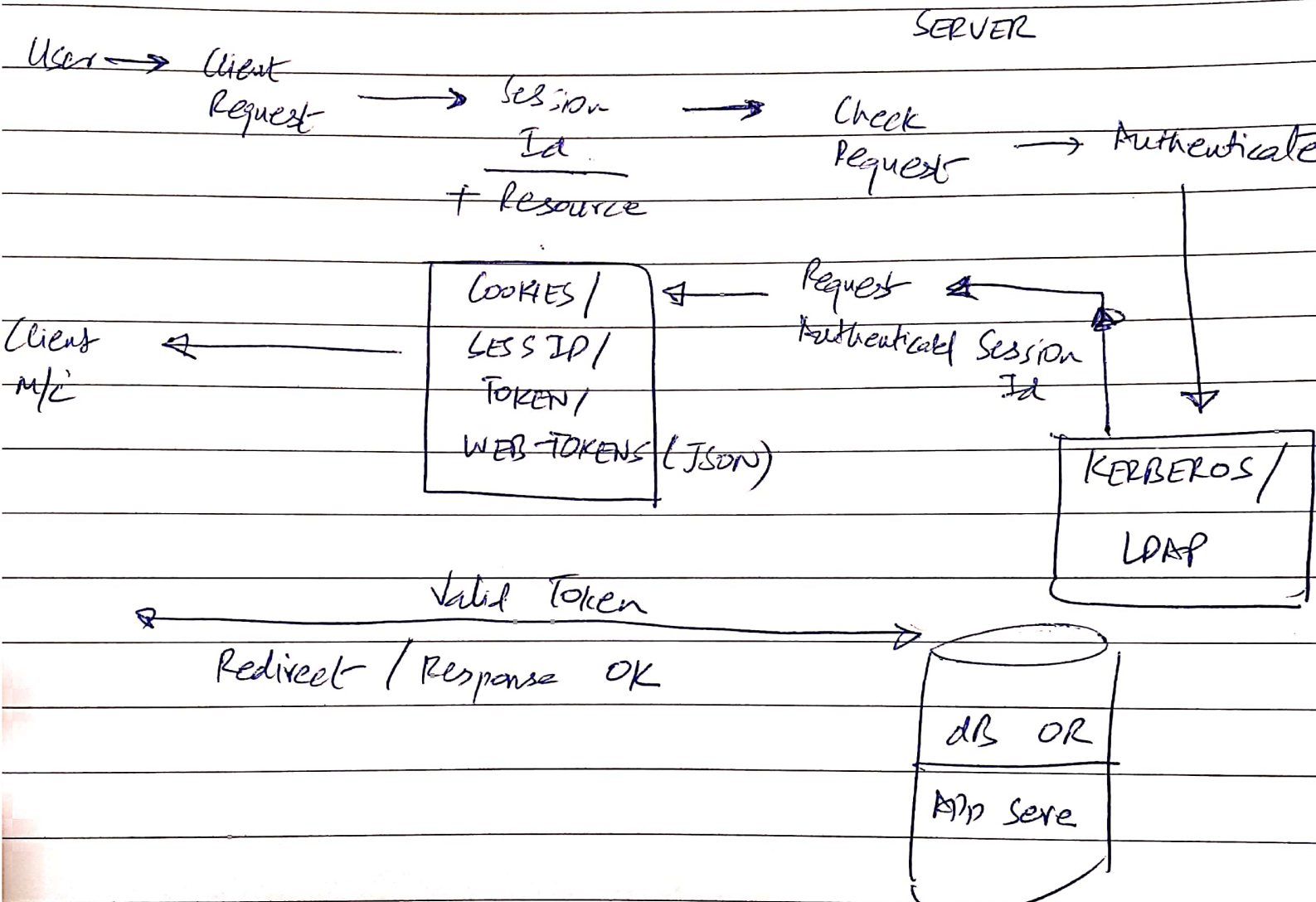
Cached web pages.

PHP / Thrift as an example





# AUTHENTICATION #



## CONSIDERATIONS

DATA SOURCES

DATA STREAMING TYPES

BATCH / OFFLINE PROCESSING  
MINI BATCH / ALMOST REAL TIME  
STREAMING / REAL TIME

DATA PREPARATIONS

DATA SEGREGATION

MODEL TRAINING

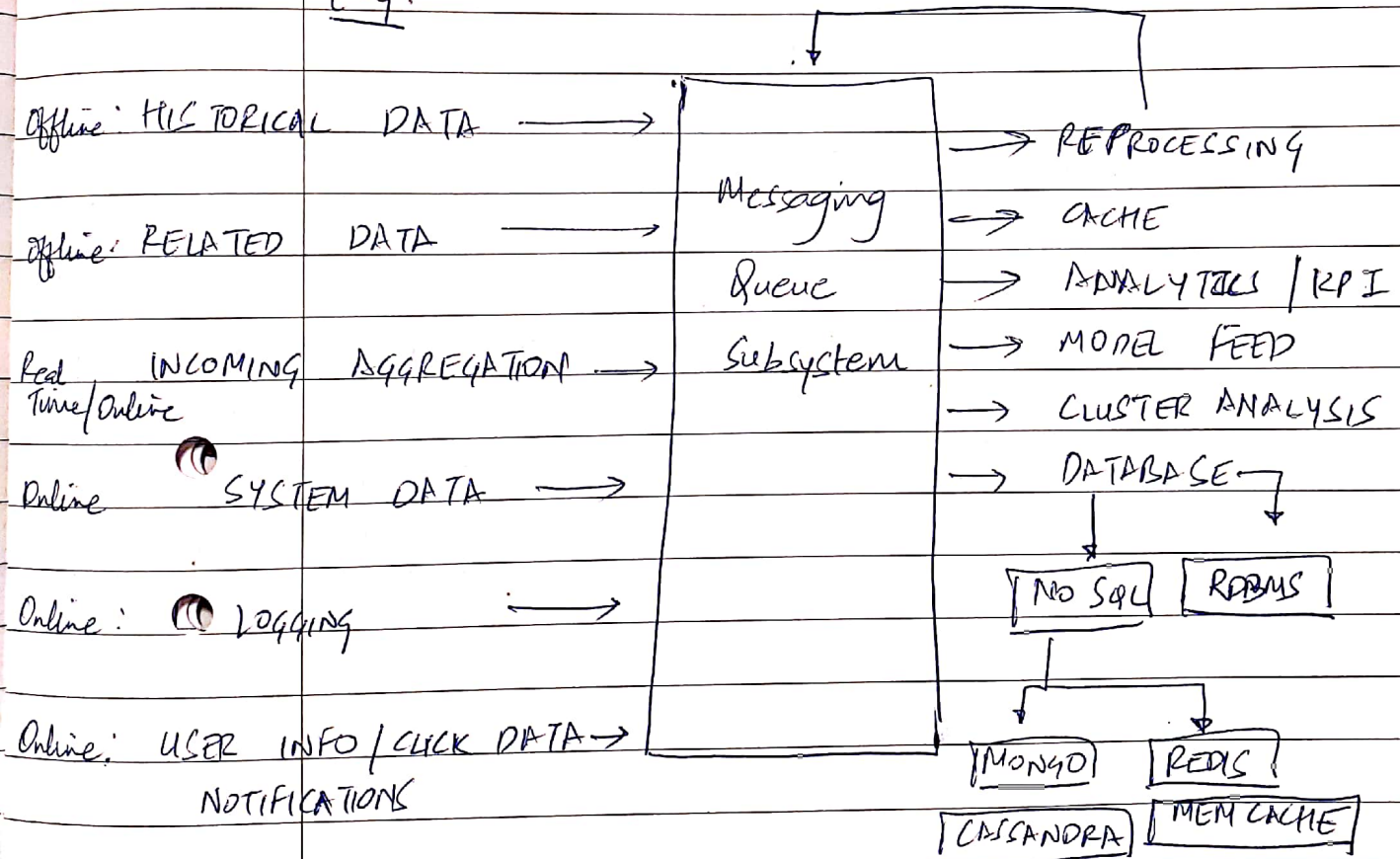
MODEL EVALUATION

DEPLOYMENT

MONITORING

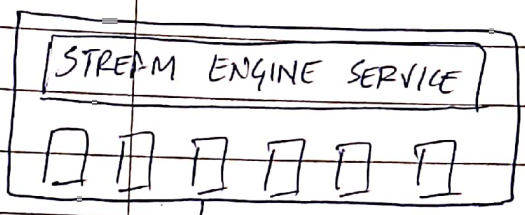
PERFORMANCE CONSIDERATIONS.

E.g.

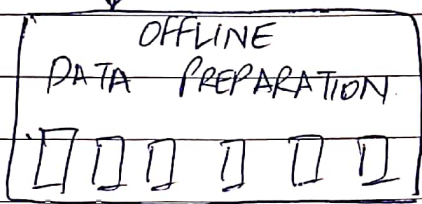
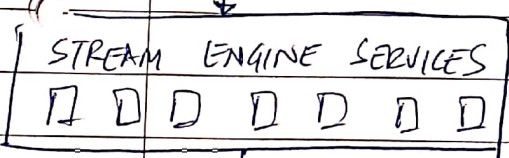
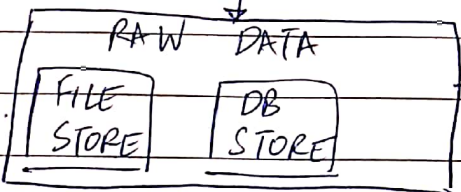


REAL TIME DATA

BATCH PROCESSING DATA



BULK DATA TRANSFER



ONLINE DETAILED DATA PREPARATION

ONLINE FEATURES DATA STORE

~~OFFLINE DETAILED DATA PREPARATION~~

OFFLINE FEATURES DATA STORE

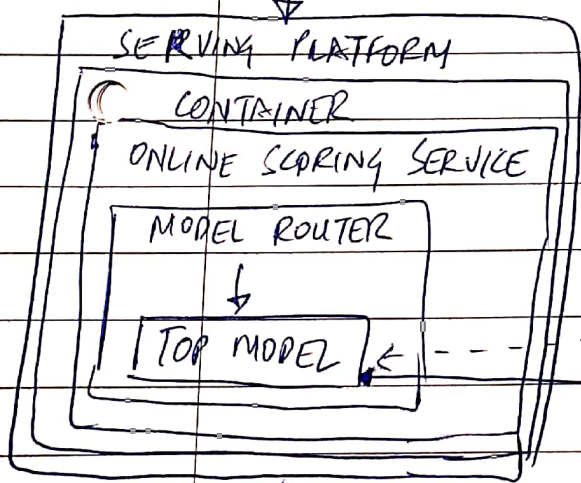
DATA SEGREGATION API

TRAIN SERVICES

TEST SERVICES

MODEL DATA STORE

MODEL EVALUATION SERVICE



CLIENT APP

MODEL METRICS STORE

