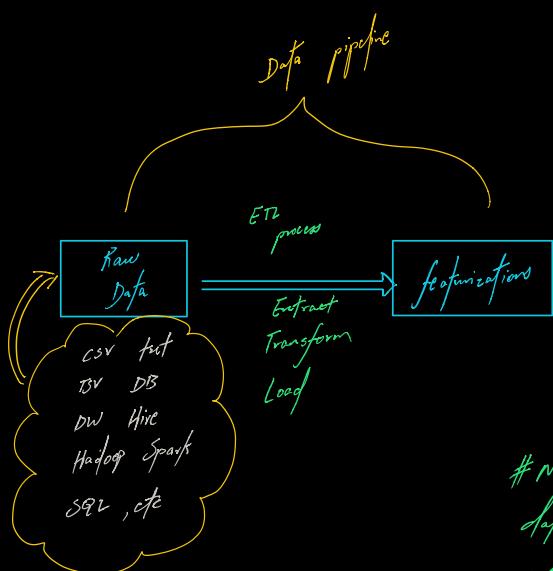
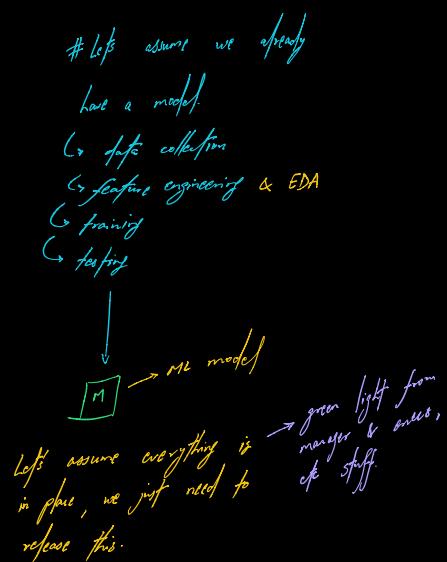


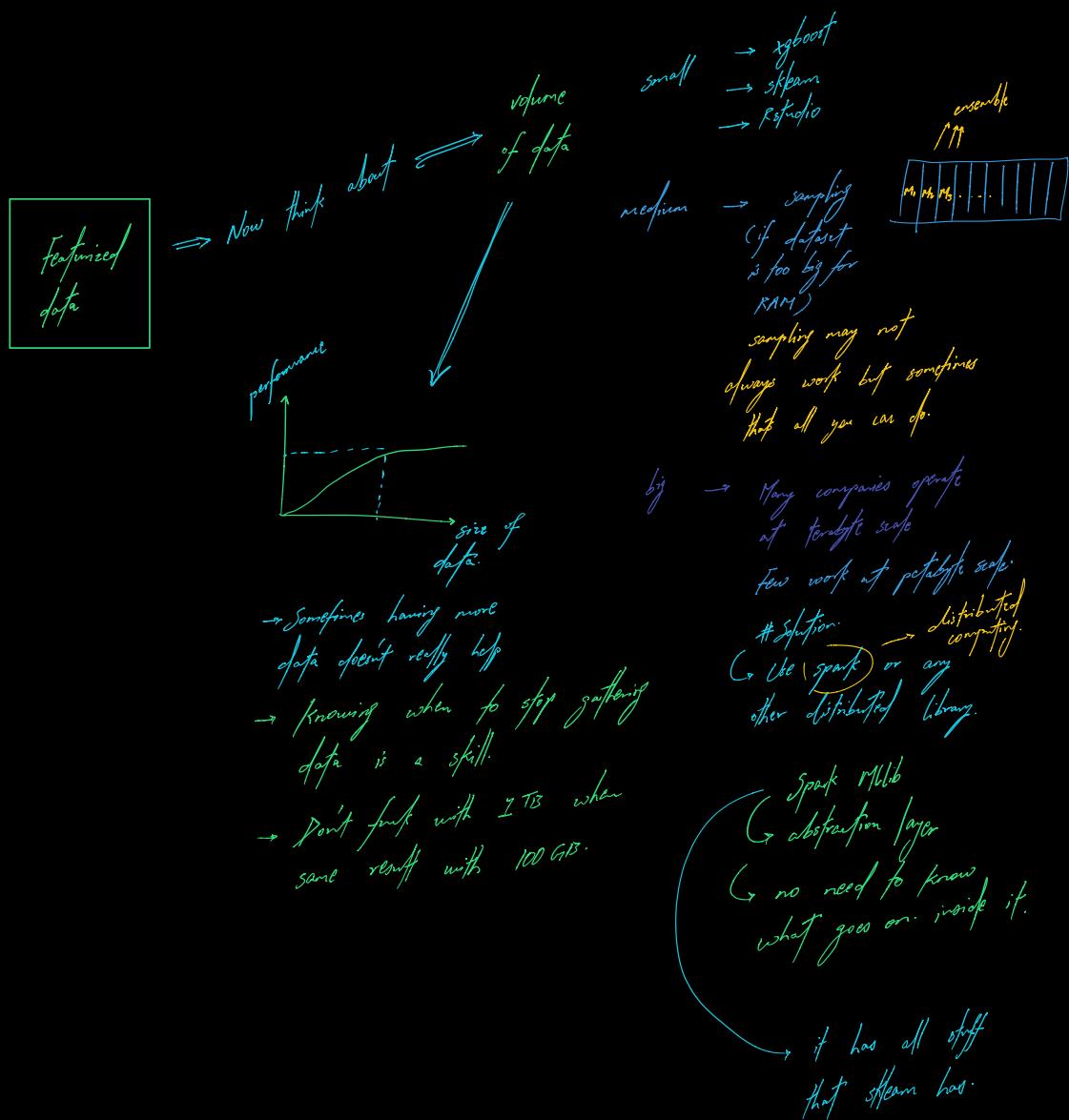
stuff covered:
 ML models
 in production
 env.
 # challenges
 # design choices
 # more difficult than the overall

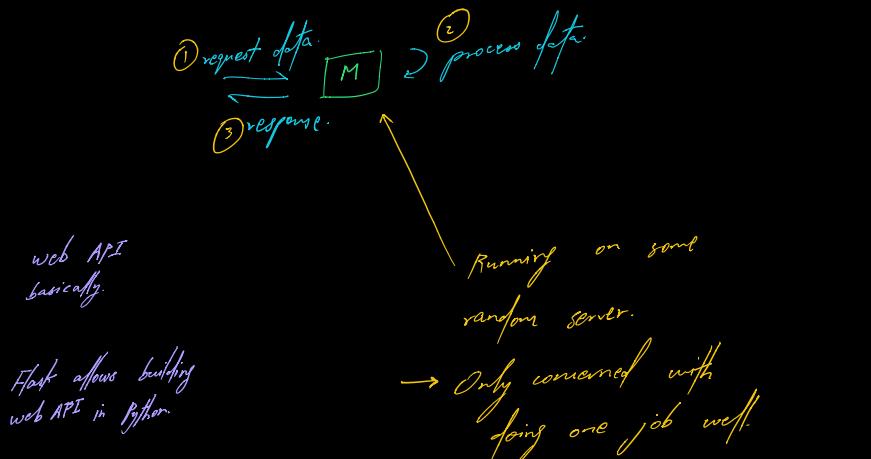


normally we have
 data engineers that
 make the ETL pipeline
 # but if company can't
 hire them, DS and us

doing this process

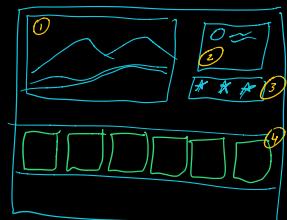
- * what's the frequency of running this ETL pipeline?
- * Big sales company might run this every hour
- * Some might want data every day.





- # Benefit of API
 - ↳ The client and server can be running on completely different languages and it would not matter
 - ↳ They have a common language to talk with each other → JSON most popular

ML use case → { "datapoint":
 { "f1": 95,
 "f2": 48
 } }



①, ②, ③, ④ all can be different APIs. That's what happens in real life.

People know MLP CNN RNN by just applies when asked to build a recommender or chatbot or phone, etc.

THAT'S WHAT FUCKING HAPPENS WITH ME!!

use profile library in Python
to sacrifice objects. → doesn't matter what
objects they are.

- * Hadoop & spark
are not important to learn for interviews.
- * People with good foundations can always
learn and catch up.
- * It doesn't hurt to learn though, in case
we have time.

selenium → gets the html for us
& then we can reason the
fact out of it.
websites don't tell when people
scrap data, they block us from
to be confident.

- * They kinda don't give a fuck about
the extra bits of knowledge we gather.

ETL pipelines can be
asynchronous → "streaming ETL pipeline"
↳ only do it if needed

It's not about GCP or AWS or
Azure, it's about how we use them.

{
↳ if some tool makes
your life simple, fucking use it...