

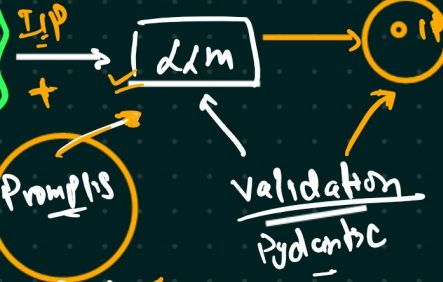
COREFUN → Application ← Production ready

① Document- Analyse

② Document- Compare

2 → more than two doc

Parse the data from doc (pdf)



③ Chat module

Single doc chat

multidoc chat

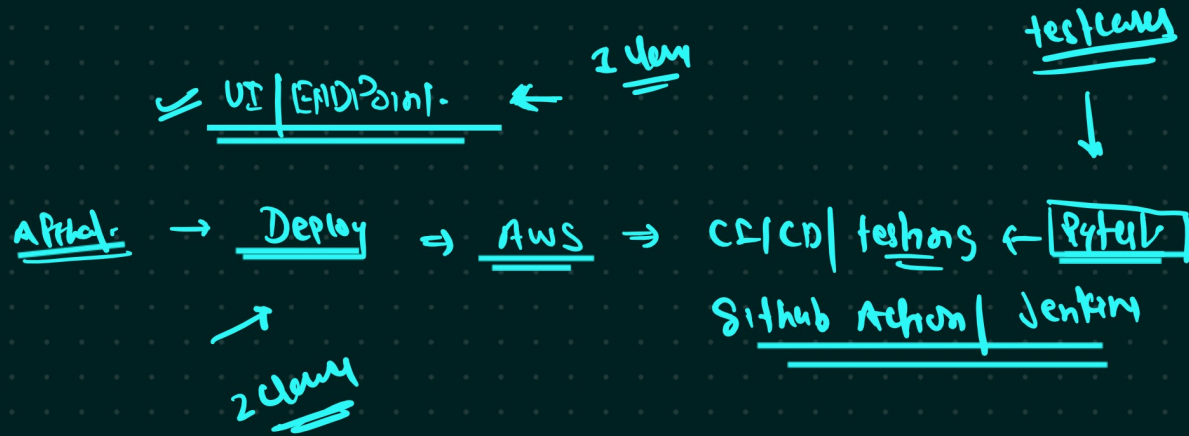
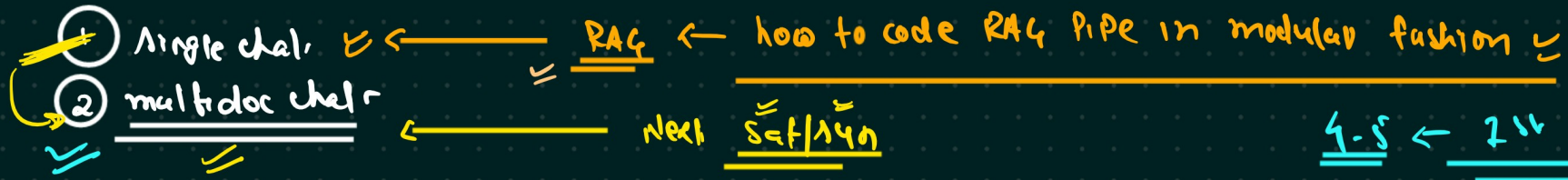
RAG

multiple pdf ⊆ multiple pdf  
multiple diff type of doc

RAG

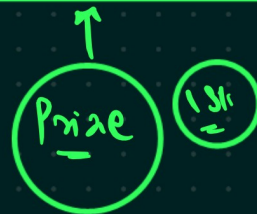
Evaluation  
history ← memory  
Catche  
Streamily.

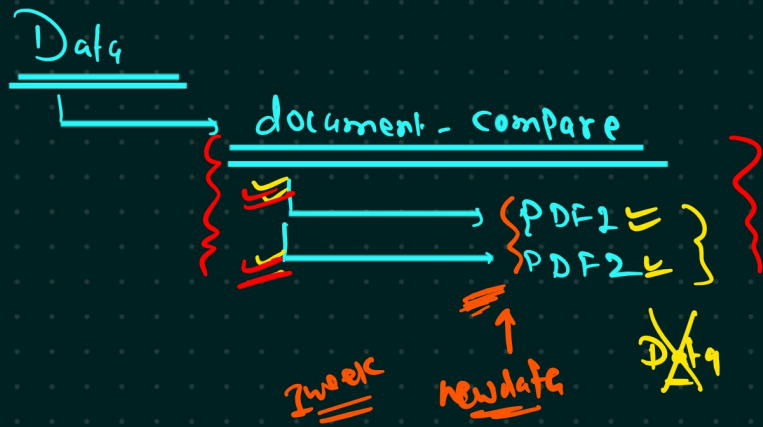
## chat module



{ Resume → you can start applying }

Interview question / Mega Assignment





Versioning

execute my pipeline

session

session folder

Data

{universal identification no.}

↑  
UUID ← module

1 month (manager) → Sunny → yesterday's data

execution → (previous session) ←  
- what data?

one day ⇒ pipeline ⇒ 10  
↓  
(30, 40, 50, 100, 250)

logic → how many total session = 3  
↑  
last 3 session

last 3 session  
↑  
latest

300 → latest 250 ← Artifact



## PyMuPDF



< PyMuPDF / Fitz >

Raw text



summary  
fetch → question

very huge  
① hallucinate  
② token  
③ cost

doc. analysis  
doc. compare



## PyPDF loader

### Conversational AI Pipeline



oip ?

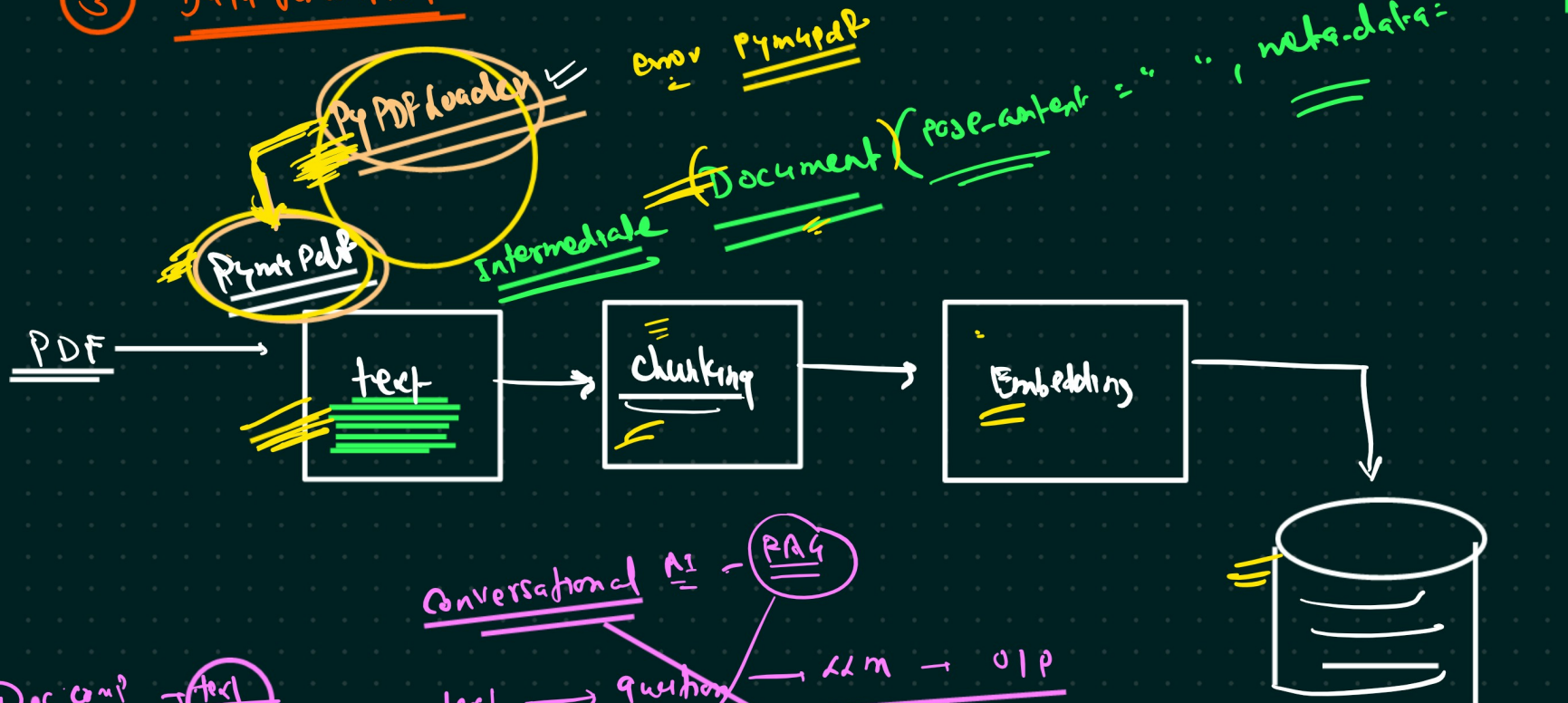
no if will not work

RAQ

1 Data ingestion

2 Data Retrieval

3 Data Generation



Doc comp - text  
Doc analysis

Conversational AI - PAG  
text → question → LLM → O/P

launcher wrapper

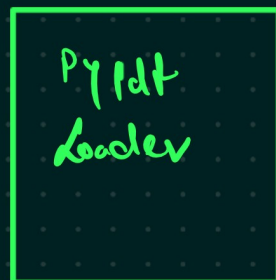


PyPDFLoader ✗

manual code

logic → from scratch

PDF



Page 1 →

Page 2 →

Page 3 →

Page 4 →

Page 5 →

