Llama Index v/s Lang Chain

Topic needs to Learn:
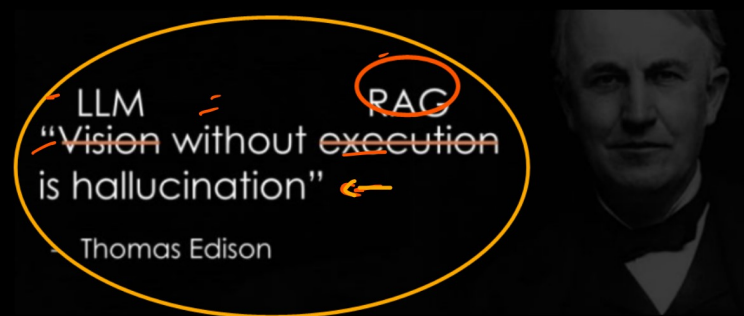
1. What is Llama Index? ✓
2. What is Lang Chain? ✓
3. What is a difference between Llama Index and Lang Chain
4. Where we should which one?
5. Can we combine both for our application building?

image/text/r

Vector Embeddings & Vector Database:

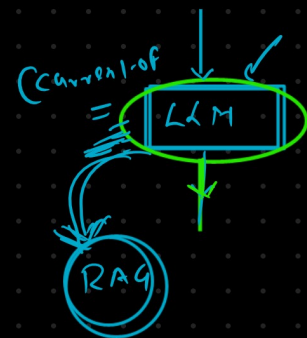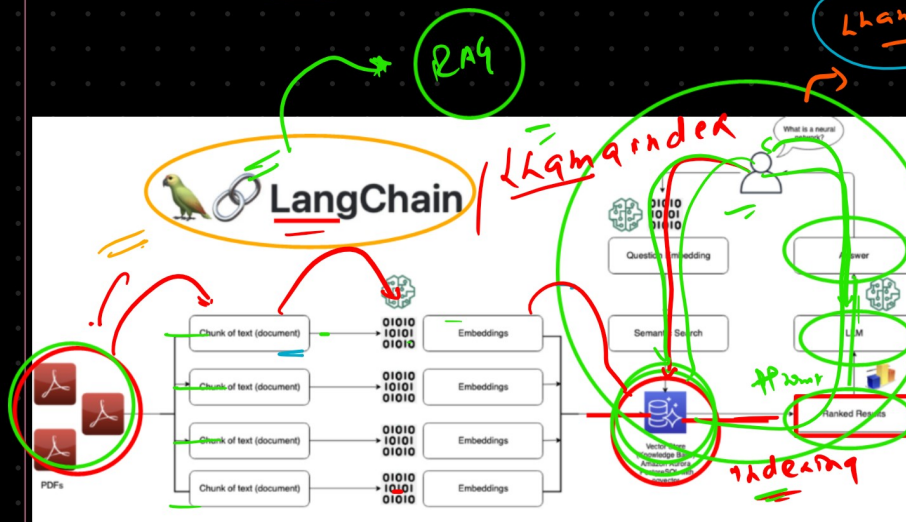Vector embedding are numerical representations of the data and for storing this vector we use a vector database.

RAG(Retrieval Augment Generation):

Retrieval augmented generation Is an architecture used to help LLMs model like GPT- 4 Gemini to provide a better response by using relevant information from additional sources and it reduces the chance that LLMs will lead to incorrect information.

LLM = RAG
"Vision without execution is hallucination"
- Thomas Edison

⇒ misleading

Lhama index | Langchain

RAG

LangChain / Lhamainder

(current of = LLM
RAG

indexing

Storing   Searching & retrival ⇒ Indexing

Llama Index:

1. Data Ingestion
2. Search
3. Retrieval
4. Indexing
5. Ranking and Structuring

Points to be remember about Llama Index:

1. Llama Index is specifically designed for building search and retrieval applications.

2. Llama Index provides a collection of features for integrating custom data into LLMs.(data ingestion)

3. Using you can connect your unstructured, structured or semi-structured data sources with llms.

4. Llama Index, can be an ideal solution if you're looking to work with vector embeddings. Using many available plugins you can load data from many sources easily, and generate vector embeddings using an embedding model.

5. It simplifies the process of querying LLMs and retrieving relevant documents based on the user's input.

Lang Chain:

Lang chain is an open-source framework designed to simplify the development of applications powered by large language models (LLMs). It goes beyond basic search and retrieval like Llama Index and provides a comprehensive toolkit for building more complex and interactive LLM applications.

This Lang Chain framework provides you lots of tools:

1. Prompt Template
2. LLM interface
3. Agents
4. Memory
5. Chains
6. Callback
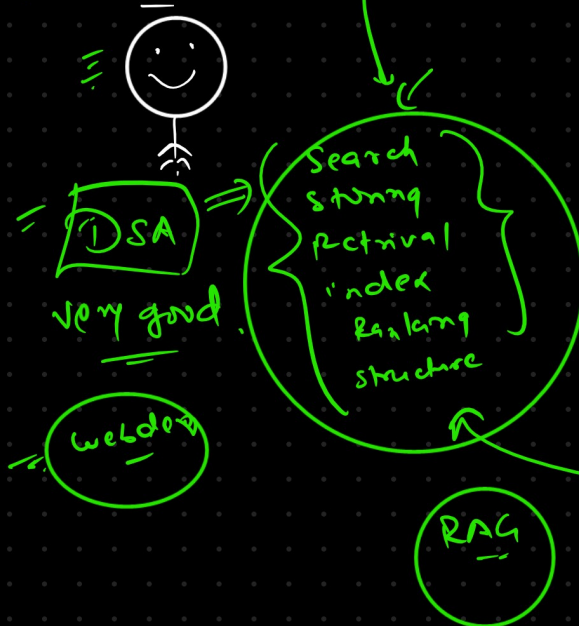7. Retrieval module
8. Lang serve
9. Lang smith
10. Lang graph

Retrival

3rd App

Chat

important (multiple)

Serving

Component

chains

DAG

Llamaindex

DSA

very good.

webdev

Search
storing
Retrival
index
Ranking
structure

RAG

Langchain

Little bit

DSA + webdev

Retrival
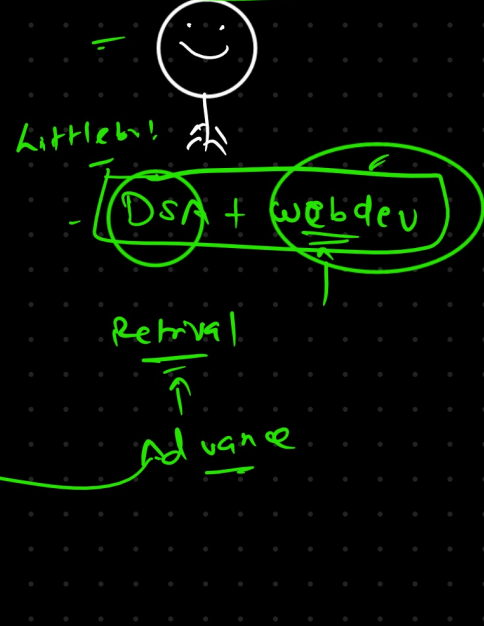*Advance*

In Conclusion I Can Say That:

1. llama index is optimized for indexing and retrieving.

2. You will be fine with just Lang chain(it is having everything but again i am saying for fast retrieval and RAG you will use llama index)

$= \{ llama\_index \} \rightarrow$ wrapped

3. You will be just fine with llama index(won't too much, searching and retrieval based on the index it is using wrapper of other framework on top of it)

4. You can combine both lama index and Lang chain for enhancing your application capability

Which one you need to use When:

Llama Index: Search and retrieval and if you want to get better search experience by building RAG llama index can be a ideal choice

Lang Chain: If you want to build end to end application with most comprehensive functionality then use Lang chain

Point to be noted: Here both tools you can use together to enhance your RAG application