**TASK-1**

**1. What is the motivation behind Retrieval-Augmented Generation (RAG)?**  
RAG helps AI give better answers by letting it search for real information instead of just guessing from memory.

**2. Explain the difference between RAG and standard LLM-based QA.**  
Regular AI answers from what it already knows. RAG first looks for helpful info (like reading a book) and then answers.

**3. What is the role of a vector store in a RAG pipeline?**  
A vector store helps the AI find the most related or similar documents to the question, like a smart search engine.

**4. Compare “stuff”, “map\_reduce”, and “refine” in LangChain.**

* **Stuff**: Put all the text together and give it to the AI at once.
* **Map\_reduce**: AI reads one by one and then combines answers.
* **Refine**: AI starts with one answer and keeps improving it with more info.

**5. What are the main parts of a simple LangChain RAG pipeline?**

* A tool to **find** the right info,
* An AI model to **answer**,
* A **prompt** to guide the AI,
* And a **chain** to connect all steps.

**TASK-2**

**RAG Pipeline Flow**

1. **User Query**  
   → The user asks a question (e.g., *"What is quantum computing?"*)
2. **Retriever**  
   → The retriever takes the user’s question and searches for the most relevant information.
3. **Vector Store**  
   → The retriever uses this to find similar documents based on meaning, not just keywords.  
   (These documents were stored earlier as vectors—numbers that represent their meaning.)
4. **LLM (Language Model)**  
   → The AI reads the documents found and uses them to generate an accurate, helpful answer.
5. **Final Answer Generation**  
   → The final answer is created and shown to the user.

**Visual Diagram (Text Style)**

less

CopyEdit

[User Query]

|

v

[Retriever] ---> [Vector Store] (Search for related info)

|

v

[Relevant Documents]

|

v

[LLM (Language Model)]

|

v

[Final Answer]