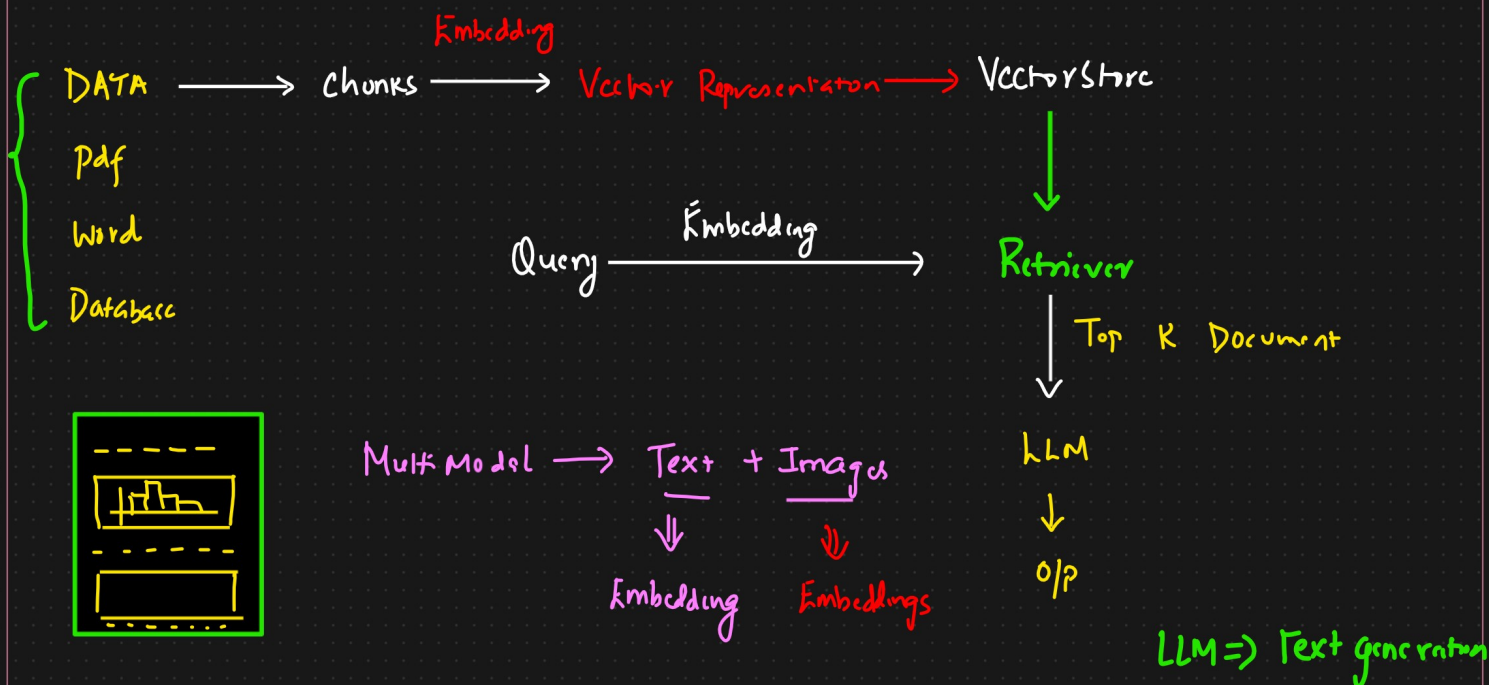


MultiModal RAG

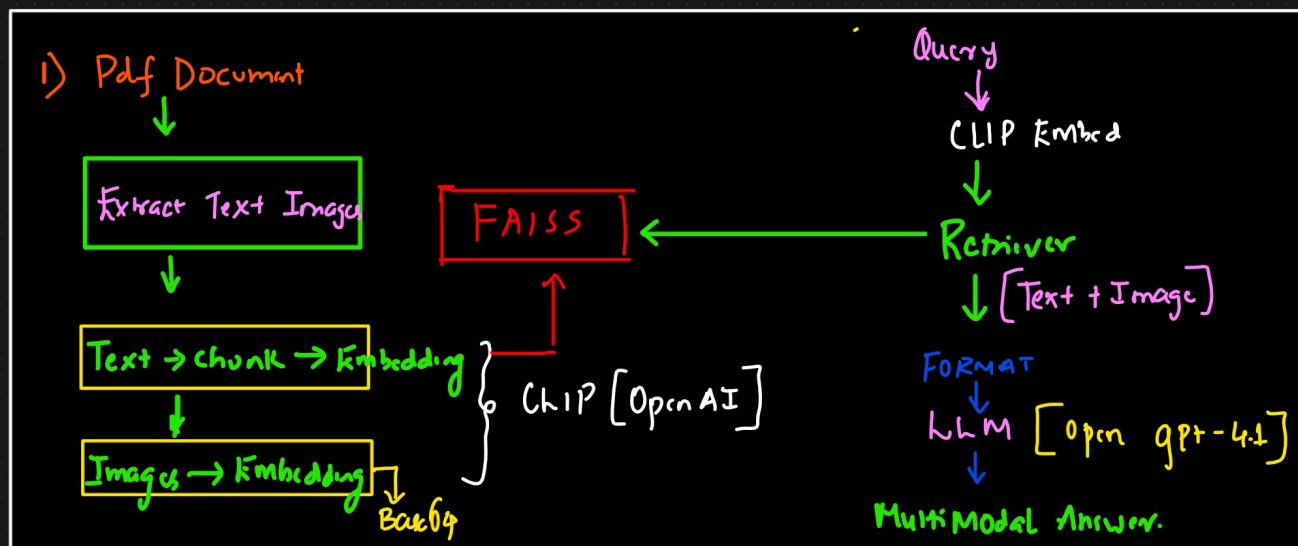


MultiModal LLM ⇒ Text + Images

⇒ OpenAI : gpt-4.1

Google Gemini flash 2.5

Steps : CLIP : Contrastive Language - Image Pre Training.



CLIP : Vision Transformer + Transformer