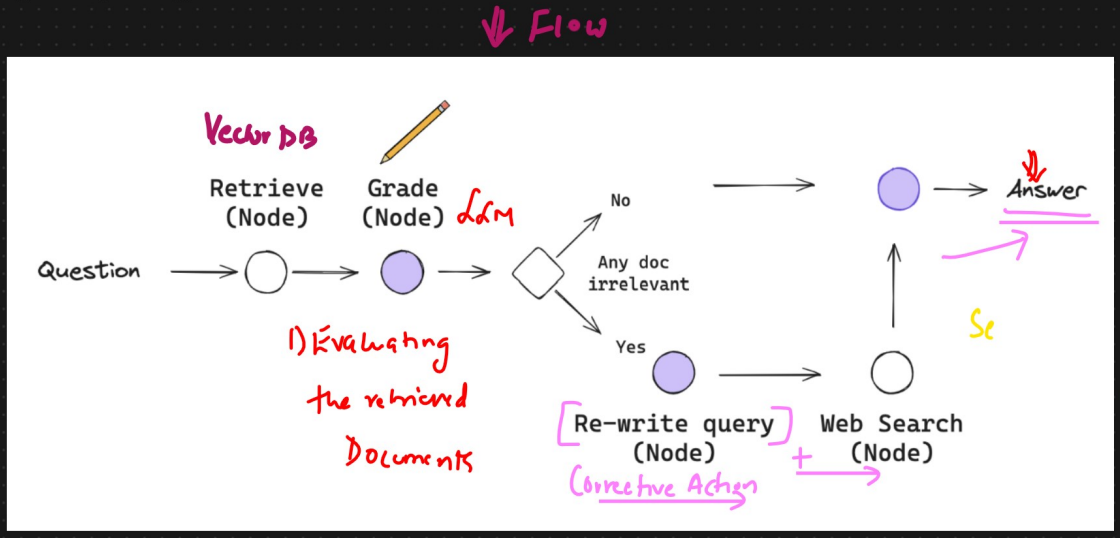
Corrective RAG (CRAG)

Corrective RAG (CRAG) is an advanced technique within Retrieval-Augmented Generation (RAG) that focuses on improving the accuracy and relevance of generated responses by incorporating mechanisms for self-reflection and self-grading of retrieved documents. It does this by evaluating the quality of retrieved documents and applying corrective actions when necessary, such as refining or replacing incorrect retrievals.



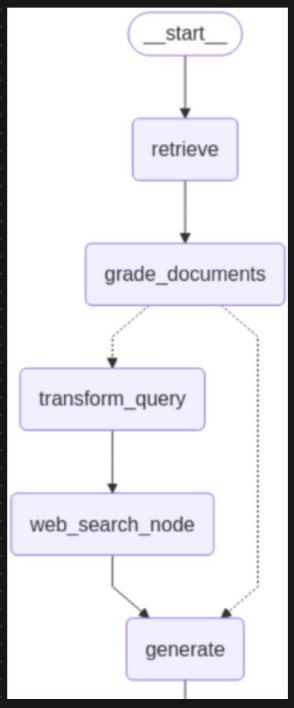
Here's a more detailed explanation:

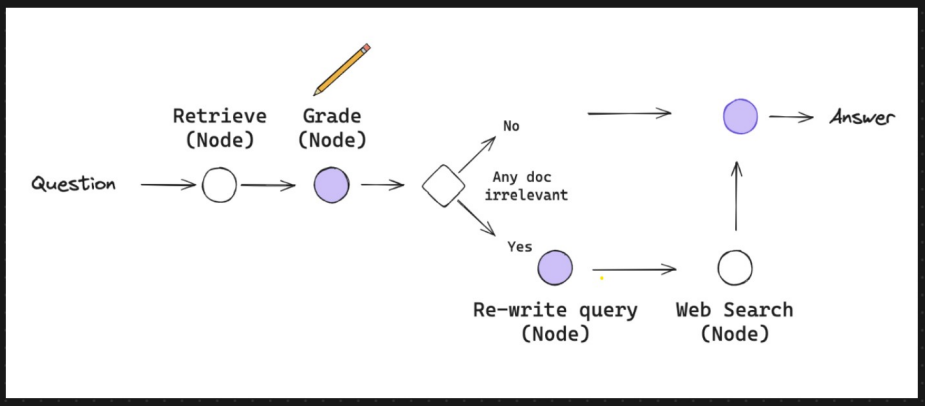
Addressing Limitations of Basic RAG:
Traditional RAG systems rely heavily on the accuracy of retrieved documents. If the retrieved information is flawed or incomplete, the generated response can also be inaccurate.

CRAG's Core Components:
1. Retrieval Evaluator: This component assesses the quality and relevance of retrieved documents.
2. Generative Model: This model generates the initial response based on the retrieved information.
3. Refinement and Correction: CRAG employs strategies like knowledge refinement or web search to address issues identified by the retrieval evaluator.

Benefits of CRAG:

1. **Improved Accuracy**: By evaluating and correcting retrieved information, CRAG helps ensure the accuracy of generated responses.
2. **Enhanced Relevance**: CRAG can identify and filter out irrelevant information, making the generated response more relevant.
3. **Increased Robustness**: CRAG can handle cases where the initial retrieval process is not perfect, leading to a more robust RAG system.


__start__
retrieve
grade_documents
transform_query
web_search_node
generate


Retrieve
(Node)
Grade
(Node)
Question
No
Any doc
irrelevant
Yes
Answer
Re-write query
(Node)
Web Search
(Node)