

MCP + RAG Execution Flow

Purpose

Explains how MCP workflows interact with RAG during runtime execution.

Execution Steps

- User query received
- Query embedding generated
- Top-K chunks retrieved
- Context injected into MCP prompt
- LLM response returned

What MCP Does Not Do

- No document ingestion
- No vector mutation
- No direct file access

End of Document – Commit under DOCS/