RAG Solution Overview

Pre-processing

Internal documents → **Chunk** → **Convert to embeddings** → **Index chunks, embeddings** → **Vector Store**

Question from chat ①

Embed Question / Semantic Search ②

Retrieve contextual data ③

Prompt + Context ④

Chatbot

Response

Orchestrator

Response ⑤

LLM

Domino

© 2024 Domino Data Lab, Inc.