

DocRAG API Report

Project Information

Project: DocRAG

Description: Multimodal Self-Hosted Retrieval-Augmented Generation System

Version: 0.1.0

System Architecture Services

Name	Technology	Language	Port
frontend	Next.js 15+	TypeScript	3000
backend	FastAPI	Python 3.14	8000
qdrant	Qdrant v1.17.0		6333
postgres	PostgreSQL 16		5432

API Endpoints

ID	Method	Path	Tag
health_check	GET	/health	Health
ingest_upload	POST	/api/v1/ingest/upload	Ingestion
query_search	GET	/api/v1/query/search	Retrieval
chat_ask	POST	/api/v1/chat/ask	Chat
chat_ask_stream	GET	/api/v1/chat/ask-stream	Chat
chat_create_session	POST	/api/v1/chat/sessions	Chat
chat_list_sessions	GET	/api/v1/chat/sessions	Chat
chat_rename_session	PATCH	/api/v1/chat/sessions/{session_id}	Chat
chat_delete_session	DELETE	/api/v1/chat/sessions/{session_id}	Chat
chat_get_history	GET	/api/v1/chat/history/{session_id}	Chat
documents_list	GET	/api/v1/documents/	Documents
documents_delete	DELETE	/api/v1/documents/{document_id}	Documents

Data Models

ChatSession

id: UUID

title: string

created_at: datetime

updated_at: datetime

ChatMessage

id: UUID
session_id: UUID
role: string
content: string
created_at: datetime

VectorChunk

id: UUID
content: string
metadata.document_id: string
metadata.file_name: string
metadata.chunk_index: integer
metadata.char_count: integer
metadata.token_count: integer