

## ● Problema escolhido

O problema escolhido foi criar um sistema automatizado de indexação e busca semântica de documentos armazenados no Google Drive, permitindo que um agente de IA responda perguntas utilizando RAG (Retrieval-Augmented Generation).

O objetivo principal é que o usuário faça upload de documentos no Google Drive e, automaticamente, esses arquivos sejam:

1. Extraídos
2. Convertidos para texto
3. Transformados em embeddings
4. Armazenados em um banco vetorial
5. Utilizados para responder perguntas de maneira contextualizada

Esse processo deve ocorrer sem intervenção manual, utilizando apenas triggers e automações.

- **Tecnologias utilizadas**

No projeto foram utilizadas:

### **1. Supabase**

- Banco vetorial utilizado para armazenar embeddings
- Tabelas para gerenciar documentos e metadados

### **2. OpenAI**

- Modelos de embeddings
- Modelo de Chat (GPT) integrado ao AI Agent

### **3. Google Drive**

- Repositório dos arquivos
- Fonte dos triggers de criação/atualização de arquivos

### **4. Automações da plataforma (Workflow)**

- Google Drive Trigger
- Data Loader
- Supabase Vector Store
- AI Agent
- Schedule Trigger (para limpeza periódica ou sincronização)

Esses componentes permitem criar um pipeline completamente automatizado.

## • O que são embeddings

Embeddings são representações matemáticas de texto geradas por modelos de IA, onde cada documento, frase ou parágrafo é convertido em um vetor de alta dimensionalidade.

Esses vetores permitem que a IA:

- compare similaridade entre textos
- faça busca contextual (semântica)
- entenda relações entre conceitos, não apenas palavras exatas

Exemplo simples:

- “carro” e “automóvel” → vetores próximos
- “carro” e “banana” → vetores distantes

Isso torna possível realizar buscas inteligentes, que entendem o significado, e não apenas palavras-chave.

## • Banco vetorial escolhido e justificativa

O banco vetorial escolhido foi o **Supabase Vector Store**.

### **Justificativas:**

Integração nativa com embeddings do OpenAI

Latência baixa e buscas rápidas

Escalabilidade para grandes volumes de documentos

Facilidade de configuração em workflows automatizados

Indexação robusta com índices vetoriais especializados (IVFFlat, HNSW)

Painel simples para monitoramento

Além disso, o Supabase funciona como uma solução completa (banco SQL + vetorial), permitindo salvar metadados, conteúdos originais e embeddings no mesmo ecossistema.

- **Arquitetura da solução**

A arquitetura observada no anexo funciona assim:

---

## **1. Ingestão de documentos**

- O Google Drive Trigger detecta quando um arquivo é criado ou atualizado.
  - Em seguida, o Download File baixa o conteúdo do Drive.
- 

## **2. Processamento e embeddings**

- O arquivo é enviado para o Default Data Loader, que converte o conteúdo para texto.
  - O texto é enviado para o OpenAI Embeddings, que gera o vetor numérico.
- 

## **3. Armazenamento**

- O vetor e os metadados do documento são salvos no Supabase Vector Store.

Existem dois pipelines:

- Um para arquivos novos
  - Outro para atualizações (com deleção e reindexação)
- 

## **4. Busca semântica**

Quando o usuário manda uma pergunta no chat:

- O AI Agent envia a consulta para o Supabase Vector Store
  - O banco retorna os embeddings mais similares
  - A IA monta uma resposta baseada nos documentos (RAG)
- 

## 5. Rotina de limpeza

- Um *Schedule Trigger* executa buscas no Drive
  - Verifica documentos removidos
  - Remove entradas antigas no Supabase
- 

Essa arquitetura garante:

- automação total
- consistência entre Drive ↔ Banco Vetorial
- alta precisão nas respostas

## • Prints do workflow funcionando

**Antes da deleção:**

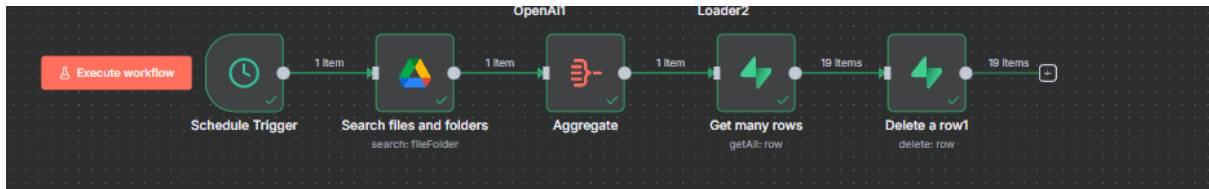
RAG Test / main PRODUCTION Connect

documents +

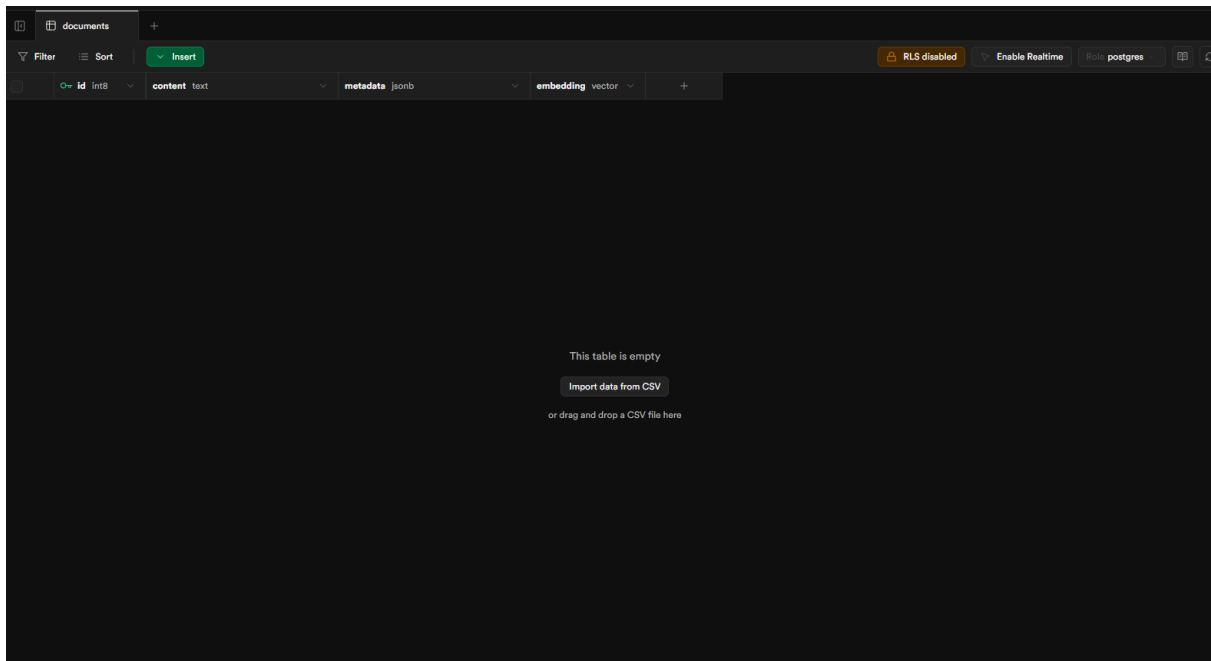
Filter Sort Insert

	id int8	content text	metadata jsonb	embedding vector	+
	282	ECONOMIA E FINANÇAS A economia	{"loc":{"lines":{"to":13,"from":1}}, "pdf":{"ini	[-0.016087785,-0.00504	
	283	empresas e mercados específicos; e maci	{"loc":{"lines":{"to":30,"from":12}}, "pdf":{"i	[-0.0053272652,0.00112	
	284	agentes interagem em mercados especifi	{"loc":{"lines":{"to":36,"from":25}}, "pdf":{"	[-0.059405487,0.015451	
	285	A teoria do consumidor explica como as p	{"loc":{"lines":{"to":67,"from":39}}, "pdf":{"	[-0.033610113,-0.00464	
	286	★ FINANÇAS Finanças é a área que admi	{"loc":{"lines":{"to":75,"from":64}}, "pdf":{"	[-0.0019982774,0.02119	
	287	● Fundos de emergência ● Investimento:	{"loc":{"lines":{"to":117,"from":78}}, "pdf":{"	[-0.003758441,0.024124	
	288	★ MERCADOS FINANCEIROS Os merca	{"loc":{"lines":{"to":153,"from":120}}, "pdf":	[-0.0023141657,0.015761	
	289	● Saber aproveitar juros compostos ● Inv	{"loc":{"lines":{"to":190,"from":156}}, "pdf":	[0.006072117,0.0411447	
	290	● Taxas de câmbio ● Inflação ● Decisõe	{"loc":{"lines":{"to":228,"from":193}}, "pdf":	[0.0023382048,0.00356	
	291	A globalização conecta economias, aume	{"loc":{"lines":{"to":263,"from":230}}, "pdf"	[-0.0025529547,-0.0022	
	292	● Pagar contas ● Planejar o futuro ● Inv	{"loc":{"lines":{"to":290,"from":266}}, "pdf"	[0.009299336,0.025904	
	293	objetivos pessoais. 1. Conceito de Investir	{"loc":{"lines":{"to":303,"from":288}}, "pdf"	[-0.014676849,0.023657	
	294	2. Importância dos Investimentos Investir	{"loc":{"lines":{"to":338,"from":306}}, "pdf"	[-0.01777,0.050834183,0.	
	295	A renda variável apresenta maior volatilid	{"loc":{"lines":{"to":369,"from":340}}, "pdf"	[-0.0005797089,0.05453	
	296	trade). Requer técnica, disciplina e análise	{"loc":{"lines":{"to":370,"from":366}}, "pdf"	[0.038593717,0.05134531	
	297	5. Análise de Investimentos Para avaliar ir	{"loc":{"lines":{"to":403,"from":373}}, "pdf"	[-0.01288191,0.06784762	
	298	● MACD ● Suportes e resistências Enqui	{"loc":{"lines":{"to":438,"from":406}}, "pdf"	[0.026078865,0.028603	
	299	● C = capital inicial ● i = taxa de juros ●	{"loc":{"lines":{"to":469,"from":441}}, "pdf"	[0.015722672,0.0307510	
	300	● Entendimento dos ciclos econômicos N	{"loc":{"lines":{"to":484,"from":472}}, "pdf"	[0.044842996,0.0313447	

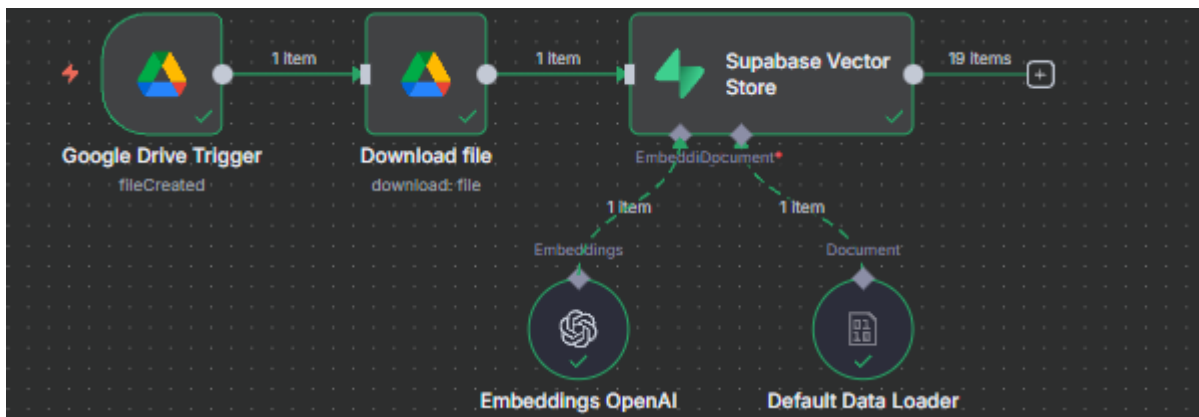
## Workflow de deleção:



## Pós execução:



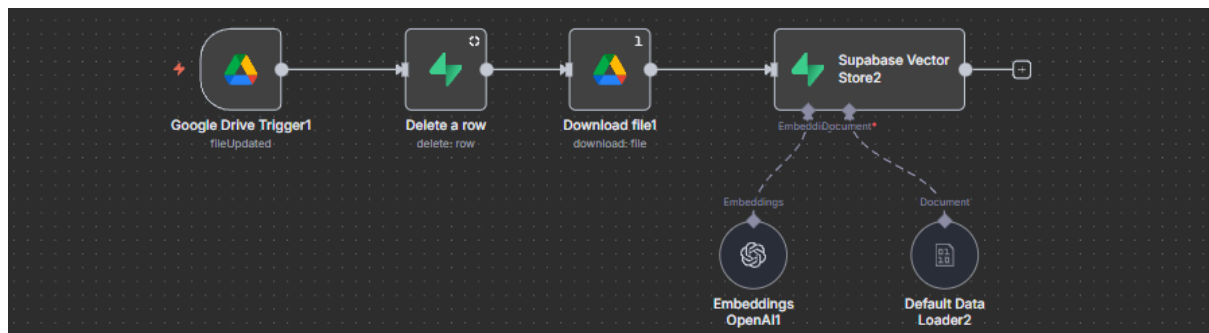
Adicionar dados para o supabase:



Pós execução:

id	content	metadata	embedding
301	ECONOMIA E FINANÇAS A economia	{\"loc\": {\"lines\": {\"to\": 13, \"from\": 1}, \"pdf\": {\"in	[-0.016096275, -0.00602,
302	empresas e mercados específicos; e maci	{\"loc\": {\"lines\": {\"to\": 30, \"from\": 12}, \"pdf\": {\"in	[-0.005278362, 0.001153,
303	agentes interagem em mercados especifi	{\"loc\": {\"lines\": {\"to\": 36, \"from\": 26}, \"pdf\": {\"in	[-0.059404127, 0.015440,
304	A teoria do consumidor explica como as	{\"loc\": {\"lines\": {\"to\": 67, \"from\": 39}, \"pdf\": {\"in	[-0.033566816, -0.00472,
305	★ FINANÇAS Finanças é a área que adn	{\"loc\": {\"lines\": {\"to\": 75, \"from\": 64}, \"pdf\": {\"in	[-0.0020383, 0.0217740,
306	● Fundos de emergência ● Investimento	{\"loc\": {\"lines\": {\"to\": 117, \"from\": 78}, \"pdf\": {\"in	[-0.0037315078, 0.02407,
307	★ MERCADOS FINANCEIROS Os merca	{\"loc\": {\"lines\": {\"to\": 153, \"from\": 120}, \"pdf\": {\"in	[-0.0023007428, 0.01576,
308	● Saber aproveitar juros compostos ● In	{\"loc\": {\"lines\": {\"to\": 190, \"from\": 156}, \"pdf\": {\"in	[0.0060724923, 0.041190,
309	● Taxas de câmbio ● Inflação ● Decisão	{\"loc\": {\"lines\": {\"to\": 228, \"from\": 193}, \"pdf\": {\"in	[0.0022976901, 0.00356,
310	A globalização conecta economias, aume	{\"loc\": {\"lines\": {\"to\": 263, \"from\": 230}, \"pdf\": {\"in	[-0.002526167, -0.00226,
311	● Pagar contas ● Planejar o futuro ● Inve	{\"loc\": {\"lines\": {\"to\": 290, \"from\": 266}, \"pdf\": {\"in	[0.009272731, 0.0259031,
312	objetivos pessoais. 1. Conceito de Investir	{\"loc\": {\"lines\": {\"to\": 303, \"from\": 288}, \"pdf\": {\"in	[-0.014676849, 0.023657,
313	2. Importância dos Investimentos Investir	{\"loc\": {\"lines\": {\"to\": 338, \"from\": 306}, \"pdf\": {\"in	[-0.017793644, 0.050935,
314	A renda variável apresenta maior volatilid	{\"loc\": {\"lines\": {\"to\": 369, \"from\": 340}, \"pdf\": {\"in	[-0.0005582873, 0.0544,
315	trade). Requer técnica, disciplina e análise	{\"loc\": {\"lines\": {\"to\": 370, \"from\": 366}, \"pdf\": {\"in	[0.03859354, 0.0512967,
316	5. Análise de Investimentos Para avaliar ir	{\"loc\": {\"lines\": {\"to\": 403, \"from\": 373}, \"pdf\": {\"in	[-0.013003356, 0.067769,
317	● MACD ● Suportes e resistências Enqu	{\"loc\": {\"lines\": {\"to\": 438, \"from\": 406}, \"pdf\": {\"in	[0.0266626, 0.02722988,
318	● C = capital inicial ● i = taxa de juros ●	{\"loc\": {\"lines\": {\"to\": 469, \"from\": 441}, \"pdf\": {\"in	[0.015701205, 0.0306408,
319	● Entendimento dos ciclos econômicos N	{\"loc\": {\"lines\": {\"to\": 484, \"from\": 472}, \"pdf\": {\"in	[0.044842996, 0.031344,

## Workflow em caso do doc ser atualizado:



## • Explicação da busca semântica

A busca semântica funciona assim:

1. O usuário envia uma pergunta.
2. A pergunta também é convertida em embedding.
3. O Supabase compara esse embedding com todos os embeddings dos documentos, medindo distância vetorial (cosine similarity).
4. Os documentos mais “parecidos” com a pergunta são retornados.
5. Esses trechos relevantes são enviados ao GPT.
6. O agente monta a resposta utilizando a base de conhecimento encontrada.

Diferente da busca tradicional por palavras-chave, a busca semântica entende contexto, mesmo que as palavras não sejam exatamente as mesmas.

## • Resultados e conclusões

Com a arquitetura montada, os principais resultados foram:

Pipeline totalmente automatizado

Indexação confiável de documentos do Google Drive

Respostas do agente de IA muito mais precisas

Busca semântica funcionando com alta relevância

Supabase demonstrou ótimo desempenho como banco vetorial

**Conclusão geral:**



A solução implementada cria uma base de conhecimento dinâmica, sempre atualizada conforme o Drive muda. O uso de embeddings somado ao Supabase proporciona uma busca inteligente e eficiente, permitindo que um agente de IA responda perguntas de forma contextualizada e precisa, viabilizando um sistema de RAG profissional.