Al-Driven Data Access: Simplifying Database Interaction



André Roque Felipe Becker Guilherme Marques



Problem: Obtaining SQL queries

- Lack of SQL expertise among business users
- Complexity of large, multi-table database schemas
- Multiple database dialects across environments
- Difficulty debugging incorrect SQL queries



Solution:

RAG Agent with SQL Generation

- Natural Language Input
 User asks question about the database in natural language
- Data Retrieval
 Tables documentation is retrieved from a Vector Database
- SQL Generation
 LLM converts question to database question
- Query Retrieval
 System executes query and store Data
- Human Friendly Response
 LLM converts query results to natural language





Features of Our Product

Natural Language to SQL

Translates natural language questions into syntactically correct SQL queries.

Answer based database information

The agent assures that the answer provided is sustained by the data in the database.

Handles Complex Schemas

Automatically retrieves and contextualizes relevant elements from large, multitable database schemas using retrieval-augmented generation (RAG) technology.

Export query Results

Provides to the user the option of downloading the query results into a .csv file.

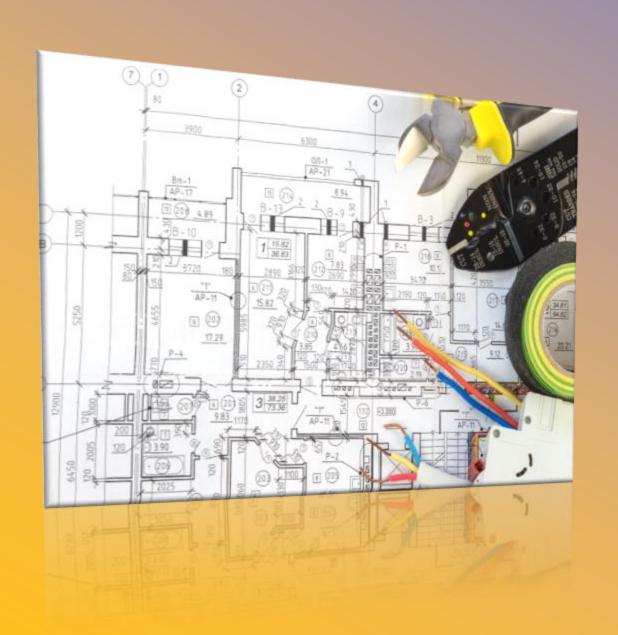
Generates dialect-specific SQL (e.g., PostgreSQL, SQLite, MySQL) automatically

Empowers business users to access data directly without needing constant technical support.

Visualise query option

The user is allowed to see the query produced to answer his question.





Technologies and Frameworks

LLM Models

- Llama 3.2:3b via Ollama
- ChatGPT 4.1

Data Processing

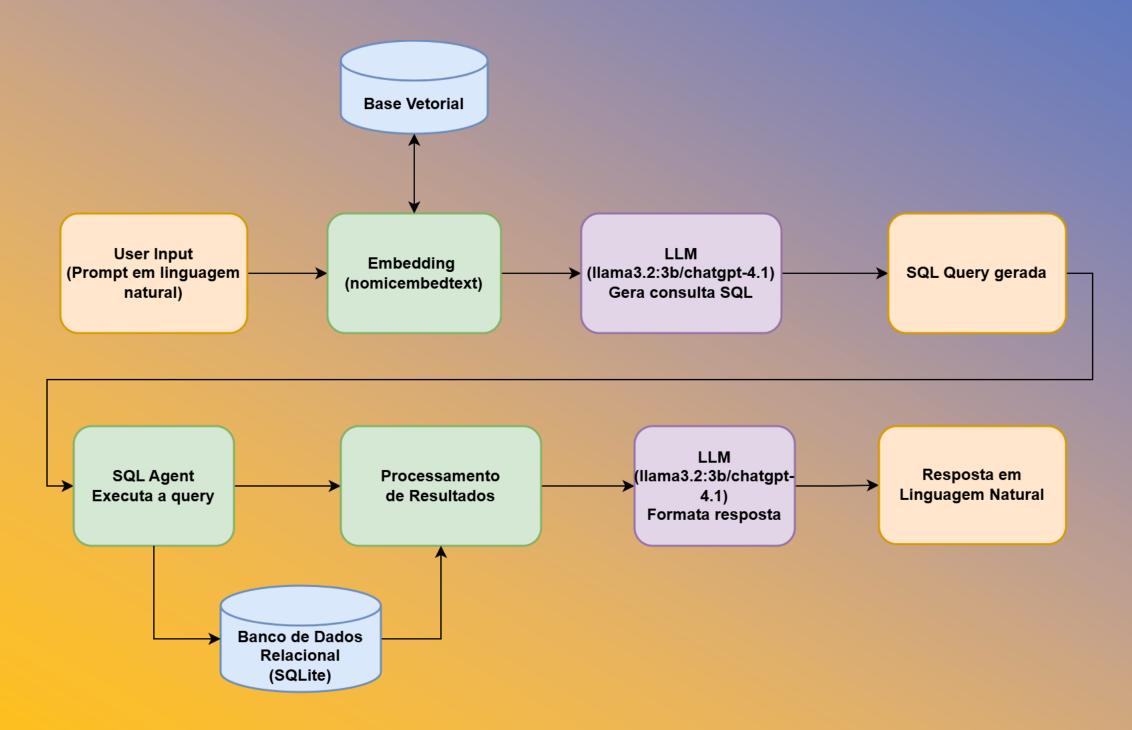
- ChromaDB Vector Database
- Nomic Embed Text via Ollama

User Interface

Streamlit Framework



Data Flow Process

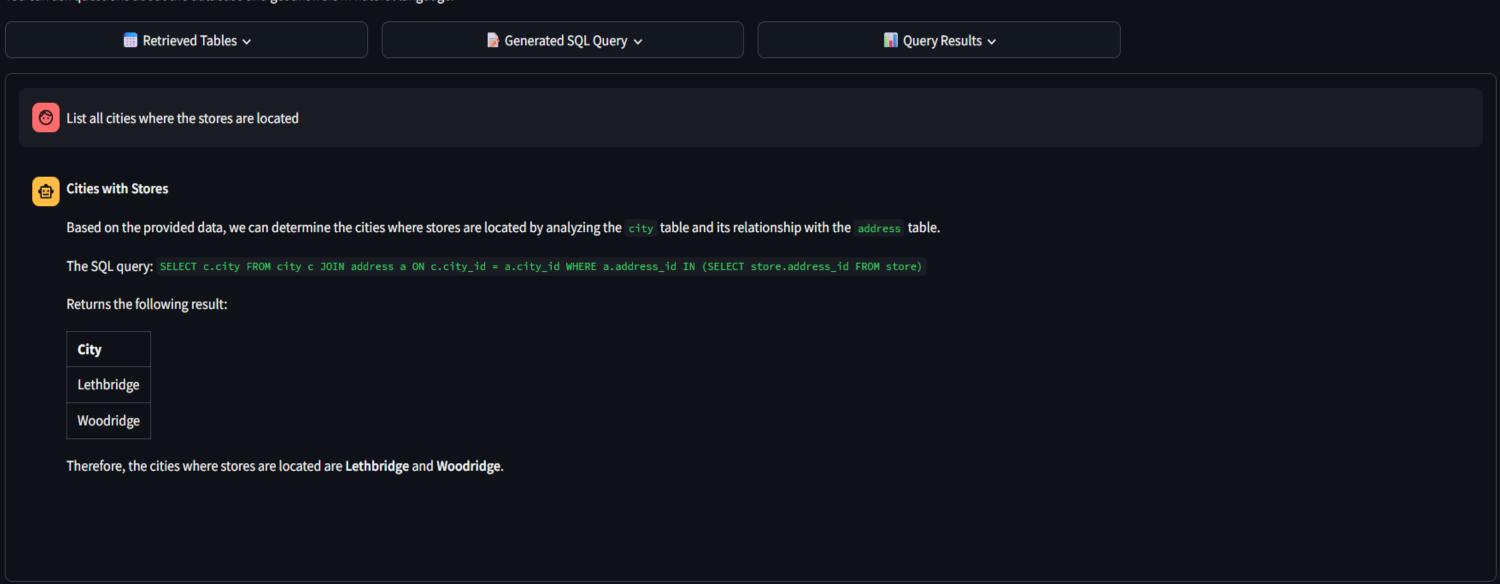






RAG SQL App Interface

You can ask questions about the database and get answers in natural language.



Ask a question about the database

Retrieved Tables ^

General

ID: 3, Distance: 0.5151309967041016,

Document: CREATE TABLE city (city_id INTEGER NOT NULL, city VARCHAR(50) NOT NULL, country_id SMALLINT NOT NULL, last_update TIMESTAMP NOT NULL, PRIMARY KEY (city_id), CONSTRAINT fk_city_country FOREIGN KEY(country_id) REFERENCES country (country_id) ON UPDATE CASCADE)

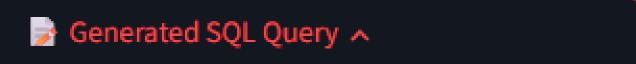
ID: 15, Distance: 0.5250715613365173,

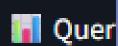
Document: CREATE TABLE store (store_id INTEGER NOT NULL, manager_staff_id SMALLINT NOT NULL, address_id INTEGER NOT NULL, last_update TIMESTAMP NOT NULL, PRIMARY KEY (store_id), CONSTRAINT fk_store_staff FOREIGN KEY(manager_staff_id) REFERENCES staff (staff_id), CONSTRAINT fk_store_address FOREIGN KEY(address_id) REFERENCES address (address_id))

ID: 1, Distance: 0.5265400409698486,

Document: CREATE TABLE address (address_id INTEGER NOT NULL, address VARCHAR(50) NOT NULL, address2 VARCHAR(50) DEFAULT NULL, district VARCHAR(20) NOT NULL, city_id INTEGER NOT NULL, postal_code VARCHAR(10) DEFAULT NULL, phone VARCHAR(20) NOT NULL, last_update TIMESTAMP NOT NULL, PRIMARY KEY (address_id), CONSTRAINT fk_address_city FOREIGN KEY(city_id) REFERENCES city (city_id) ON UPDATE CASCADE)







SELECT c.city FROM city c JOIN address a ON c.city_id = a.city_id WHERE a.address_id IN (SELECT store.address_id FROM store)



```
Query Results 
 Download CSV
Total Results: 2
Results:
   ▼0:{
      "city": "Lethbridge"
   ▼1:{
      "city": "Woodridge"
```

Future Roadmap



Scalable infrastructure deployment

- Fine tuning of hyperparameters
- Increase RAG base information
- Improving response times
- Cloud integration
- Asynchronous reports

Scheduled Background jobs

- Run syncs as scheduled tasks
- Log schema diffs before and after sync for auditability
- Resource & Rate Limits

Scalability



Enhanced Security



Automatic Sync



External Integrations

Database protection and compliance

- GDPR/LGPD compliance integration
- Input validation
- Schema controlled guardrails
- Audit logs

Connection with:

- External APIs
- Cloud databases
- BI Tools
- Data pipelines

About Capgemini

organizations to accelerate their dual transition to a digital and sustainable world, while creating tangible impact for enterprises and society. It is a responsible and delivers end-to-end services and solutions leveraging strengths from strategy and design to engineering, all fueled by its market leading capabilities in Al,









