

SQL with Gen AI

Tables, Transformed with AI

Agenda

Application Overview

3

Data Overview

4

- Tabular Database
- Document Database

4

6

Agents Overview

10

- SQL Agent
- Analyst Agent
- PDF Agent

10

13

16

Refactor Considerations

19

Application Overview

Objectives

Automate SQL query generation for efficient data retrieval, manipulation, and visualization

Core Concepts

- Leverage LLMs to translate natural language into database queries
- Retrieve relevant documents to provide context for query generation
- Implement an agentic communication pipeline for interactions

Tabular Database



Source Data: CSV files

Data Manipulation

The CSV is loaded into a structured database, with each row stored as a record. Row-level embeddings are generated and saved within the same database for similarity search.

Use cases


- The generated queries are intended for use on the tabular database
- Embeddings of rows help agents retrieve valid data points to use in queries

Chunking Strategy

- Rows are converted into sentences using column names to provide semantic clarity
- Embeddings are stored in the structured database

Access database			Create sentence per row	Create and store embedding per row
Name	Address	Rating	<ol style="list-style-type: none">1. The Name is Pasta Planet. The Address is 521 Napolean Rd, 11225. The Rating is 4.2. The Name is Griddle Spot. The Address is 184 Kindle Rd, 11223. The Rating is 5.3. The Name is Pizza House. The Address is 777 Nostrant, 11226. The Rating is 5	{row_1: [embeddings...]}
Pasta Planet	521 Nep.. 11225	4		
Griddle Spot	184 Kin... 11223	5		{row_2:[embeddings...]}
Pizza House	777 No... 11226	5		{row_3:[embeddings...]}

Document Database



Document
Database
(vectors)

Source materials: User manuals and textbooks

Data Manipulation

PDF is parsed into markdown text and then inserted into a Document Database

Use case

The document database contains supporting information that can provide added context that aids in query generation

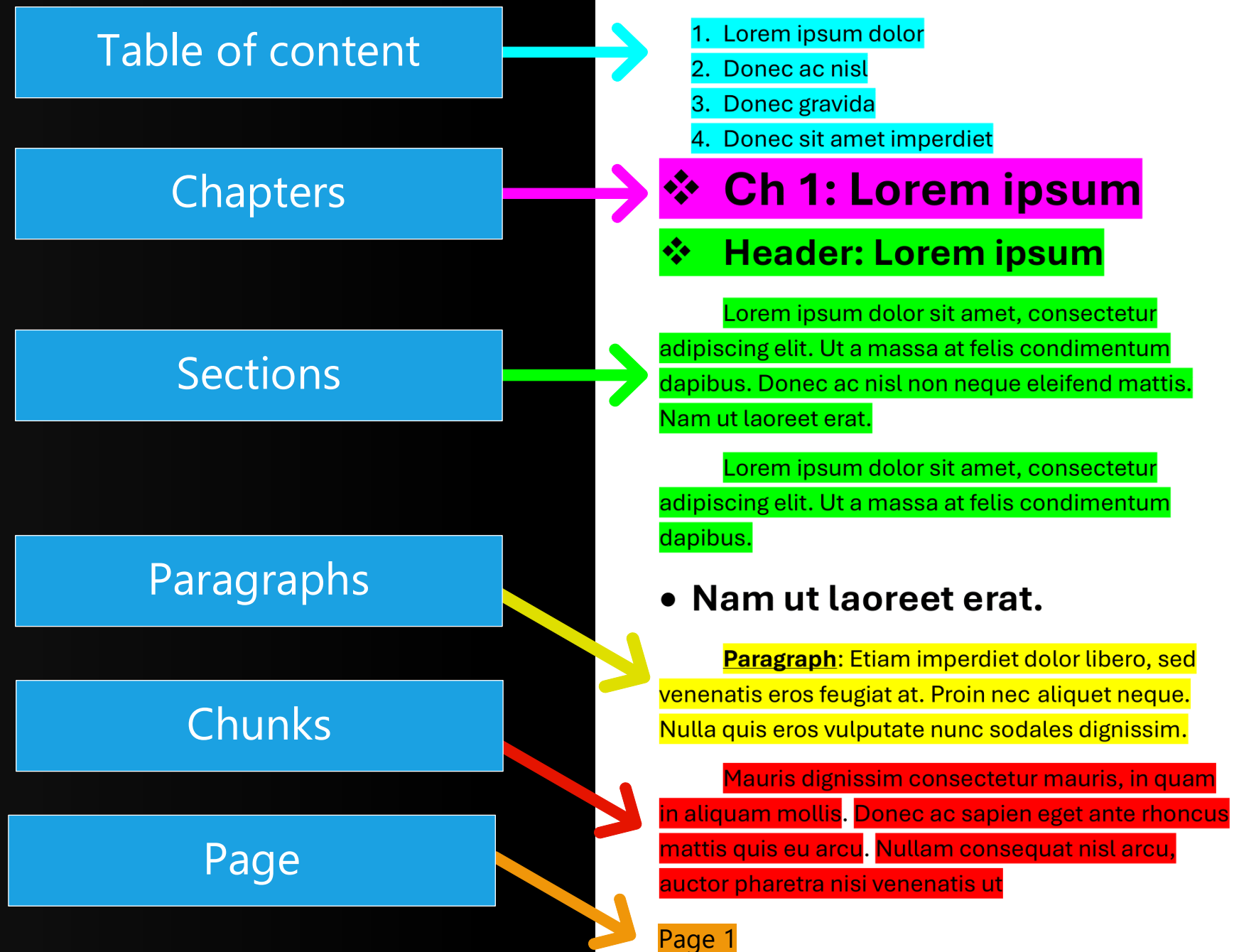
Chunking Strategy

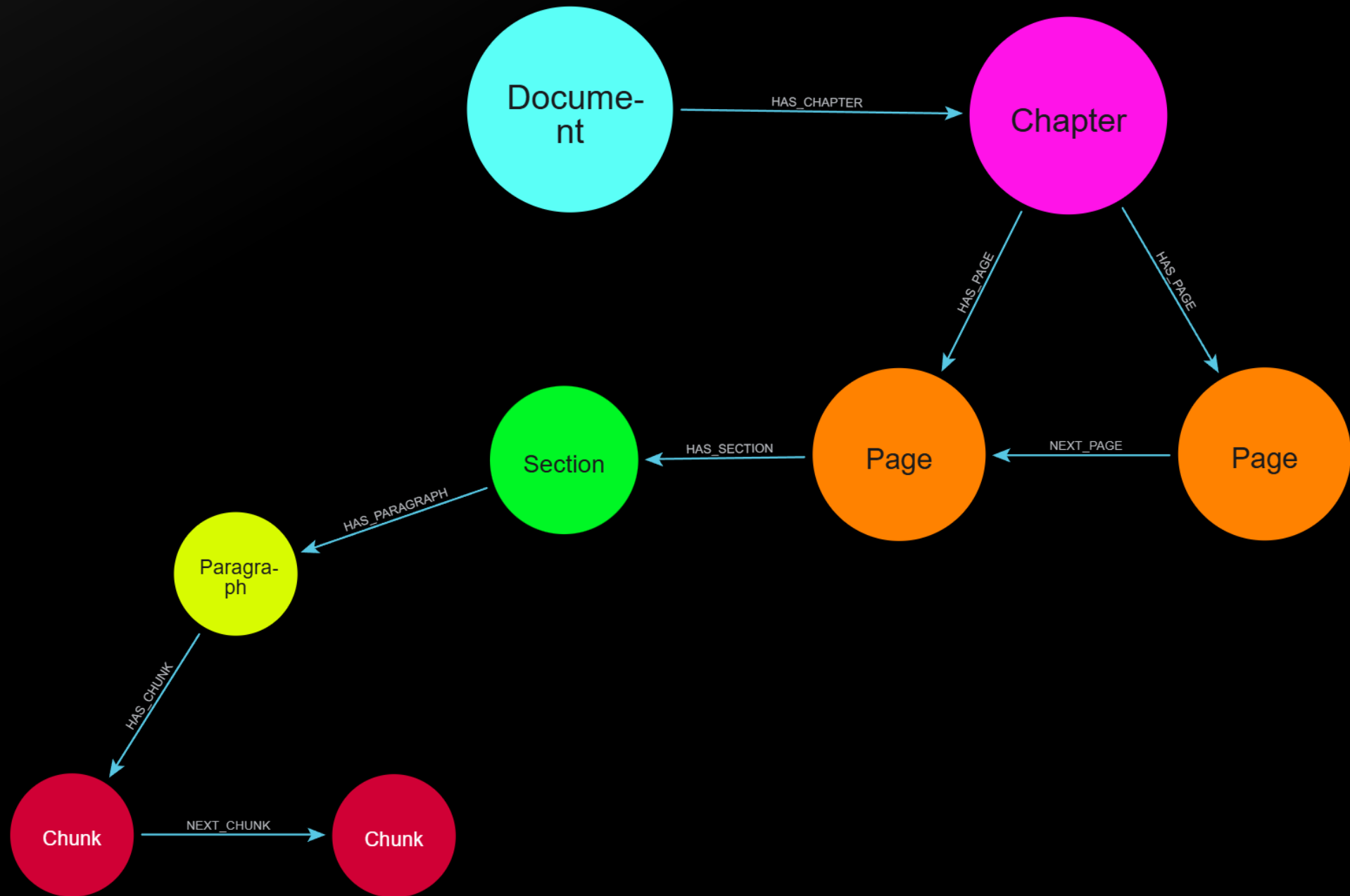
Problem with Traditional Vector Search

- Returns only the closest chunk
- No awareness of boundaries like Page, Section, or Chapter

Parent-Child Retriever Solution

Leverages graph relationships to define context-aware retrieval windows





SQL Agent

The SQL agent interprets the user's question and responds to it with SQL queries that results in an action.

- **Input:** User's question
- **Output:** Database Query
- **Database:** Reads from tabular database

Input	What Country has the highest happiness index in 2024?
Output	SELECT Country, GDP_2024 FROM "1990_2025_global_gdp" ORDER BY GDP_2024 DESC LIMIT 1;

Identify Query type

Input Question: "Which country had the highest GDP in 2024?"

- **Invoke OpenAI GPT model**
- **Prompt:** "Does the users question, relate to a retrieval or manipulation query?"

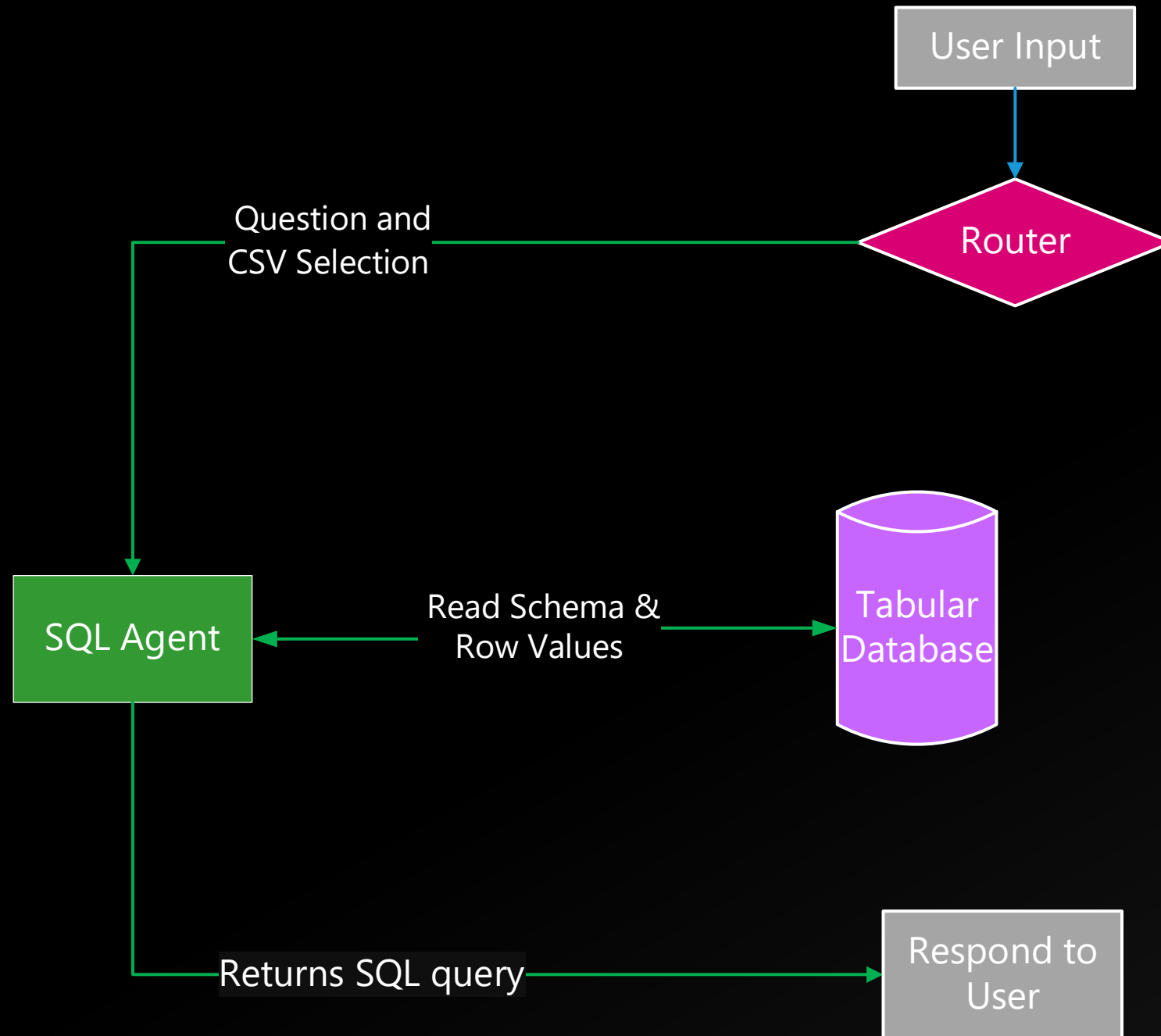
Output Query Type: retrieval

Create SQL Query

Input Question & Query Type

- **Invoke OpenAI GPT model**
- **Prompt:** "Write a SQL query that answers the users question. Test the query by observing the results for accuracy."
- **Use tool calling for tabular database schema access and query execution**

Output Generated Query:
``SELECT Country, GDP_2024 FROM
"1990_2025_global_gdp" ORDER BY
GDP_2024 DESC LIMIT 1;``

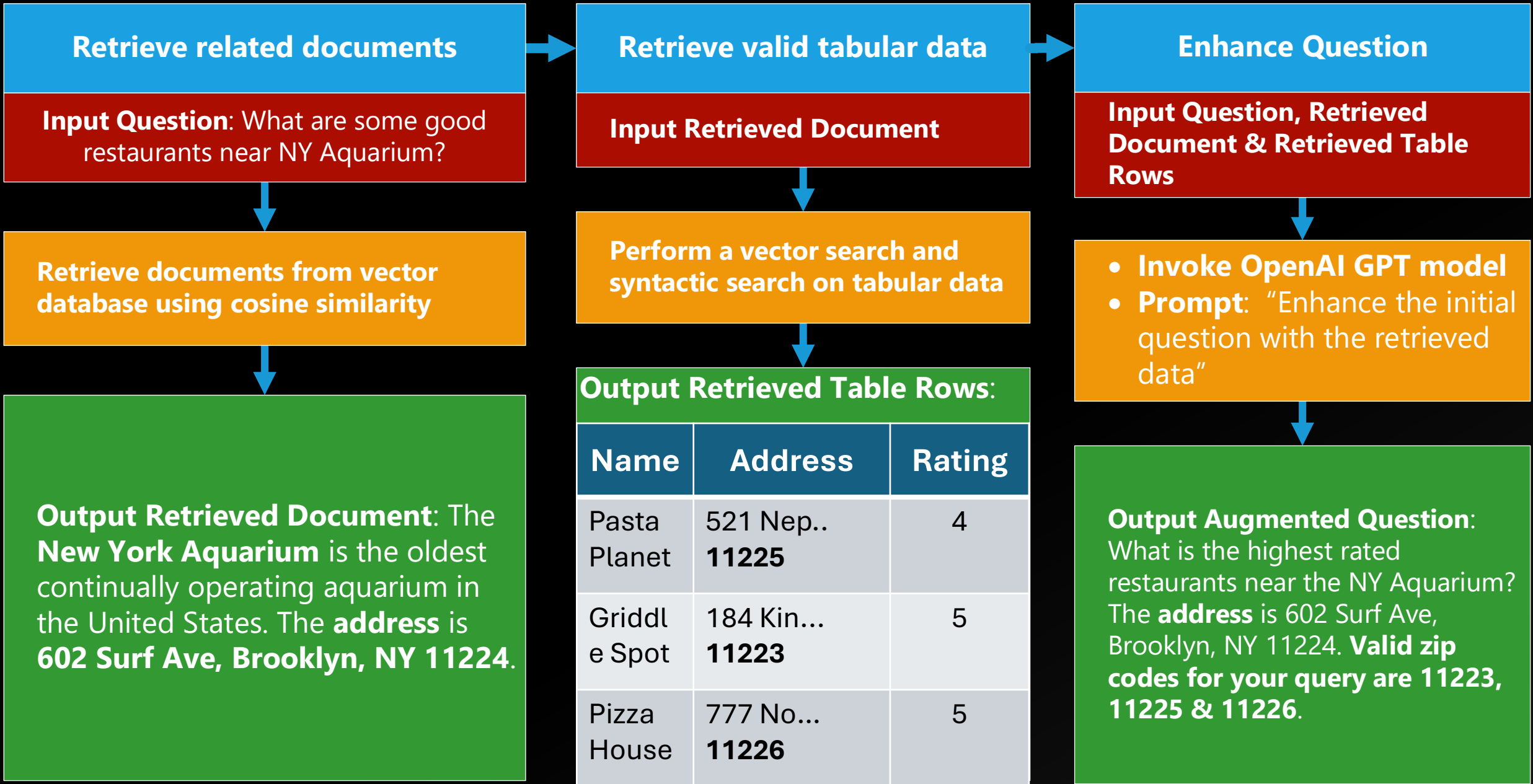


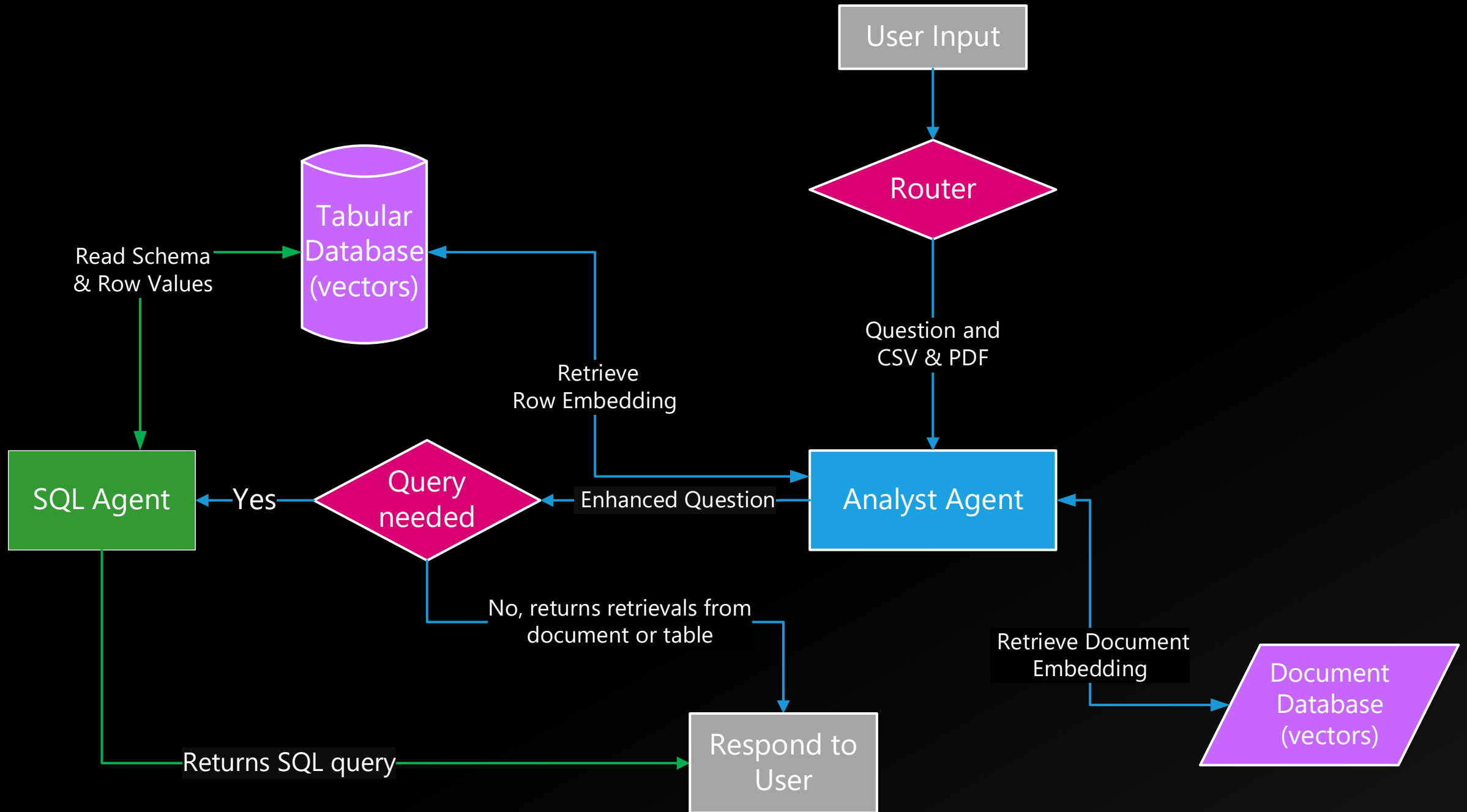
Analyst Agent

Enhances user questions with relevant context from retrieved documents and tabular database before query generation, enabling the SQL Agent to generate precise queries.

- **Input:** User's question
- **Output:** Enhanced version of user's question
- **Database:** Reads from tabular database and document database

Input	What are some good restaurants near NY Aquarium?
Output	What is the highest rated restaurants near the NY Aquarium? The address is 602 Surf Ave, Brooklyn, NY 11224. Valid zip codes for your query are 11223, 11225 & 11226.



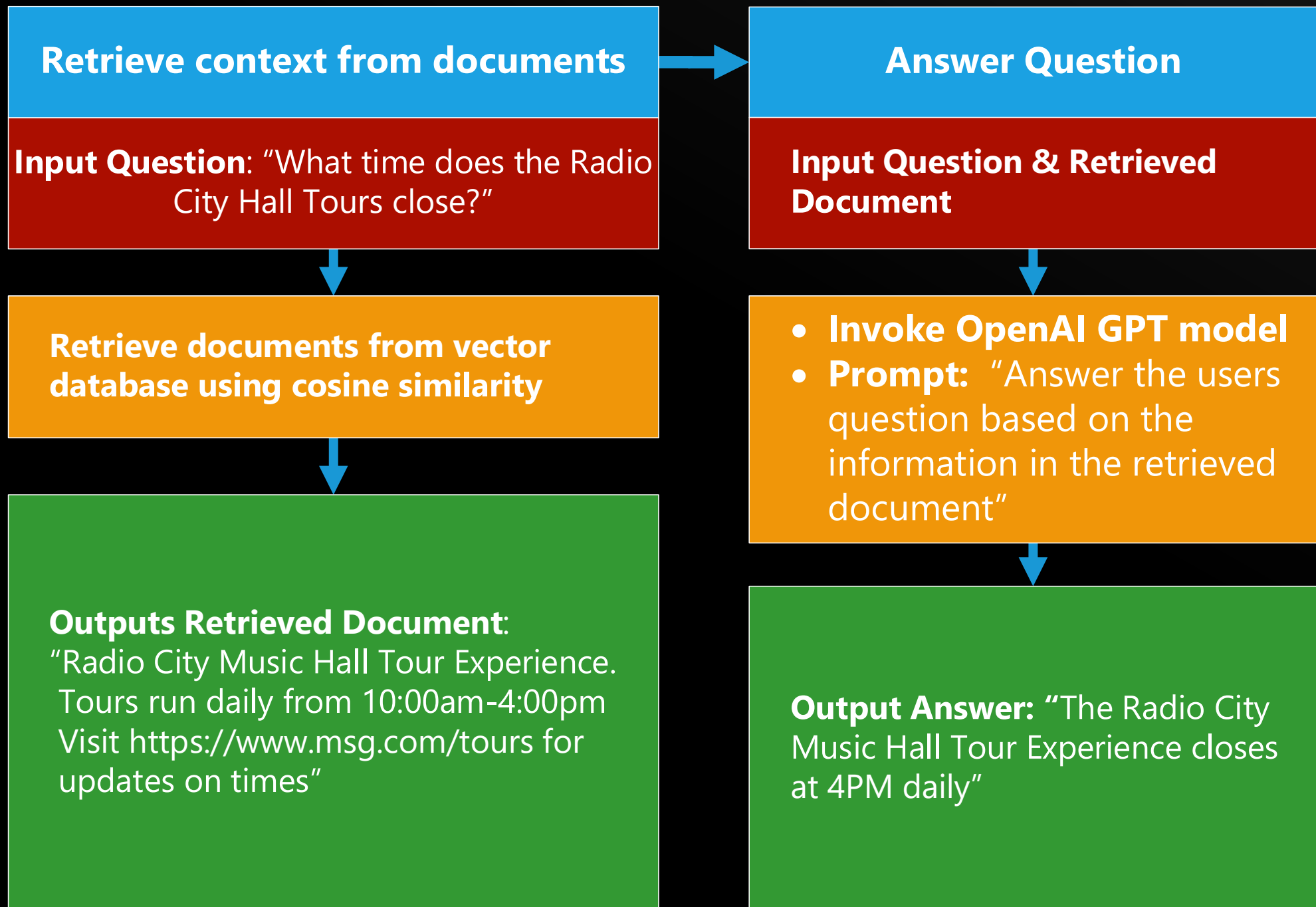


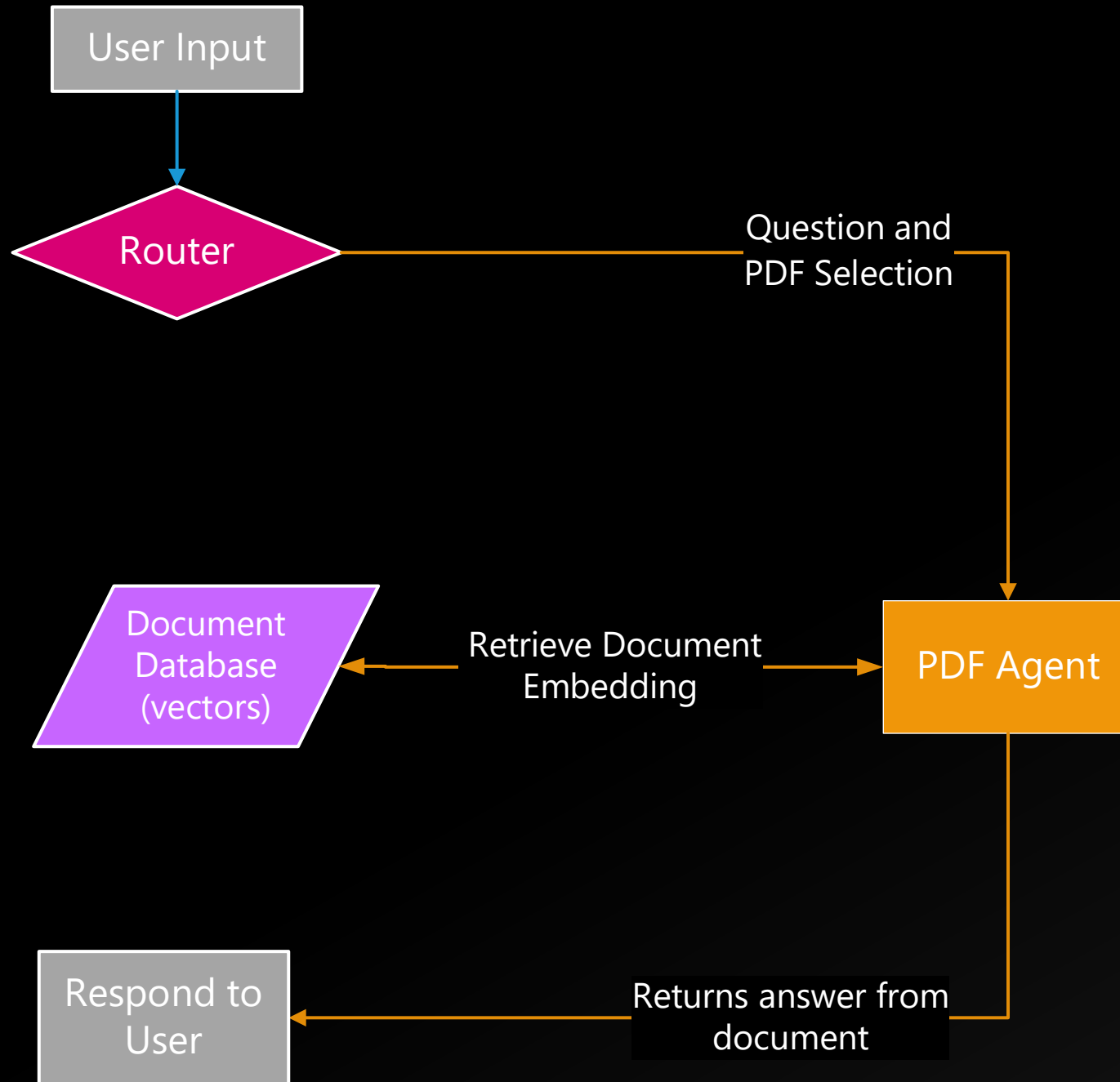
PDF Agent

Answers user questions directly using retrieved content from documents.

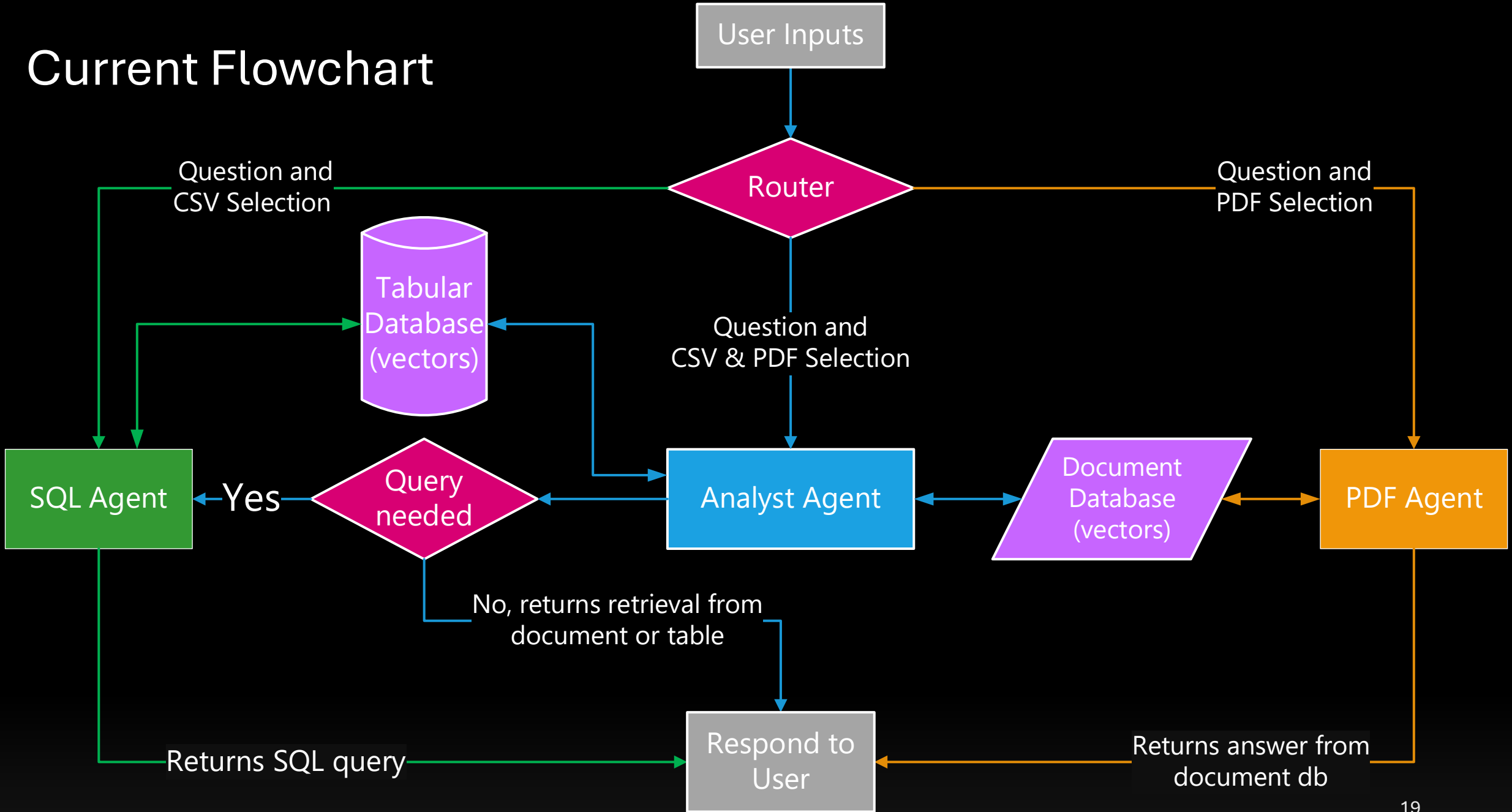
- **Input:** User's question
- **Output:** Answer and retrieved documents
- **Database:** Reads from document vector database

Input	What time does the Radio City Music Hall close?
Output	The Radio City Music Hall Tour Experience closes at 4PM daily

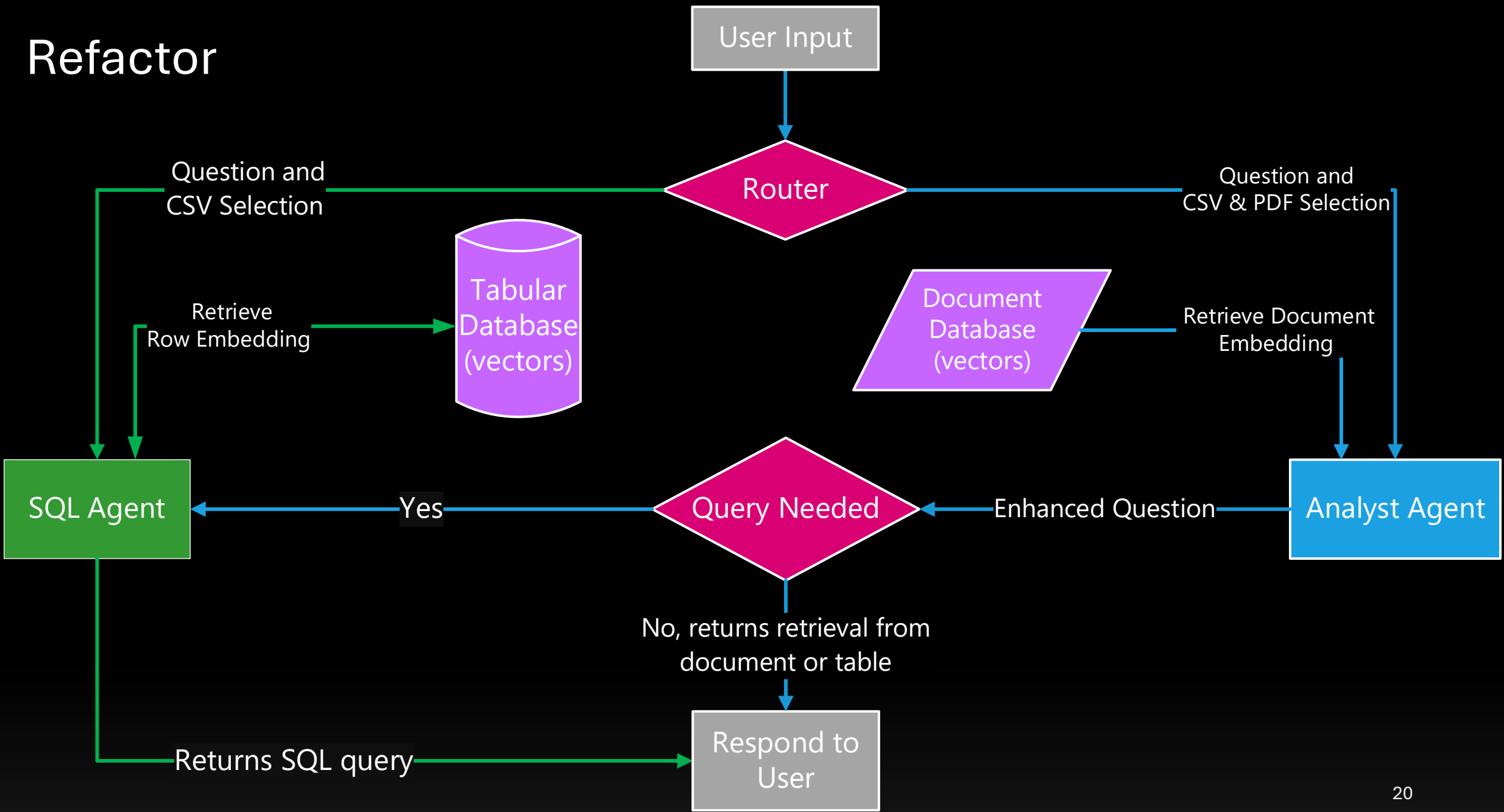




Current Flowchart



Refactor



Issue, redundant data flows

Issue: Currently, both the Analyst Agent and the SQL Agent access the tabular database, creating redundant data flows. Restricting unstructured document handling to the Analyst Agent and tabular data to the SQL Agent will improve maintainability and assign each agent to its respective data domain.

Considered changes: Reassign responsibility for tabular data validation/retrieval from the Analyst Agent to the SQL Agent. This will result in the below structure.

- Analyst Agent → Document retrieval + Question enrichment
- SQL Agent → Tabular data validation + Query generation

Issue, is PDF agent necessary?

Issue: The current role of the PDF Agent is limited to document-based question answering using retrieved context. Given that the objective of the application is structured query generation enhanced with document context, the PDF Agent introduces unnecessary redundancy

Considered changes: Remove the PDF Agent. Document retrieval and question enrichment responsibilities are already handled by the Analyst Agent.