
SQL LLM Agent

113-1 資工專題 金以凡

User Scenario

使用 SQL LLM Agent，User 只需以自然語言輸入問題，Agent 就會生成程式碼並執行相應的 SQL 查詢，最後(以自然語言)給出查詢結果。

適合：需要快速從資料庫中取得某些數據/趨勢，但不會用 SQL 的人

User Scenario

例如：

商業行銷決策人員在根據過往銷售數據制定新的策略時，會希望快速從資料庫中提取去年各地區的銷售趨勢或熱門產品資訊。

然而，由於他們不會SQL，因此需要資訊人員的協助。

現在，使用 SQL LLM Agent，他們只需以自然語言輸入問題，例如「去年各地區的月銷售額如何？請依銷售額排序」，Agent 便能快速給出回覆，讓他們能夠即時分析並制定將來的策略，提高效率。

User Scenario



User

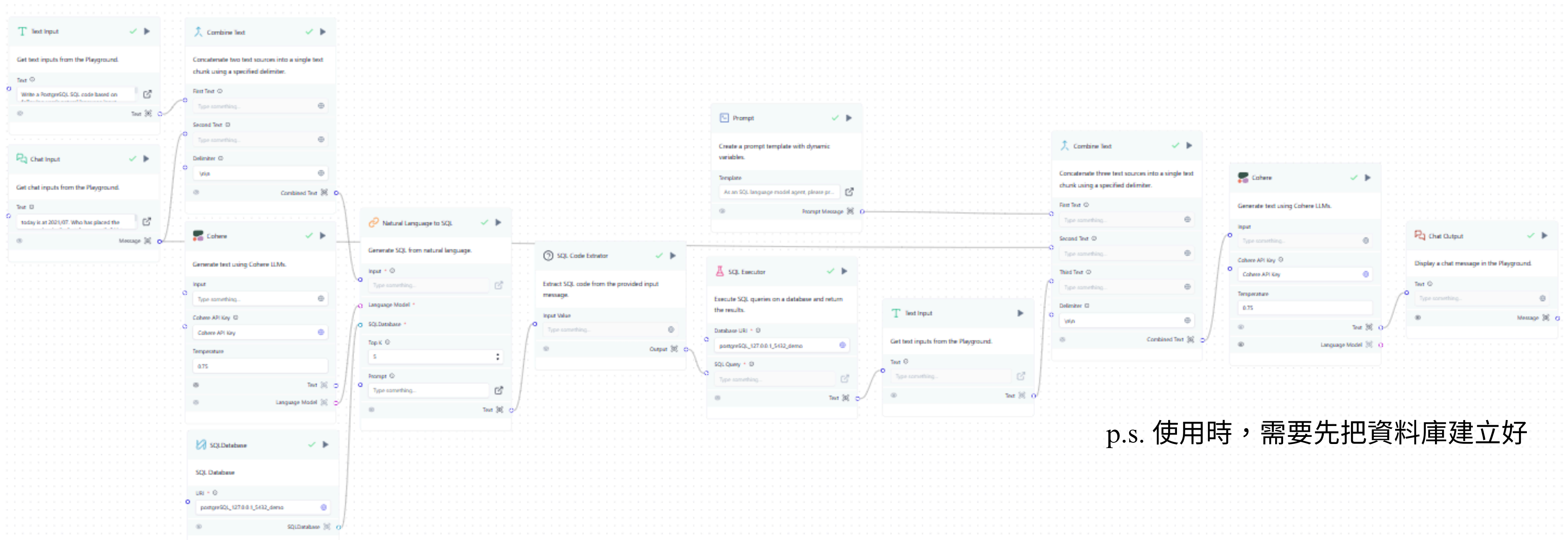
Who placed the most orders in July 2021? List the top three.



AI

In July 2021, the top three customers who placed the most orders were Hanari Carnes and Ernst Handel, both with two orders each, followed by Wartian Herkku with one order.

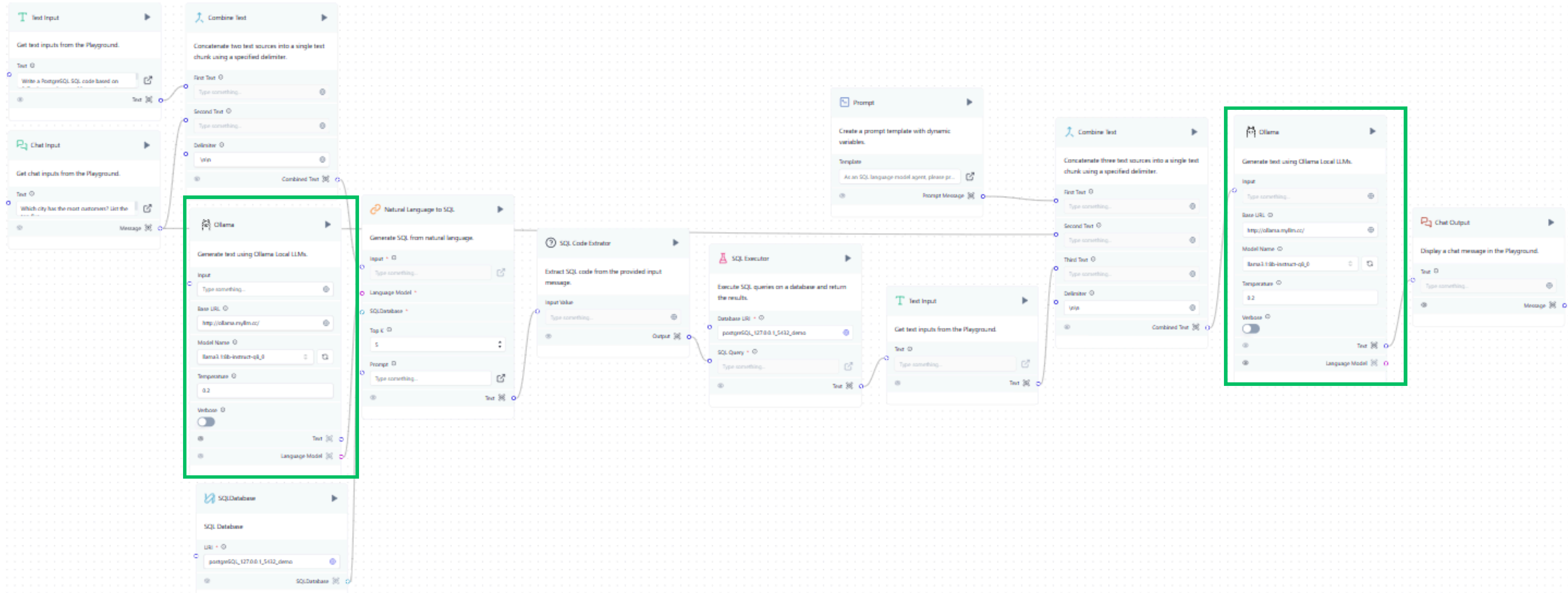
Overview (cohere.ver)



p.s. 使用時，需要先把資料庫建立好

用戶輸入 → 生成sql code → 執行sql code → 給出通順回答

Overview (ollama.ver)

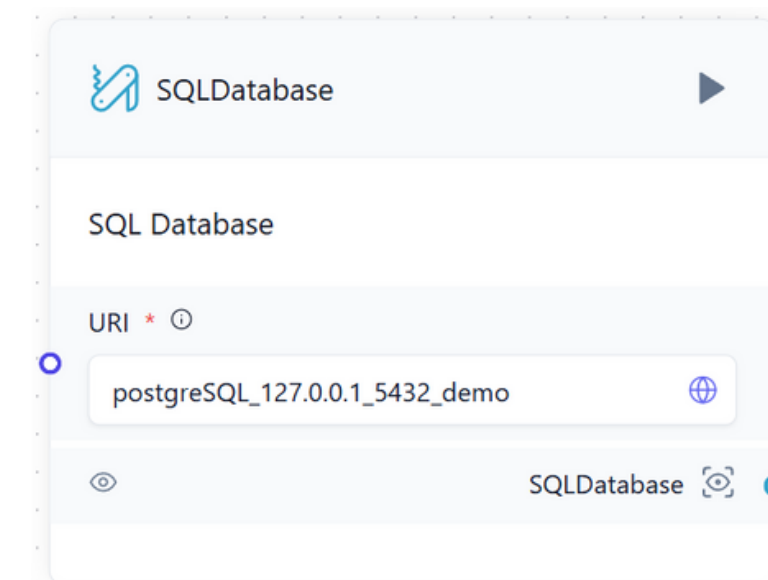


Step 0 - 預先架設資料庫

The example PostgreSQL database consists of six tables:

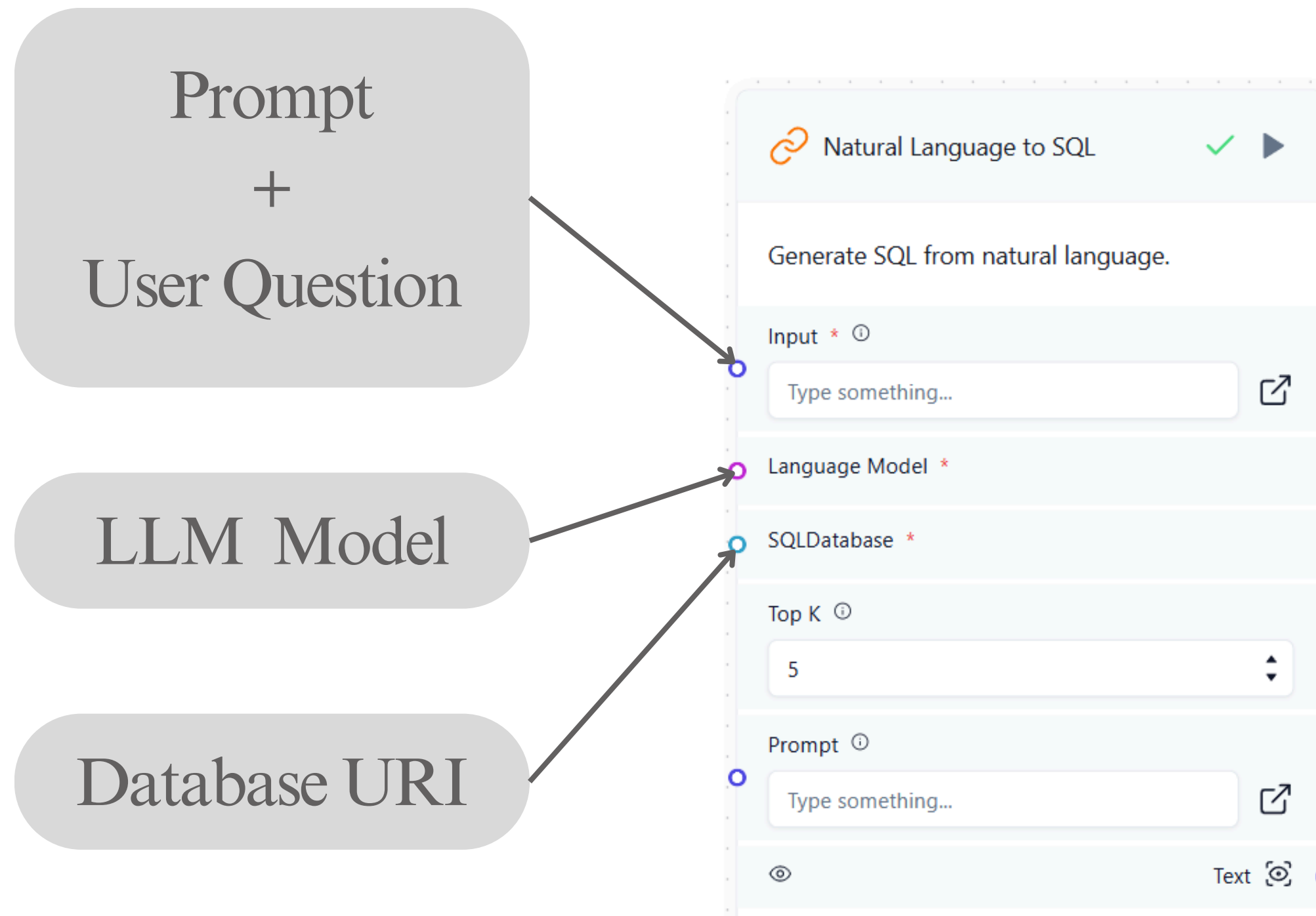
Table Name	Columns
categories	<code>category_id</code> (SERIAL, PK), <code>category_name</code> (VARCHAR), <code>description</code> (VARCHAR)
customers	<code>customer_id</code> (SERIAL, PK), <code>customer_name</code> (VARCHAR), <code>contact_name</code> (VARCHAR), <code>address</code> (VARCHAR), <code>city</code> (VARCHAR), <code>postal_code</code> (VARCHAR), <code>country</code> (VARCHAR)
products	<code>product_id</code> (SERIAL, PK), <code>product_name</code> (VARCHAR), <code>category_id</code> (INT), <code>unit</code> (VARCHAR), <code>price</code> (DECIMAL)
orders	<code>order_id</code> (SERIAL, PK), <code>customer_id</code> (INT), <code>order_date</code> (DATE)
order_details	<code>order_detail_id</code> (SERIAL, PK), <code>order_id</code> (INT), <code>product_id</code> (INT), <code>quantity</code> (INT)
testproducts	<code>testproduct_id</code> (SERIAL, PK), <code>product_name</code> (VARCHAR), <code>category_id</code> (INT)

- 本地端 (langflow+db)
- PostgreSQL
- URI (統一資源識別碼)



in langflow, 給定URI建立與資料庫連線

Step 1 - 用戶輸入到產生SQL Code



- Natural language to SQL
- LLM Model: Cohere/Ollama

Prompt:

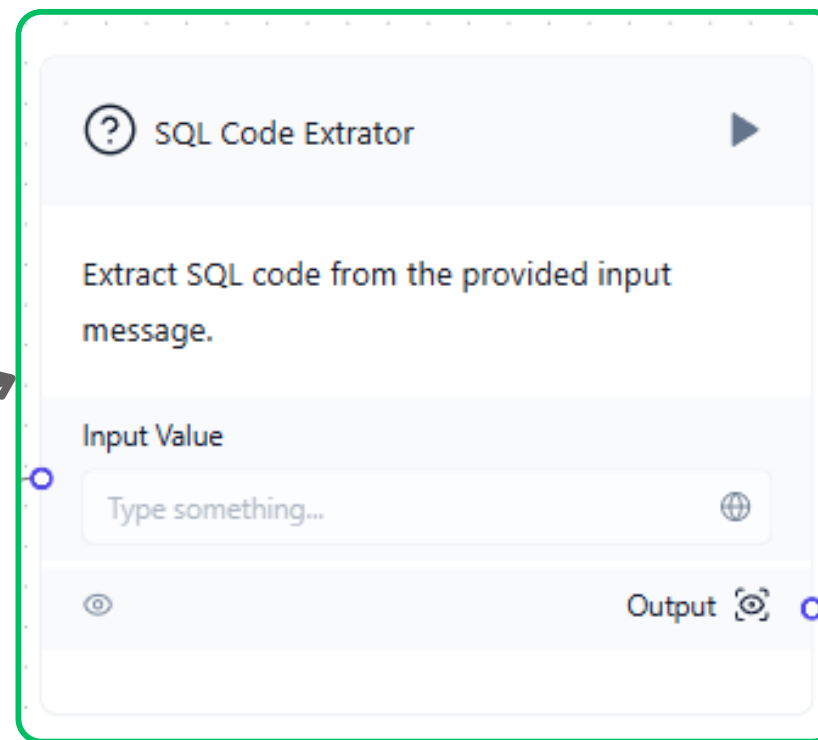
Write a PostgreSQL SQL code based on following user's natural language input.

- *Avoid using any date or time functions such as `NOW()`, `CURRENT_DATE`, `EXTRACT()`, `AGE()`, or similar date-related functions.*
- ...

Step 2 - 取出SQL Code

```
```sql
SELECT customer_name, COUNT(*)
FROM customers
 INNER JOIN orders ON customers.customer_id = orders.c
WHERE order_date BETWEEN '2021-07-01' AND '2021-07-31'
GROUP BY customer_name
ORDER BY 2 DESC
LIMIT 3;
```
```

markdown語法



```
SELECT customer_name, COUNT(*)
FROM customers
    INNER JOIN orders ON customers.customer_id = ord
WHERE order_date BETWEEN '2021-07-01' AND '2021-07-3
GROUP BY customer_name
ORDER BY 2 DESC
LIMIT 3;
```

pure sql code

Step 3 - 執行SQL Code

```
SELECT customer_name, COUNT(*)  
FROM customers  
    INNER JOIN orders ON customers.customer_id = orders.customer_id  
WHERE order_date BETWEEN '2021-07-01' AND '2021-07-31'  
GROUP BY customer_name  
ORDER BY 2 DESC  
LIMIT 3;
```

pure sql code

SQL Executor

Execute SQL queries on a database and return the results.

Database URI * ⓘ

postgresql_127.0.0.1_5432_demo

SQL Query * ⓘ

Type something...

Text ⓘ

Execution Result

Get text inputs from the Playground.

Text ⓘ

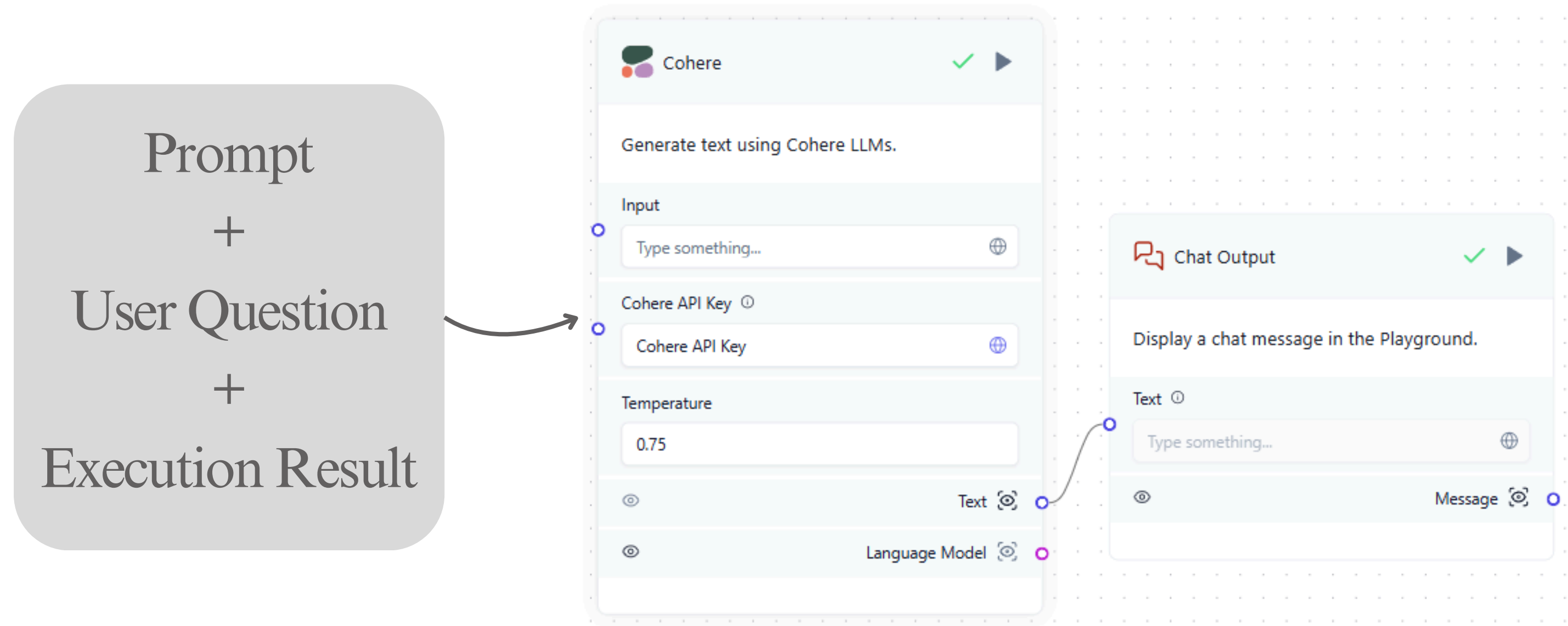
Type something...

Text ⓘ

| |
|-----------------------|
| ('Hanari Carnes', 2) |
| ('Ernst Handel', 2) |
| ('Wartian Herkku', 1) |

執行結果

Step 4 - 產生通順回答



prompt:

As an SQL language model agent, please provide a fluent answer by combining the question and the output. The user's question and execution result will be provided next.

- Make sure that your response is well-structured without bullet points, written as a continuous paragraph.*
- If there is no execution result, please respond that the query range is incorrect or no data was found.*

Experiments and Results

- 簡單的提問 e.g. How many tables are there in this database?



AI

There are 6 tables in this database.

- 複雜的提問 e.g. Who placed the most orders in July 2021? List the top three.



AI

In July 2021, the top three customers who placed the most orders were Hanari Carnes and Ernst Handel, both with 2 orders each, followed by Wartian Herkku with 1 order.

- 同時提問多個問題+錯誤文法 e.g. Most ordered product? Highest price product?



AI

The most ordered product is 'Raclette Courdavault', with a price of \$55.00 and 54 orders. The highest-priced product is 'Cote de Blaye', which costs \$263.50 and has been ordered 24 times.

落地前後性能比較

以"Which city has the most customers? List the top five."為例

Cohere

('London', 6)
('Mexico D.F.', 5)
('Sao Paulo', 4)
('Rio de Janeiro', 3)
('Madrid', 3)

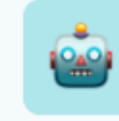


AI

The top five cities with the most customers are London, with 6 customers, followed by Mexico D.F. (5 customers), Sao Paulo (4 customers), and Rio de Janeiro and Madrid, both with 3 customers each.

Ollama

('London',)
('Mexico D.F.',)
('Sao Paulo',)
('Rio de Janeiro',)
('Madrid',)



AI

To determine which city has the most customers, we need to analyze the given data. The cities listed are London, Mexico D.F., Sao Paulo, Rio de Janeiro, and Madrid. However, there is no information provided about the number of customers for each city or any other relevant data that would allow us to accurately rank them based on customer count. Therefore, without additional context or a query range that includes this information, we cannot provide a definitive answer to which city has the most customers.

落地後遇到的瓶頸

- 生成程式碼的 LLM: 生成的 SQL Code 不夠完整
- 生成回覆的 LLM: 理解能力較弱

Demo