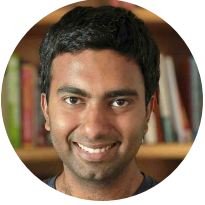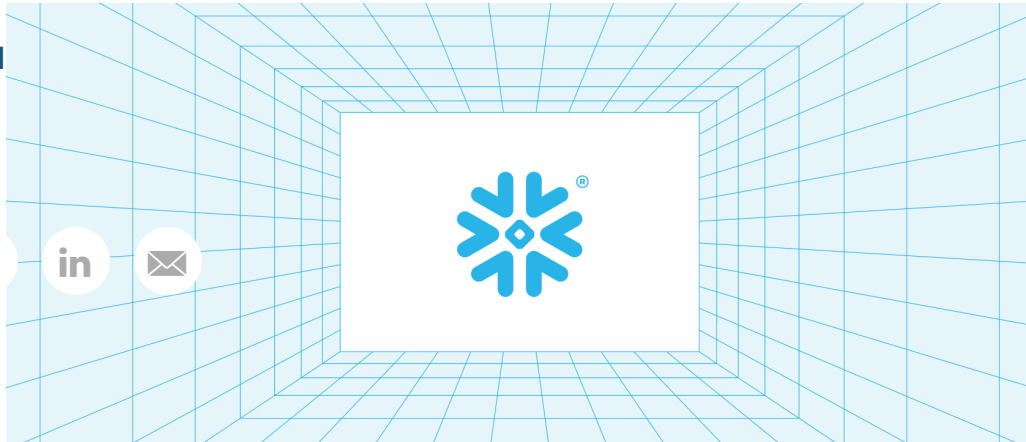NOV 12, 2024

AUTHOR

Viswajith Venugopal

Snowflake Cortex Analyst and Cortex Search: Better Together

Gen AI

Snowflake Cortex Analyst is a fully managed, LLM-powered service that provides a natural language interface for business users to receive accurate answers about structured data in Snowflake. Anyone who has written SQL queries to answer business questions knows that characterizing this problem as hard would be a wild understatement. While building Cortex Analyst, we have chipped away at this problem one step at a time and have now achieved an **accuracy of 90%** on our text-to-SQL data set, surpassing other solutions on the market.

Part of what makes text-to-SQL so hard is that writing the correct SQL query necessitates an understanding of what data is contained in the tables. What do the different columns represent? How are the different tables related? What is the data-generating process? What are the hidden caveats that could trip up the query? We address this through the creation of a **semantic model**, which captures this information.

Literals and sample values

One specific problem for generating correct SQL concerns *literals*. Let's say you want to filter on a column that contains constant strings. For the generated SQL to work, you need to find the correct literals to insert into the WHERE clause.

For example, suppose a business user asks: "How many orders did we have from California last month?" The semantic model includes an orders table with a `State` column. But what is the value in this column? Is it "California" or is it "CA"?

To support this, our semantic model specification includes a "sample_values" field on columns. If the values in the column are CA and UT, users can add those as sample values for the `State` column. The LLM generating the SQL would then know to say `WHERE STATE = 'CA'` in the generated SQL.

AUTHOR

Viswajith Venugopal

But this is an easy case, because there are a finite number of possible values. In fact, since U.S. states will be part of the world knowledge of the LLM, even putting one or two examples in the semantic model would allow the model to generalize; that is, LLMs would know to say `WHERE STATE = 'HI'` for Hawaii even if `HI` is not explicitly in the sample values.

What if the column values are not so straightforward to know? For example, if somebody asks: "How many customers ordered chicken biryani yesterday?" Let's say you have a `MENU_ITEM VARCHAR` column. What does it contain? Is it `chicken biryani`, `Chicken Biriyani`, `Biryani (Chicken)`? Or, as happens all too often, a mix of all of these? It would be too hard to capture all of these in sample values.

One option is to say `WHERE MENU_ITEM ilike '%chicken biryani%'`, but that would be brittle.

Which brings us to the more elegant solution to this problem...

Snowflake Cortex Search integration

Snowflake Cortex Search is a Snowflake service that offers high-quality, low-latency "fuzzy" search over data stored in tables. Users can create a Cortex Search service, point it to data in a Snowflake table, and within minutes, Snowflake will build an index that allows searching the data.

Cortex Search uses hybrid retrieval, combining keyword retrieval (enhanced with synonym expansion) and vector retrieval. Retrieved results are scored using a series of highly tuned scoring functions, and finally passed through a re-ranker model that adjusts the results based on query-to-document semantic similarity. All this in a few hundred milliseconds.

The solution to the problem we outlined previously now presents itself: For columns like `MENU_ITEM`, index it using a Cortex Search service, and
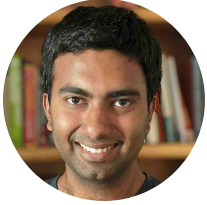
when generating the SQL, first query the Cortex Search service to fetch the literals most similar to the user's query. These can then be passed to the LLM that generates SQL.

We are excited to announce that Cortex Analyst now has first-class support for this integration. It is as simple as adding the name of the Cortex Search service associated with a column in the semantic model, similar to the below:

AUTHOR
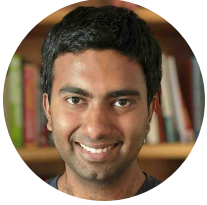
Viswajith Venugopal

```
tables:

  - name: my_table

    base_table:
      database: my_database
      schema: my_schema
      ble: my_table

    dimensions:
      - name: my_dimension
        expr: my_column
        cortex_search_service_name:
my_logical_dimension_search_service
```

Summary

Snowflake Cortex Analyst aims to solve the real-world text-to-SQL problem, making improvements one step at a time to produce ever-more-reliable SQL. Snowflake's powerful Cortex Search feature makes it possible for us to efficiently index and do fuzzy search over a column to solve the literal retrieval problem.

We invite you to try all the new **Cortex Analyst features**, including Cortex Search integration, and share your feedback. Here are some additional valuable resources to help you get started:

**BUILD Cortex Analyst Announcements**: Check out our **roundup blog** for the latest feature updates.

Quickstart Guide: Use our semantic model generator tooling to **create semantic models for Cortex Analyst**.

GitHub Samples Repository: Discover **inspiring examples** on how to put Cortex Analyst to use.

Third-Party Semantic Layers: Learn how to **translate existing third-party semantic layers for use with Cortex Analyst**.

Stay tuned as we continue to enhance support for enterprise-grade, AI-driven self-serve analytics!

AUTHOR

SHARE   f   in   ✉
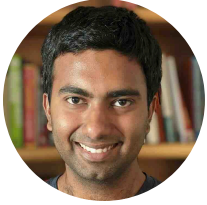
Viswajith Venugopal

RELATED CONTENT

SHARE   f   in   ✉

AUG 14, 2024

Snowflake Cortex Analyst: Behind the Scenes

Building a conversational self-service analytics product for business users is a complex and challenging endeavor. Trust and accuracy have to be at the core of such a product, as business…

See how

AUG 29, 2024

Snowflake Cortex Analyst: Evaluating Text-to-SQL Accuracy for Real-World Business Intelligence Scenarios

Being able to build a system that allows business users to ask intricate, complex and…

Here's How

AUG 08, 2024

Snowflake Cortex Search: High-Quality, Performant Search and Retrieval for Enterprise AI

Search and retrieval systems have always been a critical backbone for knowledge management in enterprises….

Discover

Snowflake Inc.

PLATFORM

Cloud Data Platform

SOLUTIONS

Snowflake for Financial Services

RESOURCES

Resource Library

EXPLORE

Blog

Trending

ABOUT

About Snowflake

Pricing

Marketplace

Security & Trust

Snowflake for Advertising, Media, & Entertainment

Snowflake for Retail & CPG

Healthcare & Life Sciences Data Cloud

Snowflake for Marketing Analytics

Webinars

Documentation

Community

Procurement

Legal

Guides

Developers

Investor Relations

Leadership & Board

Snowflake Ventures

Careers

Contact

AUTHOR
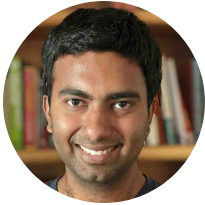
Viswajith Venugopal

SHARE

f   in   ✉

Sign up for Snowflake Communications

diana.shaw@snow

United States

By submitting this form, I understand Snowflake will process my personal information in accordance with their Privacy Notice. Additionally, I consent to my information being shared with Event Partners in accordance with Snowflake's Event Privacy Notice. I understand I may withdraw my consent or update my preferences here at any time.

SUBSCRIBE NOW