

UNLEASHING DATA-DRIVEN DECISION MAKING: HOW SNOWFLAKE CORTEX ANALYST PLAYS A PIVOTAL ROLE IN ENTERPRISE AI

Overview/Industry perspective

The convergence of data and artificial intelligence (AI) is ushering in a new era of hyperconnected ecosystems. As data volumes continue to explode, AI-driven technologies are becoming essential for extracting valuable insights and driving innovation across industries.

As enterprise systems become increasingly complex, the data and AI landscape is undergoing rapid expansion. Traditional on-premises systems are being complemented by cloud-based solutions, clickstream data, social media feeds, and third-party data aggregators. This proliferation of data sources is driving exponential growth in data consumption, fueled by diverse consumers, platforms, and analytical needs. This evolution is fostering innovation in how data is consumed, shared, and

monetized, creating new opportunities and challenges for organizations navigating the modern data landscape.

Snowflake Cortex Analyst serves as a powerful apparatus in this arena, offering users answers within the Snowflake AI Data Cloud. Cortex Analyst bridges the gap between applications and structured data - information stored in table format - by enabling natural language queries, meaning it lets you ask questions in natural language from any application. Cortex Analyst turns that into a SQL query that retrieves the response with its industry-leading text-to-SQL accuracy.

To unlock the full potential of Cortex Analyst, enterprises should continuously work to effectively prepare their data for AI initiatives – Data Ready for AI, enabling more accurate models and better decision-making.

Why do I need knowledge assistants for enterprise AI

Knowledge Assistants, specifically Text-to-SQL Assistants, helps to **democratize access to insights** to all levels of users in an enterprise – both technical and non-technical users, to easily extract insights & understand enterprise data in structured data stores like databases.

These AI assistants help data teams to focus more on supporting strategic activities for business growth and **reduce their time & effort** in tasks to get answers for ad hoc questions. These AI assistants also help business users to ask custom questions, stuff not in a dashboard, and enable more data driven decision making. This results in an improved relevance of data assets supporting business growth.

Knowledge Assistants using LLMs have issues like very low text-to-SQL accuracy, limited scalability, limited access to latest domain data, more prone to bias and hallucinations. All these issues deter data teams from implementing reliable AI solutions that can enable business teams to take full advantage of their enterprise data.



Why do we need agentic knowledge assistants for enterprise AI

Agentic knowledge assistants can significantly improve the results of text-to-SQL by leveraging LLMs and ML models' ability to understand natural language, reason over structured knowledge bases, and generate structured queries.

Agentic knowledge assistants generate reliable results by implementing capabilities like semantic interpretation, schema mapping, dynamic query generation, and iterative query refinement.

Each agent in the knowledge assistant solutions have well defined goals for their tasks, and the assistants are tailored to specific use cases. So, this approach helps to create effective and engaging conversational agents.



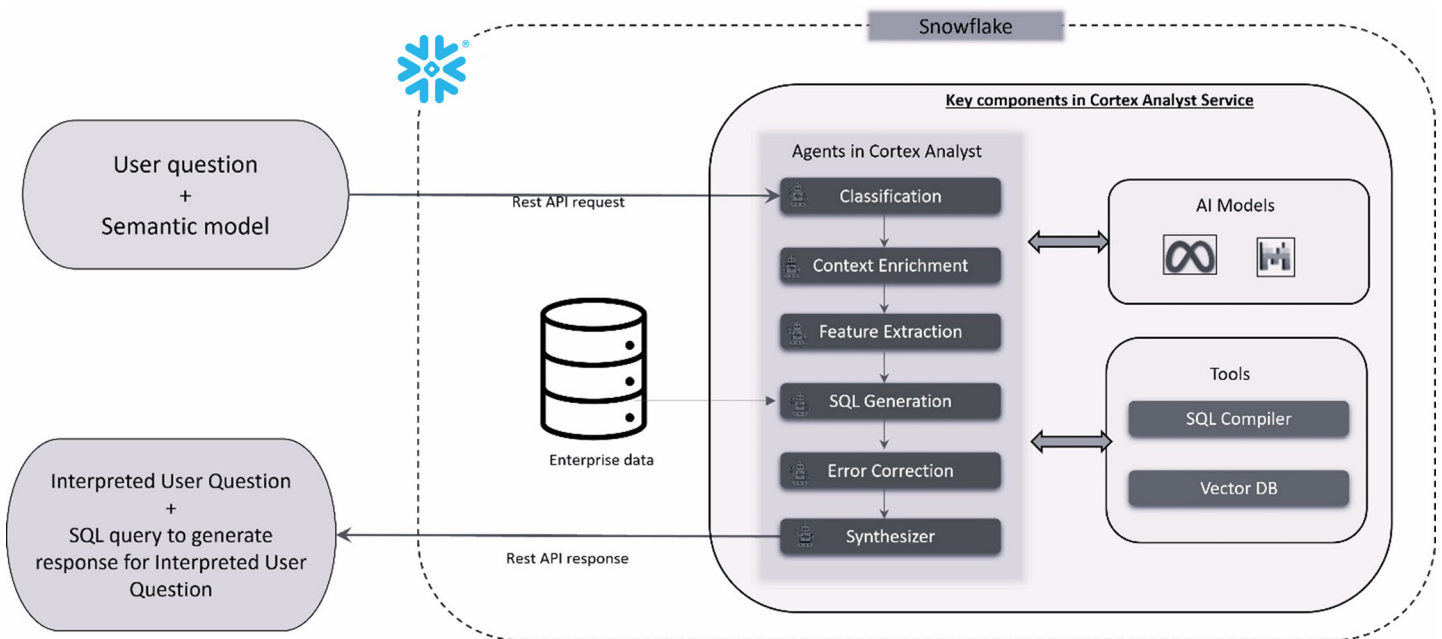
Key traits of the offering/solutions:

Cortex Analyst service abstracts the complex agentic solution components in text-to-SQL assistant with simple single rest API call. This enables easy integration of knowledge assistant i.e. text-to-SQL assistant as a service to existing business applications such as slack or existing business intelligence apps and Infosys Industry solutions like [Infosys Smart Manufacturing Intelligence Cloud](#).

In Cortex Analyst, in addition to using LLMs as the brain, semantic context is provided as input when processing our natural text queries. Adding semantic context enables agents to provide more

reliable queries considering the vocabulary of the business teams. This approach mimics the repeated work of the human analyst where the analyst would constantly analyze the data using their understanding and relationships of the different metrics and business context.

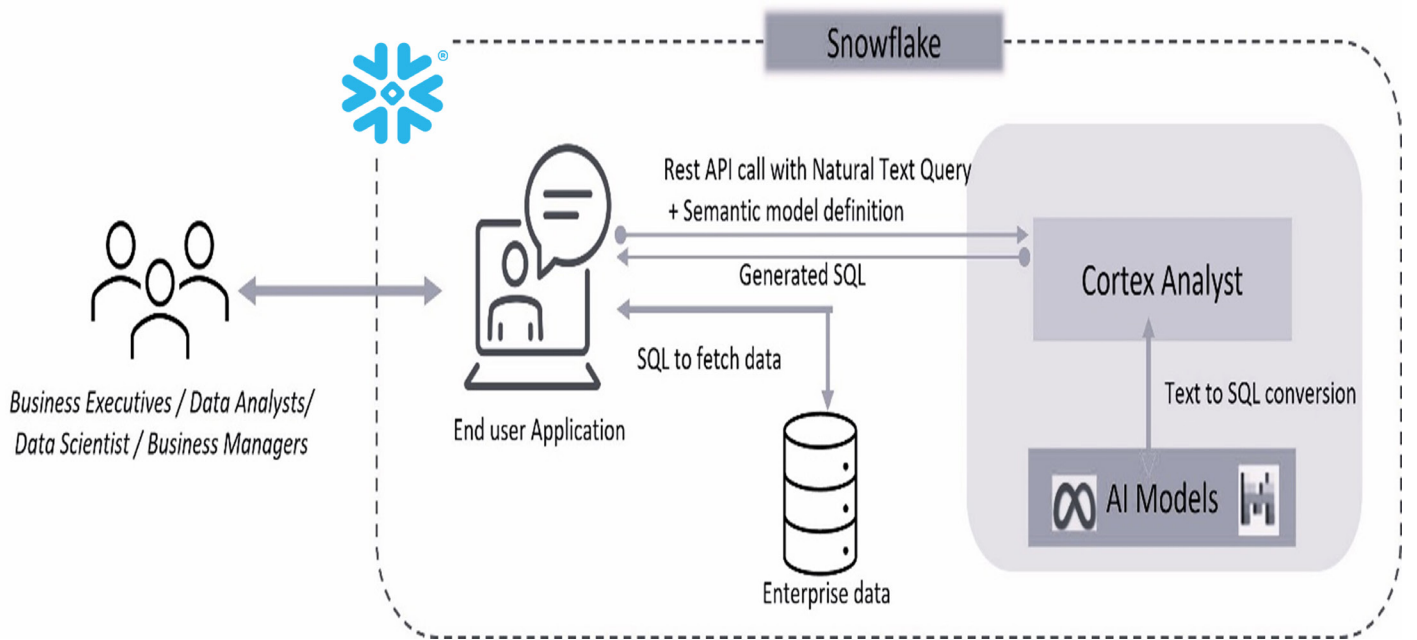
Below is the representation of inner workings of how natural text query is converted to SQLs that can generate insights through agents in Cortex Analyst service.



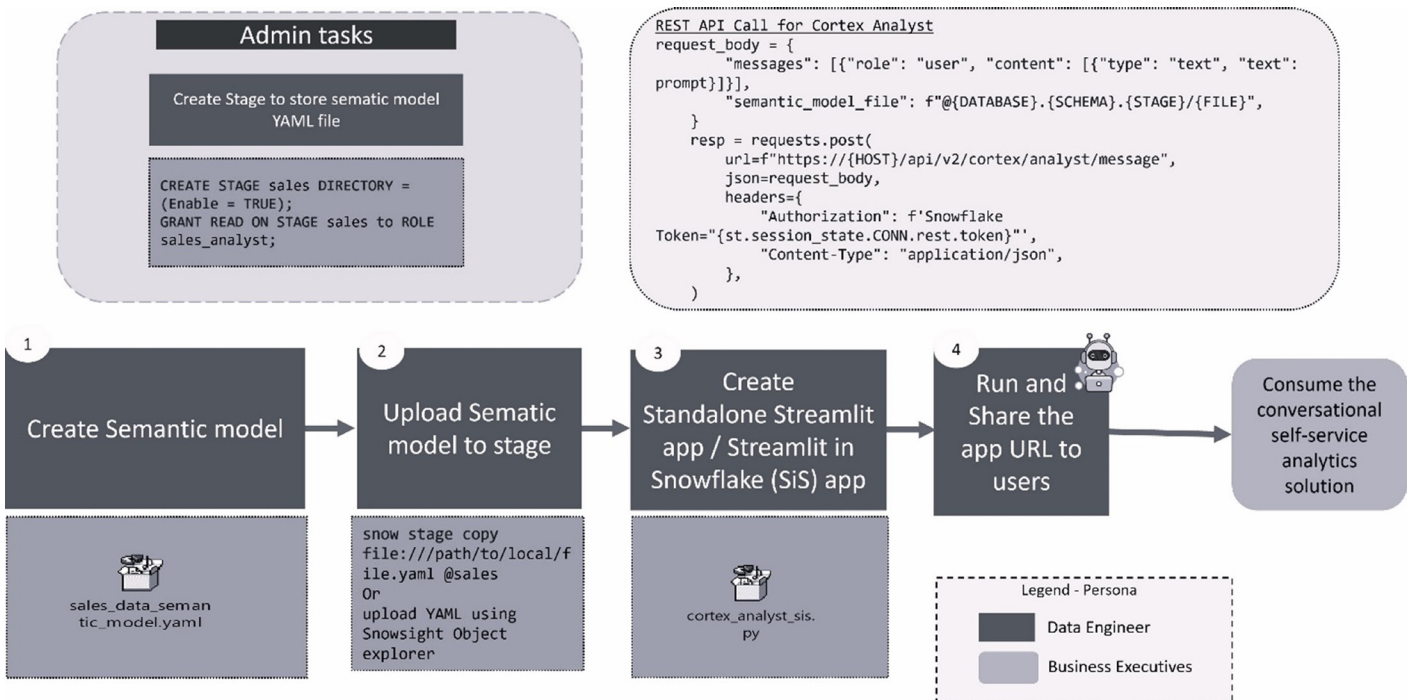
Each of the agents in the Cortex Analyst performs task specific targeted goals assigned to it and passes on the result for the next agent in the solution. Each agent is designed with guardrails at every step to reduce hallucinations and improve the quality of the query generated.



Below diagram represents the high-level view of a business intelligence application integrated with Cortex Analyst based assistants to generate reports and insights.



Below are the summary of steps involved in building Cortex Analyst based Assistant service.

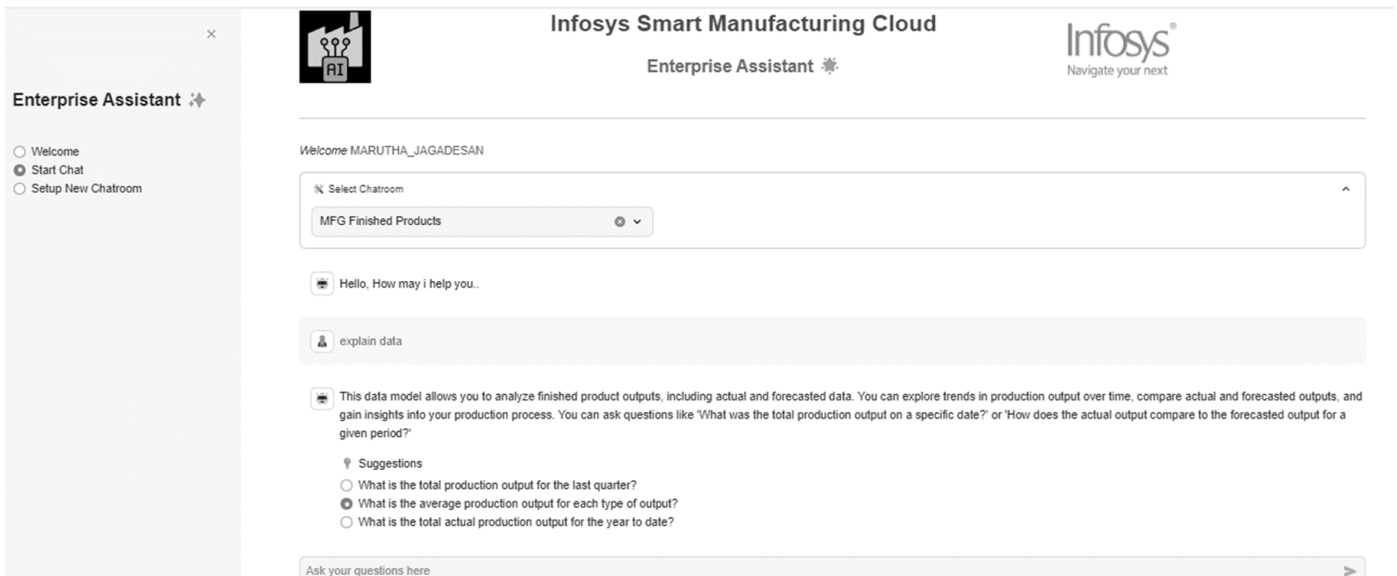


Our implementation

Chat rooms capability built using Snowflake Cortex Analyst service is integrated to one of our Industry solutions - [Infosys Smart Manufacturing Intelligence Cloud](#). This feature enhanced the user experience with the solution in Data analytics.

Below are the key feature of the chat room capability added to Infosys Manufacturing Intelligence Cloud:

-  Powered by Snowflake Cortex Analyst
-  Implementation of Agentic solution including guardrails for AI models is abstracted by Cortex Analyst service
-  Simple UI driven schema definition management
-  Automated description population for initial version of schema definition



Infosys Smart Manufacturing Cloud
Enterprise Assistant

Welcome MARUTHA_JAGADESAN

Select Chatroom
MFG Finished Products

Hello, How may i help you..

explain data

This data model allows you to analyze finished product outputs, including actual and forecasted data. You can explore trends in production output over time, compare actual and forecasted outputs, and gain insights into your production process. You can ask questions like 'What was the total production output on a specific date?' or 'How does the actual output compare to the forecasted output for a given period?'

Suggestions

- What is the total production output for the last quarter?
- What is the average production output for each type of output?
- What is the total actual production output for the year to date?

Ask your questions here



What is the average production output for each type of output?

This is our interpretation of your question:
What is the average production output for each type of output over the entire available time period?

Response

Data Line Chart Bar Chart

Category	Production Output
Actual	~280,000
Forecast	~270,000

Generated SQL

Ask your questions here

Our learnings in implementing solutions using Cortex Analyst assistants

Joins are supported with high accuracy. For complex joins, recommended is to use views with pre-joined data

Semantic model file generation can be automated, but Synonyms (column & table description) needs to be manually written

Case sensitive issue while applying the query on underlying table field values. Work-Around: By giving the sample values on semantic model

Solution is benefitted with Snowflake integrated security and governance features

Text to SQL accuracy is not good for complex queries – Work-Around: Adding Verified query repository (VQR) improves accuracy of results for standard queries

Cortex Analyst is free until November 15, 2024. Pricing is yet to be released

Why do we need fully managed agentic solution like Snowflake Cortex Analyst



Higher TCO: Using cloud-based solutions like Snowflake Cortex offer managed, tightly integrated services and AI infrastructure for NLP, machine learning, and conversational applications. This provides benefits of scalability, accelerates development and reduces cost of data movement across agents.

Governance & Security: Customer data resides within the customer's Snowflake environment and doesn't get transferred to external hosting services when using LLMs.



Productivity: his feature helps to build solutions that improve Human-AI Collaboration and augment human resource potential in organizations.

Use cases that would benefit from Cortex Analyst – Fully managed agentic text-to-SQL assistant

Organization's units / solutions that have repeated tasks for data analytics requirement would benefit from using Cortex Analyst like:

Creating personalized training programs

Feedback analysis and insights delivery

Sentiment analysis

Marketing campaign planning

Snowflake Cortex Analyst being fully managed service enables organizations to build reliable knowledge assistants on their Enterprise data and provide quality platforms for Data explorations and analytics.

Details & use cases discussed in this point of view helps to understand the possibilities of creating high accuracy Text to SQL assistants using Cortex analyst feature. The feature is evolving for better, along with fast growing AI adoption. So for latest details on the Snowflake Cortex feature, as well as best practices for obtaining more accurate results, be sure to check out the [Snowflake Cortex Analyst documentation](#).

Infosys Topaz is an AI-first set of services, solutions and platforms using generative AI technologies. It amplifies the potential of humans, enterprises and communities to create value. With a vast repository of AI assets, pre-trained AI models, 10+ AI platforms steered by AI-first specialists and data strategists, and a 'responsible by design' approach, Infosys Topaz helps enterprises accelerate growth, unlock efficiencies at scale and build connected ecosystems.

For more information, contact askus@infosys.com



© 2024 Infosys Limited, Bengaluru, India. All Rights Reserved. Infosys believes the information in this document is accurate as of its publication date; such information is subject to change without notice. Infosys acknowledges the proprietary rights of other companies to the trademarks, product names and such other intellectual property rights mentioned in this document. Except as expressly permitted, neither this documentation nor any part of it may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, printing, photocopying, recording or otherwise, without the prior permission of Infosys Limited and/or any named intellectual property rights holders under this document.