International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

# Unleashing Potential: Snowflake's Streamlit Strategy for GenAI Solutions Leveraging External Network Access and OpenAI

**Author:** Mala Kamboj, Chitresh Goyal, Neelima Ratra

**Abstract:**

This whitepaper explores the synergy between Snowflake's External Network Access, OpenAI, and Streamlit, presenting a groundbreaking approach to data analysis and interaction. By seamlessly integrating these technologies, organizations can unlock new dimensions of insight and efficiency in their data workflows. We delve into the configuration of External Network Access in Snowflake, the creation of Snowpark stored procedures for interacting with external services like OpenAI, and the development of user-friendly native applications with Streamlit. This innovative solution empowers users to access, analyze, and communicate with their data stored on Snowflake with unparalleled ease and sophistication, revolutionizing data-driven decision-making processes.

**Keywords:** Snowflake, Streamlit, Generative AI , Open AI, Snowflake External Network Access

## Introduction:

In today's data-centric landscape, organizations are on a continuous quest for innovative solutions to fully leverage their data assets. Snowflake, a leader in data warehousing and analytics, is at the forefront of this revolution. The integration of Snowflake's External Network Access feature with the capabilities of OpenAI, combined with the intuitive Streamlit framework, has opened up a new frontier in data analysis and interaction.

This whitepaper explores the synergy of Snowflake's strategy for GenAI solutions with Streamlit, alongside the advanced functionalities provided by External Network Access and OpenAI. By integrating these technologies seamlessly, organizations can unlock the full potential of their data, enabling users to derive actionable insights and make informed decisions with ease and efficiency. Throughout this whitepaper, we

will delve into the architecture, implementation, and benefits of this innovative solution. From configuring External Network Access in Snowflake, to leveraging Snowpark stored procedures for interaction with OpenAI, and developing user-friendly native applications with Streamlit, each aspect of this integration contributes to a holistic data analysis ecosystem.

**Problem Statement:**

Organizations are continually striving to unlock actionable insights from their vast data repositories, especially when stored in cloud-based warehouses like Snowflake. The traditional approaches to data analysis can be laborious and slow, often requiring manual efforts and the use of multiple, disconnected tools. Additionally, the task of securely accessing external resources for advanced analysis within a cloud environment presents its own set of technical and security challenges. There is a pressing need for a comprehensive solution that seamlessly integrates Snowflake's native capabilities with external services such as OpenAI, ensuring secure connectivity and efficient data analysis. Overcoming these challenges is crucial for organizations to unlock valuable insights and secure a competitive advantage in the ever-evolving data-centric landscape.

**Key Challenges:**

- **Complexity of External Network Access:** Configuring secure access to external network locations from within Snowflake's environment requires expertise in network configuration and authentication mechanisms, posing challenges for users with limited technical knowledge.

- **Integration with External Services:** Leveraging the capabilities of external services, such as OpenAI, for data analysis requires seamless integration with Snowflake's data warehouse, necessitating the development of custom solutions and procedures.

- **Usability and Accessibility:** Providing a user-friendly interface for querying, analyzing, and visualizing data stored in Snowflake is essential to ensure accessibility and usability for users across different roles and levels of technical expertise.

**Objective:**

The objective of this whitepaper is to showcase how organizations can overcome these challenges and unlock the full potential of their data stored in Snowflake by leveraging Snowflake's External Network Access feature and integrating external services, such as OpenAI, with Snowflake's native capabilities. Through a combination of practical examples, best practices, and implementation guidelines, this whitepaper aims to empower organizations to streamline their data analysis workflows and derive actionable insights from their data more efficiently.

**Technical Architecture:**

**Understanding External Network Access:**

Snowflake's External Network Access feature empowers users to establish secure connectivity to external network locations directly from within the Snowflake environment. This capability opens doors to a plethora of possibilities for integrating external resources seamlessly into data workflows, enhancing the versatility and power of Snowflake's data platform.

By configuring external access integrations and defining permitted network rules and authentication secrets, users can harness the full potential of this feature. With an external access integration, users gain the ability to:

- Write user-defined functions (UDFs) and procedure handlers that seamlessly access external locations.
- Define and manage access permissions for locations on networks external to Snowflake.
- Utilize secrets, representing stored credentials, within handler code to authenticate securely with external network locations.
- Specify and manage which secrets are permitted for use with external network locations, ensuring robust security and access control.

In this whitepaper, we delve into the comprehensive capabilities of Snowflake's External Network Access feature and explore its integration with user-defined functions (UDFs) and stored procedures. By leveraging

this feature alongside OpenAI and the Streamlit framework, organizations can revolutionize their data analysis workflows, unlocking unprecedented insights and efficiencies.

**Streamlit: Empowering Data Analysis**

Streamlit is a powerful and user-friendly framework for building interactive web applications. With its intuitive interface and rich set of features, Streamlit enables developers to create dynamic and responsive applications with minimal effort. From data visualization to machine learning models, Streamlit offers a seamless platform for showcasing insights and driving actionable decisions.

**Creating the Snowflake Native App**

The Snowflake Native App is designed to provide users with a user-friendly interface for querying, analyzing, and visualizing data stored in Snowflake. Leveraging the capabilities of Streamlit, the app offers a seamless and intuitive experience for exploring and interacting with data.

**Key Features of the Snowflake Native App:**

- **Querying Data:** Users can execute SQL queries directly within the app, leveraging Snowflake's powerful querying capabilities to retrieve and manipulate data.
- **Data Analysis:** The app provides tools for performing advanced data analysis, including filtering, aggregation, and visualization, empowering users to derive actionable insights from their data.
- **Visualization:** With built-in support for interactive charts and graphs, the app enables users to visualize their data in meaningful ways, enhancing comprehension and decision-making.
- **User-Friendly Interface:** Streamlit's intuitive interface makes the app easy to navigate and use, even for users with limited technical expertise.

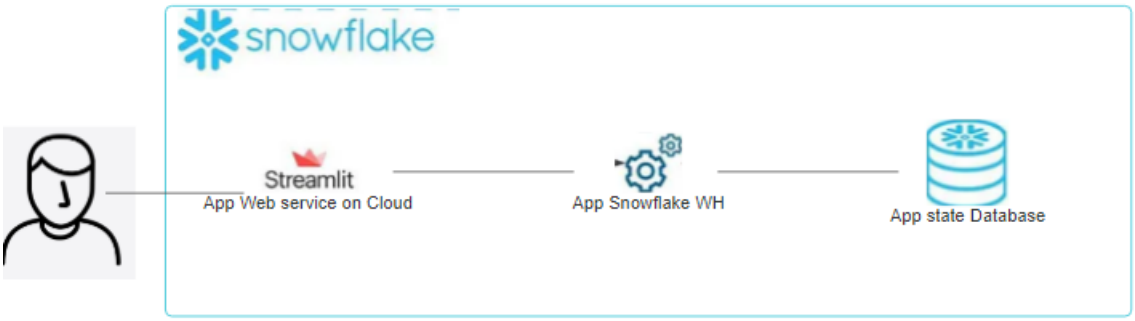
**Architecture of the Snowflake Native App:**

The architecture of the Snowflake Native App comprises three main components:

**Frontend:** The frontend of the app is built using Streamlit, providing the user interface for interacting with the data.

**Backend:** The backend logic, including data retrieval and processing, is handled by Snowflake's querying engine.

**Snowflake Database:** The Snowflake database stores the data used by the app, providing a secure and scalable storage solution.




Architecture of the Snowflake Native App

**Benefits of the Snowflake Native App:**

- **Enhanced Accessibility:** The user-friendly interface of the app makes it easy for users to access and analyze data, regardless of their technical expertise.
- **Improved Usability:** Streamlit's intuitive features streamline the data analysis process, enabling users to derive insights more efficiently.
- **Empowered Decision-Making:** By providing a platform for interactive data exploration and visualization, the app empowers users to make data-driven decisions with confidence.

  The Snowflake Native App, powered by Streamlit, offers a powerful and intuitive solution for analyzing data within the Snowflake environment. By combining the scalability and security of Snowflake with the ease of use of Streamlit, organizations can unlock new possibilities for data-driven innovation and decision-making.

**Technical Steps:**

1. **Setting up Snowflake Environment:**

This step involves creating and configuring a Snowflake environment, which includes tasks such as:

- Creating a Snowflake account: Registering for a Snowflake account and selecting the appropriate plan.
- Setting up warehouses: Configuring virtual warehouses to process queries and data loading tasks.
- Creating databases and tables: Defining databases and tables to store and organize data.
- Configuring security settings: Setting up user roles, permissions, and authentication methods to ensure data security.

2. **Configuring External Network Access:**

In this step, External Network Access is enabled in Snowflake to allow connectivity to external services. This involves:

- Defining network rules: Specifying the network locations and protocols allowed for external access.
- Managing authentication secrets: Creating and managing authentication secrets (such as API keys) to authenticate with external services.
- Creating external access integrations: Configuring integrations that combine network rules and authentication secrets to enable external access.

3. **Developing Snowpark Stored Procedures:**

Snowpark stored procedures are developed to leverage External Network Access for interacting with external services. This step includes:

- Writing code using Snowpark: Using Snowpark, Snowflake's native programming environment, to develop stored procedures.
- Authenticating with external services: Incorporating authentication logic within stored procedures to authenticate with external services using the configured secrets.
- Interacting with external data: Implementing logic to retrieve data from external services, perform data processing or analysis, and store results in Snowflake.

```
CREATE OR REPLACE PROCEDURE sp_data_analysis_genAI(table_name STRING, prompt_ques STRING)
returns STRING
LANGUAGE PYTHON
RUNTIME_VERSION = 3.9
HANDLER = 'genai_table_analysis'
EXTERNAL_ACCESS_INTEGRATIONS = (openai_access_integration)
PACKAGES = ('snowflake-snowpark-python','aiohttp', 'aiosignal', 'async-timeout', 'attrs', 'certifi',
'charset-normalizer', 'colorama', 'frozenlist', 'idna', 'multidict', 'requests', 'tqdm', 'urllib3', 'yarl')
IMPORTS = ('@test_stage/openai.zip')
SECRETS = ('openai_api_key'=openai_api_key)
AS
$$
import _snowflake
import logging
import openai
import sys
from snowflake.snowpark.functions import listagg

def genai_table_analysis(session,table_name,prompt_ques)->str:
    logging.basicConfig(level=logging.DEBUG)
    logger = logging.getLogger(__name__)
    df_table = session.table(table_name)
    df_columns = ','.join(df_table.columns)
    select_cols = df_columns.replace(',','||\' | \'||')
    logger.info(select_cols)
    tb_data = session.sql('SELECT '+select_cols+' AS col FROM '+table_name+' limit 50 ')
    output_set = df_columns.replace(',',' | ')
    logger.info(output_set)
    for rows in tb_data.toLocalIterator():
        output_set += str(rows[0]) + '\n'
        logger.info('data loaded - ' + output_set)
    text = prompt_ques + '\n\n' + output_set
    openai.api_key = _snowflake.get_generic_secret_string('openai_api_key')
    response = openai.ChatCompletion.create(
                    model="gpt-3.5-turbo-0613",
                    messages=[{ "role": "user",
                    "content":text}],
                    temperature=0,)
```

**Snowpark Stored Procedure with initial setup of external access**

### 4. Creating Streamlit Native App:

A Streamlit native app is developed to provide a user-friendly interface for querying, analyzing, and visualizing data stored in Snowflake. This involves:

- Setting up Streamlit environment: Installing Streamlit and configuring the development environment.
- Developing interactive dashboards: Using Streamlit's Python-based framework to create interactive data exploration tools, dashboards, and visualizations.
- Integrating with Snowflake: Incorporating Snowflake queries and data retrieval logic within the Streamlit app to fetch and display data from Snowflake databases and tables.

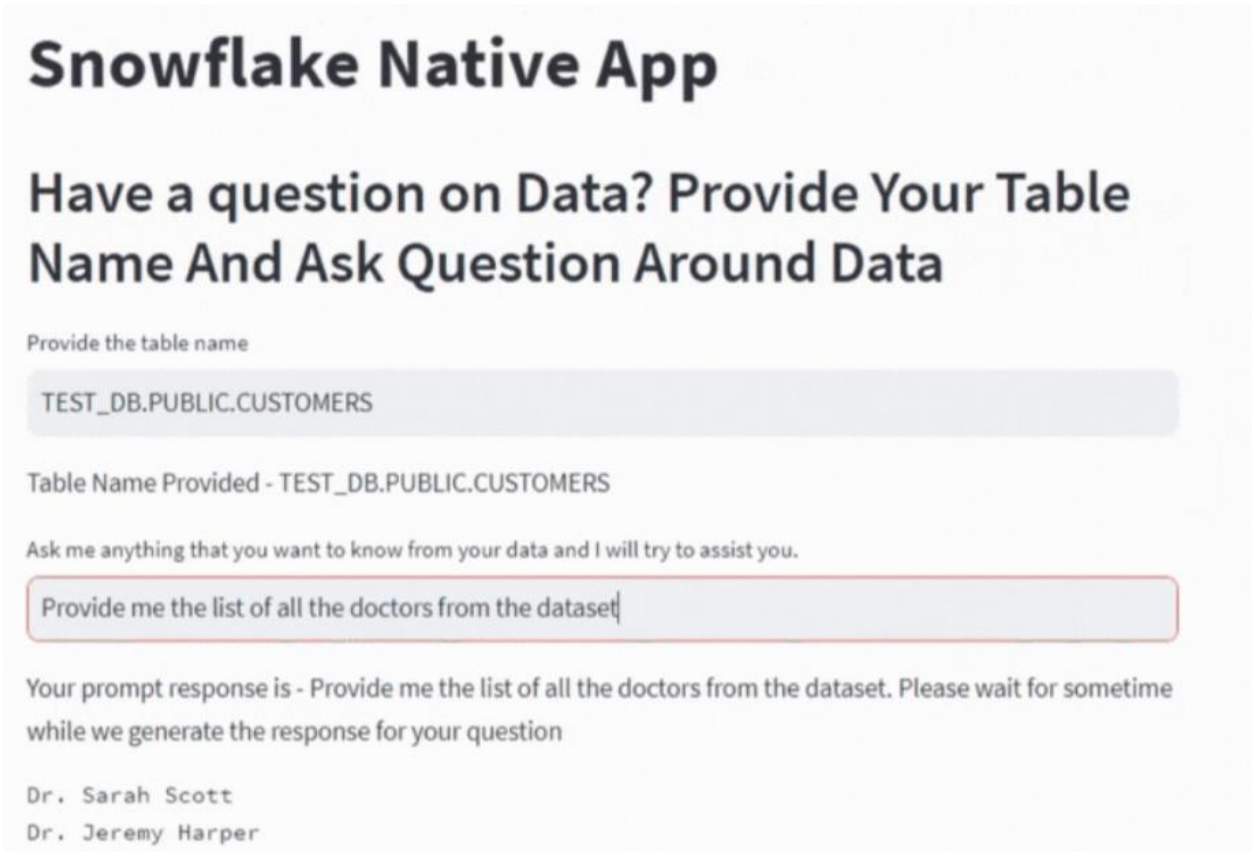
```python
import streamlit as st
import time
from snowflake.snowpark.context import get_active_session

session = get_active_session()

st.title("Snowflake Native App")
st.header(
    "Have a question on Data? Provide Your Table Name And Ask Question Around Data"
)

table_name = str(st.text_input("Provide the table name")).upper()
if table_name != "":
    st.markdown(f"Table Name Provided - {table_name}")
    prompt = st.text_input(
        "Ask me anything that you want to know from your data and I will try to assist you."
    )
    if str(prompt) != "":
        st.markdown(
            f"Your prompt response is - {prompt}. Please wait for sometime while we generate the response for your question"
        )
        op = session.call(
            "TEST_DB.PUBLIC.sp_data_analysis_genAI", f"{table_name}", f"{prompt}"
        )
        st.text(op)
```

**Code For Streamlit Native App**



**Output of Native App**

5. **Integration and Testing:**

The developed Snowpark stored procedures and Streamlit native app are integrated with the Snowflake environment, and thorough testing is conducted to ensure:

- **Functionality**: The integrated solution performs as expected and meets the specified requirements.
- **Performance**: The solution operates efficiently, with acceptable response times for data retrieval and analysis tasks.
- **Security**: Data access controls and authentication mechanisms are properly implemented to protect sensitive information.
- **Usability**: The user interface of the Streamlit app is intuitive and user-friendly, facilitating easy data exploration and analysis.

These steps collectively enable organizations to harness the power of Snowflake's capabilities and external services like OpenAI for advanced data analysis and decision-making.

**Solution Feature:**

The key feature of this application is, its ability to democratize data analysis by providing users with a powerful yet user-friendly tool for exploring and analyzing data stored in Snowflake, leveraging both internal and external resources to unlock new insights and possibilities.

Integration of Snowflake with Streamlit and OpenAI offers a powerful tool for data analysis, enabling users to:

- Access and analyze data stored in Snowflake with ease.
- Leverage the capabilities of OpenAI for advanced analytics.
- Visualize data and derive insights through an intuitive interface.

**Conclusion:**

This whitepaper has delved into the transformative integration of Snowflake's native capabilities with External Network Access, OpenAI, and Streamlit, showcasing a revolutionary approach to data analysis and interaction. By leveraging Snowflake's robust data management platform and the External Network Access feature, users can securely connect to external services, thereby broadening the horizons for data analysis.

The inclusion of Snowpark stored procedures further amplifies this capability, enabling the execution of complex data processing tasks within Snowflake, while harnessing the power of external resources like OpenAI for advanced analytics. This integration empowers users to extract valuable insights and actionable intelligence from their data with unprecedented ease and efficiency.

Furthermore, the development of a Streamlit native application built on Snowflake provides a user-friendly interface for querying, analyzing, and visualizing data, making data-driven insights accessible across organizations. This intuitive platform allows users of diverse backgrounds to seamlessly interact with their data, thereby facilitating informed decision-making and driving business innovation. By embracing these integrated technologies, organizations can unlock the full potential of their data, paving the way for growth and competitiveness in the dynamic landscape of data-driven decision-making.

**References:**

1. **Snowflake basics:** https://docs.snowflake.com/en/user-guide/tutorials/snowflake-in-20minutes
2. **Snowflake Native Apps**: https://docs.snowflake.com/en/developer-guide/native-apps/native-apps-workflow
3. **External Network Access:** https://docs.snowflake.com/en/developer-guide/external-network-access/external-network-access-overview
4. **Snowpark**: https://docs.snowflake.com/en/developer-guide/snowpark-container-services/overview
5. **Generative AI in Snowflake:** https://www.snowflake.com/trending/generative-ai/
6. **GenAI and LLM:** https://www.snowflake.com/blog/generative-ai-llms-summit-2023/