AskDB– Turning Questions into SQL and Insights

Project Report

Upendra Kumar | 28th Jan 2025

Table of Contents

1. Project Overview

AskDB is a GenAI-powered chatbot that enables business users to interact with databases using natural language.

It eliminates the need for SQL expertise by generating accurate SQL queries, retrieving data from AWS Athena, and providing clear summaries in plain English.

This project aims to bridge the gap between business users and technical data analytics tools, fostering greater data accessibility.

2. Problem Statement

1.Business users often lack technical skills to write SQL queries.

2.Dependence on technical teams causes delays in decision-making.

3.Existing tools are complex and not user-friendly for non-technical users.

4.A user-friendly system is needed to allow natural language interaction with databases, ensuring faster and more efficient insights.

3. Objectives

1.Enable natural language querying for database interaction.

2.Automate SQL query generation tailored to the right database and table.

3.Retrieve data from AWS Athena quickly and efficiently.

4.Provide data insights in simple, natural language for easy comprehension.

5.Develop a scalable and user-friendly system that can handle complex queries.

4. Design and Architecture

High-Level Architecture:

1. **Input Layer**: User submits queries via the chatbot interface.
2. **Processing Layer**:
    1. Natural Language Processing (NLP) converts user queries into SQL.
    2. Database and table mapping ensure correct SQL generation.
3. **Execution Layer**:
    1. SQL queries are executed on AWS Athena.
    2. Results are retrieved and processed.
4. **Output Layer**:
    1. Summarizes the retrieved data in plain English.
    2. Displays results to the user via the chatbot interface.

5.Technology Stack

1.Backend: Python
2.Database: AWS Athena
3.LLM: OpenAI GPT, AWS Bedrock
4.Cloud Services:
    1.AWS S3 (Data Storage)
    2.AWS Lambda (Serverless Functions)
    3.AWS Athena (Query Execution)
5.Libraries and Frameworks:
    1.Pandas
    2.SQLAlchemy
    3.LangChain
    4.Streamlit

6.Implementation Details

Query Understanding:

Leveraged NLP models to interpret user input and extract intent.

SQL Generation:

Mapped user intent to the correct database schema.

Automatically generated SQL queries using Python and SQLAlchemy.

Data Retrieval:

Executed SQL queries on AWS Athena.

Retrieved results in real-time.

Summarization:

Used a summarization model to generate natural language insights.

Deployment:

Hosted the system on AWS Infrastructure.

7.Challenges Faced

1. **NLP Accuracy**: Handling ambiguous or incomplete user queries.
2. **Database Mapping**: Ensuring correct schema and table selection.
3. **Performance**: Optimizing SQL queries for large datasets.
4. **Error Handling**: Addressing invalid queries or connection failures.
5. **Model Fine-Tuning**: Adapting generative models to handle domain-specific terms.

8.Lessons Learned

1.Effective use of GenAI can simplify complex technical tasks for non-technical users.
2.Importance of error handling to improve user experience.
3.Scalable cloud infrastructure is essential for real-time query execution.
4.NLP models need continuous training and fine-tuning to adapt to user-specific needs.

9.Future Scope

1. **Multi-Database Support**: Expand compatibility beyond AWS Athena to MySQL, PostgreSQL, etc.
2. **Data Visualization**: Add real-time charts and graphs to complement natural language summaries.
3. **Voice Integration**: Enable voice-based input for queries.
4. **Multi-Language Support**: Allow queries in regional and international languages.
5. **Advanced AI Models**: Use domain-specific fine-tuned models for better accuracy.

10.Conclusion

AskDB successfully bridges the gap between business users and databases by enabling natural language querying and simplifying data insights. This project demonstrates the potential of combining GenAI with cloud technologies to create user-friendly, scalable, and efficient solutions for data-driven decision-making.

With future enhancements, AskDB has the potential to become an **indispensable tool** for businesses worldwide.

THE END