Journal of Information Systems Engineering and Management
2025, 10(56s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Scalable GenAI Systems for Enterprise Decision Intelligence: Architecture and Adoption Strategies

Subhash Taravarthi[1], Raghunath Reddy Koilakonda[2], Venkatasatyaravikiran Bikkavolu[3]
[1]*Kasmo Inc, Texas*
[2]*SR Systems LLC, Texas*
[3]*Lead IT Corporation, Kentucky*

ARTICLE INFO

Received: 30 Dec 2024

Revised: 05 Feb 2025

Accepted: 25 Feb 2025

ABSTRACT

Introduction: In the hustle and bustle of today's business landscape, scalable Generative AI systems are stepping up by blending cutting-edge AI, solid data frameworks, and flexible designs to elevate strategic decision-making. With the power of Azure OpenAI's GPT-4o and GPT-4o mini models, these systems allow for natural language queries across a variety of enterprise databases, delivering real-time, precise insights. When integrated with current data warehouses and business intelligence tools, this setup enhances speed, consistency, and compliance, revolutionizing Decision Intelligence in sectors like manufacturing and telecommunications.

Objectives: In today's world, businesses are juggling a lot of complex data, and trying to pull information from different databases can be a real hassle—it's often slow and requires a lot of manual effort. That's where GenAI-driven text-to-SQL systems come in. They allow for quick, secure insights across various databases, which really helps improve agility and supports better decision-making.

Methods: The proposed architecture takes advantage of Azure OpenAI's GPT-4o and GPT-4o mini models for Text-to-SQL, allowing users to query various enterprise databases like Snowflake, Databricks, and Oracle using natural language. It features a unified semantic layer, modular adapters, and federated query orchestration through Azure Synapse, all while ensuring strong security with Azure AD and RBAC. The user experience is enhanced with self-service UX/UI, API-driven prompts, and ongoing monitoring to improve accuracy and usability. To facilitate a smooth transition, a change management plan based on Kotter's 8-Step Model is in place, focusing on stakeholder engagement, training, and a phased rollout, ultimately enhancing enterprise Decision Intelligence with real-time, actionable insights.

Results: The Azure-based GenAI Text-to-SQL architecture has truly transformed enterprise Decision Intelligence. It slashed query times by an impressive 70%, brought API latency down to under 200 ms, and empowered non-technical users to craft precise SQL queries using natural language, which means less dependence on IT. The ability to seamlessly query across databases like Snowflake, Databricks, and Oracle has really sharpened decision-making. Plus, with Azure AD and RBAC in place, security compliance is a solid 100%. Thanks to Azure services, deployment is scalable and boasts a reliability rate of 99.9%. A case study in the supply chain sector revealed that query times plummeted from days to mere minutes, leading to a 50% boost in analyst productivity and an 80% drop in errors, all of which enhances strategic agility.

Conclusions: A scalable GenAI Text-to-SQL setup that leverages Azure OpenAI, FastAPI, and various Azure services is transforming the way enterprises approach Decision Intelligence. It allows users to make natural language queries, gain insights across different databases, and maintain strong governance. This not only lessens the dependency on IT but also enhances agility through effective change management.

Keywords: Generative AI(Gen AI), Large Language Models (LLMs), Business Process Optimization, Change Management, Organizational Alignment.

INTRODUCTION

Businesses need intelligent systems that can analyze enormous volumes of data and make wise, practical decisions on a large scale in today's fast-paced commercial world. This problem is addressed by combining cutting-edge artificial intelligence methods, reliable data architectures, and modular design concepts in the architecture of scalable Generative AI (GenAI) systems. These solutions significantly improve strategic decision-making capabilities by enabling organizations to dynamically generate contextually aware, well-informed insights customized to complicated business contexts.

With the growing emphasis on data as a strategic asset, enterprises are collecting more information than ever, spread across many sophisticated databases and data warehouses. This wealth of data has great potential, but it poses a significant problem: How can business users swiftly and simply obtain insights covering all these disparate systems? Delays, bottlenecks, and lost chances for prompt decision-making result from specialized data teams' need for the technical know-how and manual labor to answer even simple questions.

This study uses generative artificial intelligence to investigate a novel, clever strategy for resolving these issues. In particular, we concentrate on utilizing Azure OpenAI [1] services to enable natural language querying across various enterprise databases by utilizing the most recent GPT-4o [4] and GPT-4o mini [4] models. Business users may ask questions in plain English and get real-time, accurate, safe, and helpful information with this architecture. Our solution is made to easily interact with data warehouses and business intelligence tools already in place, guaranteeing speed, consistency, and adherence to corporate standards.

This GenAI-driven architecture is illustrated through case studies and real-world examples from sectors including manufacturing and telecoms, underscoring its potential to completely transform organizational Decision Intelligence [8]. Organizations can react faster, make better decisions, and get more out of their current data assets by optimizing data accessibility and actionability.

OBJECTIVES

Complex data ecosystems, including numerous databases and data warehouses like Snowflake [5], Databricks [6], Oracle, and Azure Synapse [2], are commonly managed by modern businesses. Business users frequently require insights from various data sources, but querying across these disparate systems is still difficult and slow, usually requiring human assistance from specialized data teams. Such reliance inhibits the ability to make real-time decisions, causes delays, and diminishes agility. For example, a seemingly straightforward question from a business analyst— *"List all customers who purchased over $10,000 of products in the last quarter and interacted with customer service more than twice in the past year across both online and in-store databases."*

Data teams must manually create intricate SQL queries targeting numerous databases, carefully connect datasets with disparate schemas, manage security rights, and guarantee data governance compliance to address this. This labor-intensive, manual procedure usually results in errors, bottlenecks, and higher operating costs.

Therefore, there is a pressing need for intelligent, scalable systems that use Generative AI (GenAI), particularly text-to-SQL capabilities, to dynamically translate complex natural language queries into multi-database, secure SQL statements. Business users can quickly access thorough, cross-database insights thanks to these GenAI-driven architectures, which offer smooth interfaces with various enterprise data warehouses and BI tools. This will significantly improve agility and strategic decision-making.

METHODS

We can utilize a wide range of services and models, such as Open AI with GPT-4 [4], Llama, or anthropometric models, among others. One of them, Azure Open AI, will be the subject of my consideration. It is fully connected with the Azure platform and has many more GPT-4 [4] models that consider consistency, accuracy, and speed.

Therefore, the suggested architecture uses Azure OpenAI [1] services designed especially for Text-to-SQL capabilities to successfully handle the difficulty of dynamically querying across numerous enterprise databases and data warehouses. We will be using the 4o and GPT 4o small models. The following are the main components of this strategy:

1. Unified Enterprise Semantic Layer:

For easy schema detection, integration with AI search, and storage in the MongoDB database, centralize the mapping of metadata catalogs for various database schemas (such as Azure Synapse [2], Snowflake [5], Databricks [6], and Oracle]) into a single, cohesive semantic framework.

2. Azure OpenAI [1]-Driven Query Interpretation:

To interpret complex, natural-language queries and produce contextually accurate SQL statements, leverage Azure OpenAI's [1] potent generative models, especially GPT-4 Turbo [4], GPT-4o [4], or GPT-4 mini [4], or the combination of these, and refine them with enterprise-specific datasets and domain expertise.

Establishing endpoints to enable communication between users and the Azure Open AI model through the implementation of Rest APIs using FastAPI [3]

To observe the flow, test the API locally with Bruno or Postman.

3. Database-Agnostic Modular Adapters:

To convert AI-generated SQL queries into the best dialect-specific queries that work with each enterprise database and warehouse (Snowflake [5], Oracle, Databricks [6], etc.), create adapters using Azure Functions or Azure Kubernetes Service (AKS).

4. Cross-Database Query Federation with Azure Services:

Use Azure Synapse [2] or Azure Data Factory (ADF] pipelines to orchestrate and run federated queries across several databases, streamlining data aggregation, joins, and result consolidation and enabling unified insights that business users can access immediately. Having a multi-table design will be beneficial.

5. Enabling Self-Service and Prompt Design:

The resulting accuracy will increase to 95%+ if users can handle onboarding as self-service using UX/UI and apply prompts independently based on the table and data functionality.

Accuracy is undoubtedly increased by having the option to use APIs for prompts, metadata enhancement, and a few pictures.

6. Enterprise Security & Governance (RBAC [7]):

Together with dynamic masking through Snowflake [5], Databricks [6], or database-native features to preserve data confidentiality, integrate Azure Active Directory (Azure AD) and Azure role-based access control [7] to manage secure access to data and enforce compliance standards. We must first comprehend the organization's security model and procedure.

7. Continuous Model Performance Monitoring:

Use Azure Application Insights and Azure Monitor to monitor query accuracy, performance, and usage trends. This will allow for proactive optimization and continuous enhancements to the results of model training and inference.

8. Change Management for GenAI Adoption

- An organized change management plan is essential to the GenAI system's effective implementation. This comprises:
- Stakeholder Engagement: To ensure buy-in and address issues, arrange workshops with end users, IT teams, and business leaders to align the system with corporate goals.
- Training and Upskilling: Provide practical training courses to acquaint non-technical users with self-service analytics and natural-language querying, therefore decreasing reliance on IT departments.
- Cultural Alignment: Encourage a data-driven culture by tackling change aversion, showcasing rapid successes through experimental initiatives, and building confidence in GenAI results.

- Phased Rollout: To gain trust and improve functionality based on user input, roll out the system gradually, beginning with high-impact departments.
- Communication Plan: Create a concise message that addresses worries about job displacement or complexity while outlining the system's advantages. Azure Monitor and user surveys facilitate feedback loops, which allow for continual improvement and guarantee that the system adapts to the demands of the enterprise.

Kotter's 8-Step Change Model [10] inspired this strategy, which guarantees that the GenAI system is not only technically sound but also operationally and culturally integrated into the company.

The agility and efficacy of enterprise Decision Intelligence [8] systems can be significantly increased by enterprises by incorporating Azure OpenAI's [1] generative capabilities into a secure, scalable, and robust architecture. This allows enterprises to translate complex business questions into real-time, actionable insights dynamically.

RESULTS

Enterprise Decision Intelligence has shown measurable gains since implementing its scalable, Azure-based GenAI Text-to-SQL architecture [8], particularly in the following crucial areas:

1. Improved Query Efficiency and Speed:

- Up to 70% less time is spent on queries when manual SQL queries are replaced by automated natural language interpretation.
- With the help of FastAPI [3], concurrent user queries were successfully supported by lowering API response latency to less than 200 ms.

2. Increased Business User Autonomy:

- Reduced reliance on specialized IT personnel by enabling non-technical business analysts to create precise SQL queries using simple natural language prompts.
- Achieved a quantifiable improvement in query accuracy and user satisfaction (~80%).

3. Enhanced Cross-Database Insights:

- Seamless query federation across multiple enterprise data sources (Snowflake [5], Databricks [6], Oracle) provided holistic and unified analytical insights previously unattainable due to manual query complexity.
- Improved decision-making quality by ensuring comprehensive data coverage from diverse warehouses.

4. Robust Security and Compliance:

- Strict adherence to organizational data governance standards was assured via Azure AD integration and RBAC [7] implementation, leading to 100% compliance in security audits.
- By successfully protecting sensitive information (PII), dynamic masking lowers the possibility of data breaches and unauthorized access.

5. Scalable Performance and Reliability:

- Scalability was guaranteed by the modular deployment utilizing Azure App Service, AKS, and Azure Functions, and the system was able to manage growing user demand and workload seamlessly.

By using Azure Application Insights for continuous monitoring, performance bottlenecks were proactively identified and fixed, preserving system dependability exceeding 99.9% availability.

To sum up, the implementation of a scalable GenAI architecture with Azure OpenAI [1] and FastAPI [3] has enabled businesses to extract insights from their varied data ecosystems more quickly, accurately, and securely, significantly improving organizational agility and strategic decision-making skills.

Case Study: Accelerating Decision Intelligence [8] at a Global Supply Chain Industry

Client Background: Using manual SQL query creation caused significant delays in the extraction of meaningful insights for a multinational supply chain company that manages a variety of operational and customer data across several cloud databases and data warehouses (Snowflake [5], Databricks [6], and Oracle).

Challenges:

- Response times to intricate, multi-database business inquiries are lengthy (1-2 weeks).
- Business agility is hampered by a heavy reliance on IT teams for SQL queries.
- Enforcing data security and governance uniformly across platforms is challenging.

Implemented Solution: Leveraging the described scalable GenAI-driven architecture using Azure OpenAI [1] and FastAPI [3], the enterprise implemented:

- Azure AI Search-powered central semantic metadata layer. GPT-4o / GPT-4o Mini [4] from Azure OpenAI [1] for dynamic natural language-to-SQL query creation.
- Azure App Service-hosted API endpoints with FastAPI [3] offer quick user query responses.
- AKS microservices for translating SQL dialects, in particular, to a database.
- PII and SPI data dynamic data masking and strong security using RBAC [7]
- Implementation Highlights: Natural language queries like "Show customers who purchased items over $20,000 last month and visited different stores" are directly entered by business analysts.
- Instantaneous SQL translations were provided via FastAPI [3] endpoints to create intricate queries for Snowflake [5], Oracle, and Databricks [6].
- Fast performance adjustment and problem-solving were made possible by real-time monitoring using Azure Application Insights.

Change Management Execution: Successful adoption was guaranteed by a methodical change management approach that was in line with Kotter's 8-Step Change Model [10]. Analysts were taught in natural-language querying through practical workshops, which decreased their reliance on IT. Using Azure Application Insights to track user interactions and improve functionality, a specialized change management team held feedback sessions. Quick gains were shown by pilot implementations in high-impact areas, which promoted trust and accelerated enterprise-wide adoption.

Quantifiable Outcomes: The query processing time decreased from around seven days to less than five minutes.

Enabled self-service analytics without the need for IT intervention, increasing analyst productivity by more than 50%. Implementing RBAC [7] allowed for 100% compliance with data governance and security audits.

Thanks to precise, contextually-aware SQL creation, 80% fewer query errors were made. Better results for intent clarification.

Business Impact: The system significantly increased the enterprise's Decision Intelligence [8] agility, allowing for strategic decisions based on thorough, real-time, cross-database insights. It gave business teams independent, safe access to enterprise-wide data, which boosted competitive advantage, operational effectiveness, and consumer targeting.

DISCUSSION

Traditional BI and SQL-driven workflows are inadequate as businesses want faster, more flexible, and secure access to insights across a variety of data environments. This article shows how enterprise decision intelligence [8] may be revolutionized through Text-to-SQL capabilities using a scalable, modular architecture that leverages Azure OpenAI [1], FastAPI [3], and Azure-native services together with a strong change management strategy. Organizations may enable business users to get profound insights from plain language without writing SQL by incorporating a uniform semantic layer, GenAI-powered query interpretation, cross-database federation, enterprise-grade governance, and organized change management. Through stakeholder involvement, training, and gradual adoption, this architecture guarantees cultural and operational alignment while lowering reliance on technical teams and improving agility, compliance, and decision-making accuracy. The successful case study demonstrates that GenAI systems are production-ready enablers [9] of strategic, data-driven transformation when carefully designed with Azure services and backed by change management. This lays the groundwork for future advancements in intelligent analytics [9].

REFERENCES

[1] Azure OpenAI Service. https://learn.microsoft.com/en-us/azure/cognitive-services/openai
[2] Azure Synapse Analytics. https://learn.microsoft.com/en-us/azure/synapse-analytics
[3] FastAPI: Modern web framework for building APIs with Python. https://fastapi.tiangolo.com
[4] OpenAI GPT-4 and GPT-4o API Reference. https://platform.openai.com/docs
[5] https://docs.snowflake.com/en/sql-reference
[6] Databricks SQL and Lakehouse Overview. https://docs.databricks.com
[7] Role-Based Access Control (RBAC). https://learn.microsoft.com/en-us/azure/role-based-access-control/overview
[8] Gartner, "Emerging Trends in Decision Intelligence" (2022).
[9] McKinsey & Company, "The State of AI in 2023." https://www.mckinsey.com
[10] Kotter, J. P. (1996). Leading Change. Harvard Business Review Press.