Enhancing Business Analytics with LLM-SQL Integration

PROJECT SYNOPSIS

BACHELOR OF ENGINEERING
Computer Engineering

SUBMITTED BY

Neeti Kurulkar
Komal Ambhore
Tejal Dhoot
Soham Sarade
Shrey Doshi

Under the guidance of
Prof. Dr. Mrs. S. A. Itkar

Department of Computer Engineering
P. E. S. Modern College of Engineering,
Pune.

2025-2026

Contents

List of Figures

List of Tables

1 Title

Enhancing Business Analytics through LLM-SQL Integration

2 Domain

Generative AI

3 Keywords

Text-to-SQL, Large Language Models, Structured Data Reasoning, Schema Linking, Financial Analytics, Natural Language Interface, Few-Shot Prompting, Multi-Agent Systems

4 Team

Group Id: A5

Team Members:

1. Komal Ambhore - 31003
2. Tejal Dhoot - 31018
3. Shrey Doshi - 31020
4. Neeti Kurulkar - 31040
5. Soham Sarade - 31071

5 Literature Review

S.No	Paper Title	Authors (Journal, Year)	Key Contribution	Technique Used	Relevance to Project	Source
1	*MAG-SQL: Multi-Agent Generative NL2SQL with Soft Schema Linking*	Wenxuan Xie, Gaochen Wu, Bowen Zhou *arXiv preprint*, 2024	Multi-agent system for decomposing questions and refining SQL	Soft schema linking, question decomposition, sub-SQL refinement	Supports modular SQL for complex financial queries	[1]
2	*Few-shot Text-to-SQL with Intermediate Representation*	Luyao Chen, Qinyuan Ye, Lili Mou, Huan Sun *ACL*, 2023	Splits generation into structure and content stages for better accuracy	Few-shot prompting, intermediate representation	Matches your approach to few-shot modular SQL generation	[2]

S.No	Paper Title	Authors (Journal, Year)	Key Contribution	Technique Used	Relevance to Project	Source
3	*OpenSearch-SQL: Enhancing Text-to-SQL with Dynamic Few-shot and Consistency Alignment*	Xiangjin Xie, Guangwei Xu, Lingyan Zhao, Ruijie Guo *SIGMOD*, 2025	Dynamically selects best few-shot examples using relevance graphs	Graph-based few-shot selection	Optimizes prompt reuse across financial domains	[3]
4	*Is Long Context All You Need? Leveraging LLM's Extended Context for NL2SQL*	Siyao Qiu, Ming Ding, Yifan Xu et al. *arXiv preprint*, 2024	Evaluates prompt size, schema injection in long-context LLMs	Schema/value injection, column sampling	Informs prompt size and schema choice in system design	[4]
5	*RSL-SQL: Robust Schema Linking in Text-to-SQL Generation*	Zhenbiao Cao et al. *arXiv preprint*, 2024	Enhances schema linking with pruning and voting-based self-correction	Bidirectional linking, pruning, context filtering	Improves robustness for financial database schema linking	[5]
6	*Eliciting Reasoning in Language Models with Cognitive Tools*	Danqing Wang, Ximing Lu, Bill Lin, Xiang Ren *EMNLP*, 2023	Models structured reasoning via modular tool operations	Chain-of-thought, tool calling, cognitive flow	Supports structured SQL reasoning with intermediate steps	[6]

6 Objective

1. To automatically translate user queries into accurate and executable SQL statements
2. To extract insights from financial statements such as P&Ls, balance sheets, and KPIs
3. To eliminate the need for SQL or BI tool expertise among finance professionals.
4. To improve accessibility and decision-making by democratizing data access

7 Problem Statement

To develop a natural language interface that allows users to ask complex financial questions without SQL or BI expertise, by leveraging LLMs to translate those questions into executable SQL queries and returning insights conversationally.

8 Scope

1. Focus on financial reports: P&L, balance sheets, KPI tables
2. Natural language input mapped to structured SQL-compatible schemas
3. Readable insights generated using structured data reasoning.

9 Brief Description

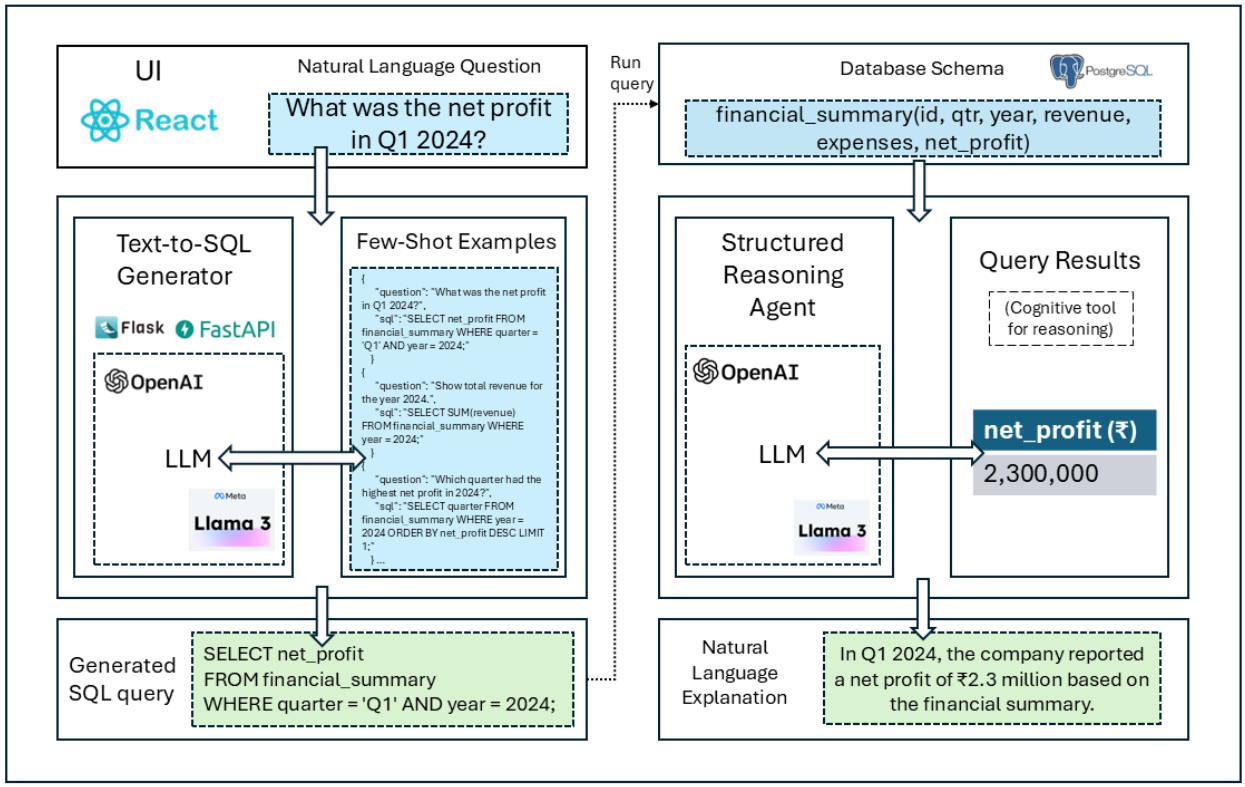

Figure 1: Architecture Diagram

1. Enables users to query financial data (like P&Ls and balance sheets) through a chat-based interface using natural language.
2. Uses Large Language Models to translate questions into SQL queries and returns human-readable insights.
3. Financial statements are parsed and stored in a structured database format for efficient querying.
4. Techniques like few-shot prompting and schema linking ensure accuracy, making the tool accessible to non-technical users.

10 Technical Details

1. **Input**
   (a) Financial statements (CSV, XLSX - structured data)
2. **Processing Pipeline**
   (a) Few-Shot Text-To-SQL Prompting
   (b) Schema Linking
   (c) Reasoning refinement with cognitive tools
3. **Output**
   (a) Results of SQL query and Natural Language explanation.
4. **Technologies**
   (a) **Backend**: PostgreSQL + Flask/FastAPI
   (b) **LLMs**: GPT-4 or open-source alternatives(Llama)
   (c) **Frontend**: React

11 Probable Date of Completion

April 2026

12 References

References

[1] W. Xie, G. Wu, and B. Zhou, "Mag-sql: Multi-agent generative approach with soft schema linking and iterative sub-sql refinement for text-to-sql," *arXiv preprint arXiv:2408.07930*, 2024.

[2] L. Chen, Q. Ye, L. Mou, and H. Sun, "Few-shot text-to-sql with intermediate representation," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL)*, 2023.

[3] X. Xie, G. Xu, L. Zhao, and R. Guo, "Opensearch-sql: Enhancing text-to-sql with dynamic few-shot and consistency alignment," *Proceedings of the ACM on Management of Data (SIGMOD)*, vol. 3, no. 3, pp. Article 194, 24 pages, 2025.

[4] S. Qiu, M. Ding, Y. Xu, Z. Wang, Z. Liu, and J. Tang, "Is long context all you need? leveraging llm's extended context for nl2sql," *arXiv preprint arXiv:2405.20695*, 2024.

[5] Z. Cao, Y. Zheng, Z. Fan, X. Zhang, W. Chen, and X. Bai, "Rsl-sql: Robust schema linking in text-to-sql generation," *arXiv preprint arXiv:2411.00073*, 2024.

[6] D. Wang, X. Lu, B. Y. Lin, and X. Ren, "Eliciting reasoning in language models with cognitive tools," in *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2023.