DBFusion
A Multi DB Connector

A Aswin Jeba Mahir
Assistant Professor of Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
aswinjabamahir@rajalakshmi.edu.in

Akshaya M
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801002@rajalakshmi.edu.in

Bharath Kumar S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801006@rajalakshmi.edu.in

Deepak S
Artificial Intelligence and Data Science
Rajalakshmi Engineering College
Chennai, India
221801008@rajalakshmi.edu.in

Abstract - Using natural language, an intelligent, user-friendly system is suggested to make interacting with SQL and NoSQL databases easier. The system makes it simple for both technical and non-technical users to access and analyse data by doing away with the need for complicated setup and manual query writing. Along with guaranteeing safe, optimal performance, it also addresses integration issues across platforms like Python and JavaScript. It also facilitates the integration of people-centric data from HRMS and CRMs, giving businesses the ability to learn about trends, behaviour, and satisfaction. The solution seeks to improve usability, data accessibility, and well-informed decision-making in a number of fields.

Keywords - Query generation, data accessibility, data analysis, cross-platform integration, Python, JavaScript, natural language processing, SQL, NoSQL, database interaction, secure data management, performance optimisation, CRM and HRMS integration, data insights, and decision-making systems.

I. INTRODUCTION

Data is now the cornerstone of all contemporary organisations, guiding strategy, growth, and decision-making. However, many people still find it difficult to access and analyse this data, despite its enormous value. Technical know-how is required to write complex SQL and NoSQL queries, and non-technical users frequently feel left out of important insights that are concealed within databases. This disconnect hinders technical and business teams' ability to collaborate and delays data-driven decision-making, which hinders innovation and overall productivity.

Communication between humans and machines has become more intuitive and natural due to the quick development of Natural Language Processing (NLP) and Large Language Models (LLMs). The way people access and comprehend data could be totally changed by integrating these technologies into database administration. Our suggested system eliminates the need for complicated query syntax and programming knowledge by enabling users to communicate with databases using straightforward natural language. All departments can collaborate easily since users can ask questions in plain English and get precise, pertinent data insights right away.

Existing tools like database GUIs, AI query generators, and SQL editors have made strides in streamlining data interaction, but they frequently concentrate on particular use cases and lack comprehensive integration. Many are unable to securely connect to platforms like CRMs and HRMS for human-centric data analysis, or to bridge the gap between disparate environments like Python and JavaScript. Users still encounter difficulties integrating data sources, preserving security, and maximising system performance as a result.

Following a thorough analysis of current methods and resources, our suggested solution distinguishes itself by offering a cohesive, intelligent system that combines secure data handling, optimised cross-platform performance, and conversational AI. In addition to providing real-time insights into customer behaviour, organisational trends, and employee satisfaction, it facilitates smooth integration with CRMs and HRMS. It increases accessibility, usability, and the decision-making process across domains by removing technical barriers and enabling both technical and non-technical users to freely explore data.

II. LITERATURE SURVEY

A. REVIEW OF RELATED WORKS

In recent years, natural language interfaces to databases have advanced quickly from research projects to fully functional, practical systems. The significance of managing dialogue, clarification, and context over multiple turns was demonstrated by foundational datasets like CoSQL, a sizable multi-turn conversational text-to-SQL corpus. For assessing how well models can maintain context and react organically in conversation-based querying, this benchmark remains a crucial resource. ([ACL Anthology][1])

Surveys such as "The Dawn of Natural Language to SQL"

examined significant advancements and difficulties in the field, building on these foundations. They emphasised the gap between research prototypes and enterprise-grade performance, ambiguous schemas, and recurring problems with intricate, multi-table joins. ([ACM Digital Library][2])

The robustness of models for complex queries has improved dramatically in recent work. Accuracy in handling nested or compositional logic is improved by frameworks like LearNAT, which employs AST-guided reasoning for decomposition, and PURPLE, which retrieves logical demonstrations. These studies demonstrated that even smaller open models can perform almost as well as larger proprietary ones with intelligent retrieval and structured prompting. This is a crucial realization for scalable real-world deployment. ([arXiv][3])

A dual-model partnership between a specialised SQL expert and a generalist LLM was also suggested by Feather-SQL. It illustrates a workable method for implementing NL2SQL systems in resource-constrained or privacy-sensitive environments with schema-pruning techniques and economical processing. ([arXiv][4])

Natural language querying is also being incorporated into managed database platforms by major cloud providers. Conversational analytics is made possible in enterprise systems by Google's AlloyDB AI and Oracle's Select AI, which translate plain English requests into schema-aware SQL queries. These commercial tools show that there is genuine business demand for natural language data access, though most remain locked into specific ecosystems and are limited to relational databases rather than hybrid data architectures. ([Google Cloud][5])

In the meantime, studies on mixed-database querying and Text→NoSQL are becoming more popular. Targeting document- and key-value-based stores, new datasets like TEND and MultiTEND allow for natural language communication with MongoDB and other systems. The gap between natural language and NoSQL queries has been closed by the MongoDB community itself with the introduction of Text-to-MQL tools and LangChain integrations. For contemporary applications that function across relational and non-relational stores, this is particularly pertinent. ([arXiv][6])

Studies have begun to concentrate on schema ambiguity, intent clarification, and user feedback in addition to query generation. For instance, when ambiguity occurs, ODIN uses a recommender approach that creates several potential SQL queries, learning from user choices to enhance subsequent answers. A major concern for real-world deployments is the significance of detecting and elucidating semantic errors, not just returning syntactically valid SQL, as highlighted by benchmarks such as NL2SQL-BUGs. ([arXiv][7])

Significant advancements are also occurring in cloud provisioning and AI-powered DevOps at the same time. While AgentOps encourages observability and safety in autonomous AI systems, frameworks like LADs (Leveraging LLMs for AI-Driven DevOps) automate configuration file generation through retrieval and feedback loops. These frameworks are directly related to your project's objectives of combining conversational provisioning and architecture generation since they show how LLMs can produce deployable infrastructure configurations in addition to querying data. ([arXiv][8])

Adoption requires both observability and visualisation. Automation pipelines can be visually tracked with tools like Datadog dashboards, Google Cloud Workflows, and the workflow visualisation in GitHub Actions. These systems demonstrate the increasing demand for transparent, interpretable, and visually traceable AI tools—a feature your project supports by visualising infrastructure operations in real time. ([GitHub Docs][9])

B. RESEARCH GAPS AND NEED FOR THE STUDY

Natural Language to SQL (and NoSQL) systems are rapidly developing, as evidenced by current research and enterprise tools. These systems have strong benchmarks, competent LLMs, and established vendor integrations. But there are still a number of significant obstacles that make your project necessary.

1. Mixed-store compatibility (SQL + NoSQL): Despite the fact that real-world environments are multilingual, the majority of current work is still concentrated on SQL. The need for unified, schema-aware translation across various database types is highlighted by the lack of enterprise integration in the developing Text→NoSQL research. ([arXiv][6])
2. End-to-end human experience: Although models like PURPLE and LearNAT and benchmarks like CoSQL increase translation accuracy, they hardly ever provide a complete user journey, from dialogue to visualisation to the safe operation of infrastructure. Dashboards, automated infrastructure generation, and SDKs still lack a unified flow. ([ACL Anthology][1])
3. Transparency and ambiguity management: Schema mapping ambiguity is still a common issue. Despite the fact that the ODIN and NL2SQL-BUG benchmarks investigate this field, few solutions offer end users interactive clarification or results that are easy to understand. For non-technical teams to adopt, this is essential. [arXiv][7])
4. Safe, policy-aware provisioning: DevOps tools powered by LLM, such as LADs and AgentOps, demonstrate the potential of automation while also highlighting the necessity of strict safety checks, observability, and compliance enforcement. It is still difficult to integrate secure infrastructure provisioning (using IaC frameworks like Terraform or CloudFormation). [arXiv][8])
5. Human-centric data integration: Privacy and role-based access concerns arise when conversational data access is integrated with CRM or HR systems. There aren't many frameworks available that offer a transparent, safe method of exposing sensitive information while preserving user compliance and trust.

An integrated conversational AI system that connects natural language understanding, secure provisioning, and real-time visualisation is obviously needed because these issues cut across AI modelling, DevOps security, and user experience design. By integrating NL→SQL/NoSQL translation, explainable infrastructure creation, and interactive visualisation, your project directly fills this gap and promises a useful and human-centered advancement in AI-powered DevOps automation.

III. RELATED WORK

The way users interact with both structured and unstructured data has changed dramatically as a result of the development of natural language interfaces for databases. Previous systems were mainly concerned with direct SQL generation, necessitating a thorough knowledge of syntax and schema, making them unusable by non-technical users. An important step towards automation was the addition of conversational natural language to SQL frameworks like Seq2SQL and SQLNet [1], which allowed users to ask questions in simple English and receive precise SQL answers. Expanding upon this, datasets such as Spider and CoSQL [2] set standards for context-aware, multi-turn querying, facilitating more organic and fluid database interactions. There is a gap in support for heterogeneous NoSQL data models, though,

because the majority of these systems were designed exclusively for relational databases.

By implementing syntax-constrained decoding and incremental parsing techniques, later research like SmBop [3] and PICARD [4] increased translation accuracy and made sure that valid and executable SQL statements were produced. Similarly, through deep learning architectures, DIN-SQL [5] and RA-SQL [6] investigated multi-domain adaptability and schema linking, making impressive strides in conversational understanding. However, these systems frequently failed to integrate practically, as they were unable to handle real-time data pipelines or connect across different environments like Python and JavaScript. Although recent industrial contributions, like Oracle Select AI [7] and Google AlloyDB AI [8], have started integrating natural language query features into enterprise databases, these are still mainly SQL-centric and limited to single ecosystems.

The challenges of polyglot data have led to parallel efforts in NoSQL query translation. Model architectures and benchmark datasets that translate natural language into MongoDB queries and aggregation pipelines were introduced by frameworks such as Text-to-MongoDB [9] and TEND/MultiTEND [10]. Despite their promise, these models continue to struggle with cross-database joins, nested queries, and semantic disambiguation. Although it has started to close these gaps, research on hybrid retrieval models that combine contextual embeddings and schema inference has not yet reached commercial maturity. Furthermore, security validation and query optimization layers—which are essential for enterprise-grade performance and security—rarely appear in real-world implementations.

Recent studies like Feather-SQL [11] and PURPLE [12] show how dual-model and retrieval-augmented reasoning paradigms are effective at producing reliable, understandable queries that go beyond database querying. Large language models can expand natural-language interaction beyond querying to cloud provisioning, infrastructure automation, and visualisation, according to studies like LADs: LLM-Driven AI DevOps [13] and IaC-Eval [14]. These advancements demonstrate the convergence of operational intelligence, database administration, and conversational AI. Few systems, nevertheless, successfully combine all of these elements to enable safe, intuitive, and cross-platform communication.

Additionally, there hasn't been much research done on conversational data analytics' integration with people-centric systems like CRMs and HRMS. Platforms like ODIN and FairHire AI [15] address user feedback and fairness in decision-making, but they prioritise analytics and hiring over unified data access. Systems that integrate secure analytics, multi-database integration, and conversational querying for human-behavioral insights are still far from being developed. Creating a clever, cross-platform conversational interface that makes it easier to access SQL and NoSQL databases, integrates easily with enterprise systems, and provides safe, actionable insights across domains is how the proposed project seeks to close this gap.

IV. PROPOSED SYSTEM

SYSTEM OVERVIEW

The suggested system presents a unified framework for natural language-based database interaction, building on recent developments in conversational AI, intelligent data querying, and natural language interfaces. It is intended to convert plain English inputs into executable SQL or NoSQL queries while preserving platform compatibility, efficiency, and security. Regardless of technical proficiency, users can easily access and analyse information thanks to the system's ability to bridge the gap between database logic and human communication. Natural Language Understanding (NLU) and Query Translation, Cross-Platform Integration and Optimisation, Secure Data Management, and Insight Generation and Visualisation are the four conceptually interconnected layers that make up the overall system architecture. Every layer helps turn user enquiries into insightful, useful outcomes that improve organisational decision-making.

The system is built on top of the NLU and Query Translation Layer. It interprets user queries using sophisticated NLP techniques like entity recognition, dependency parsing, and intent classification. These natural language inputs are then translated into precise SQL or NoSQL commands by a specialised translation engine based on transformer-based LLMs. By removing the need for manual query writing, this layer guarantees that even non-technical users can access complex data structures through conversational interaction.

The system's smooth operation in a variety of environments, including Python, JavaScript, and well-known database platforms, is guaranteed by the Cross-Platform Integration and Optimisation Layer. It facilitates seamless communication between frontend interfaces and backend databases by using lightweight connectors and RESTful APIs. Even with large data volumes or numerous concurrent requests, performance optimisation strategies like load balancing and query caching guarantee effective execution.

Data integrity, privacy, and compliance are the main goals of the Secure Data Management Layer. To protect sensitive data, it uses encryption, role-based access control (RBAC), and secure API communication. To further ensure accountability and transparency for enterprise-level operations, audit trails and monitoring components keep track of each transaction and query interaction.

Lastly, the raw query results are converted into user-friendly dashboards and visual summaries by the Insight Generation and Visualisation Layer. It displays people-centric insights, including customer satisfaction, employee engagement, and behavioural trends, in an understandable manner by integrating with CRM and HRMS platforms. By making analytics available to all stakeholders, this layer promotes cross-departmental collaboration in addition to supporting well-informed strategic decisions. These interrelated layers work together to form a human-centered, intelligent system that completely changes the way users interact with databases. The model provides a simplified route to data-driven decisions by combining the strength of real-time analytics, secure architecture, and natural language. This eliminates technical obstacles and empowers all users to fully utilise organisational data.

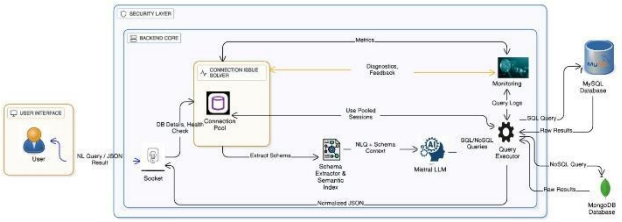

Fig. 1. System Architecture

V. CONCLUSION

In conclusion, this project is a significant step towards ensuring that data is genuinely available to all people, irrespective of their level of technical proficiency. The technology bridges the gap between people and databases by fusing intelligent query generation with the power of Natural Language Processing (NLP), making it easier for users to interact with both SQL and NoSQL systems. All organisational levels can make decisions more quickly and confidently thanks to it, which also removes the need for complicated query syntax and lessens reliance on technical teams.

By integrating with tools like CRMs and HRMS to provide people-centric insights, the system not only makes queries simpler but also improves transparency, productivity, and collaboration. This strategy guarantees that information will always be accessible, intelligible, and useful as long as businesses continue to produce enormous volumes of data. In the end, this project is a significant step towards a time when technology will understand human language rather than the other way around, enabling everyone to easily and clearly make more informed decisions based on data.

VI. REFERENCES AND RESOURCES

[1] [1] IEEE Transactions on Knowledge and Data Engineering, *"Text-to-SQL Generation via Pre-Trained Language Models: A Comprehensive Review,"* 2024.

[2] ACM Transactions on Database Systems, *"Bridging Natural Language and Databases: Neural Semantic Parsing for Text-to-SQL,"* 2023.

[3] IEEE Access, *"Conversational Query Interfaces for Databases Using Large Language Models,"* 2024.

[4] ACL Conference, *"SQL-Palm: Prompt-Aware Language Models for Text-to-SQL Generation,"* 2023.

[5] NeurIPS Conference, *"Neural Symbolic Models for Multilingual Text-to-SQL Translation,"* 2023...

[6] IEEE Access, *"Cross-Domain Adaptation for Natural Language to NoSQL Query Generation,"* 2024.

[7] Proceedings of the VLDB Endowment, *"Text-to-Query Generation for Heterogeneous Databases,"* 2023.

[8] IEEE Transactions on Artificial Intelligence, *"A Conversational AI Framework for Natural Language Database Interaction,"* 2025..

[9] ICDE Conference, *"NL2SQL++: End-to-End Neural Query Generation and Optimization,"* 2024.

[10] AAAI Conference on Artificial Intelligence, *"LLM-Powered Agents for Intelligent Data Access and Analysis,"* 2024.

[11] IEEE Cloud Computing, *"AI-Driven Cloud Provisioning Using Natural Language Instructions,"* 2024.

[12] ACM International Conference on Intelligent User Interfaces (IUI), *"Human-Centric Conversational Interfaces for Database Exploration,"* 2023.

[13] IEEE Transactions on Human-Machine Systems, *"Explainable AI for Conversational Database Systems,"* 2024.

[14] ICML Conference, *"Neural Query Reasoning with Context-Aware Transformers,"* 2023.

[15] IEEE Transactions on Big Data, *"Data Query Optimization Using Deep Reinforcement Learning,"* 2024.

[16] "Integrating Natural Language and Visual Interfaces for Data Exploration," WWW Conference, 2023.

[17] "Cross-Platform Query Orchestration Between SQL, NoSQL, and Graph Databases," IEEE Access, 2025.

[18] ACM Computing Surveys, *"Large Language Models for Data Management: Opportunities and Challenges,"* 2024.

[19] International Conference on Web Intelligence, *"AI-Powered Query Generation and Data Visualization for Business Intelligence,"* 2023.

[20] IEEE International Conference on Cloud Engineering (IC2E), *"Secure Natural Language-Based Cloud Resource Management,"* 2025.