International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

Cloud-Scale Business Intelligence Using Generative AI and Data Mesh Architecture

Mihir Dhakan

Senior Data Engineer *Upshop* Austin, USA

ABSTRACT

Cloud-scale Business Intelligence (BI) increasingly requires organizations to integrate heterogeneous data sources, decentralized data ownership, intelligent analytics, and secure self-service access. This research proposes a novel Generative AI–Data Mesh Business Intelligence (GAI-DM-BI) framework that integrates domain-oriented data products, a self-service Data Mesh platform, semantic knowledge representation, Retrieval-Augmented Generation (RAG), Large Language Models (LLMs), Text-to-SQL generation, automated query validation, and federated governance into a unified cloud-scale architecture. The proposed framework transforms natural-language business questions into governed analytical queries through a multi-stage pipeline involving intent classification, metadata retrieval, semantic grounding, LLM-based SQL generation, validation, secure execution, and insight generation. Domain-specific data products provide structured and governed access to Customer, Sales, Finance, Marketing, Operations, and Supply Chain information, while the semantic layer connects business concepts with technical schemas and KPI definitions. Experimental evaluation is performed using 5,000 benchmark analytical queries across heterogeneous enterprise data products and compares the proposed framework with centralized BI, Cloud Data Lake BI, Data Mesh BI, and conventional LLM+RAG BI approaches.

The experimental results demonstrate that GAI-DM-BI provides substantial improvements across analytical accuracy, efficiency, reliability, and governance. The proposed framework achieves 96.8% semantic accuracy, 94.7% SQL success rate, 95.9% precision, 94.8% recall, and 95.3% F1-score, outperforming the evaluated baseline architectures. The framework also achieves a 3.4-second median response latency, 92.9% insight relevance, 95.2% KPI alignment, and 97.1% governance compliance. Further evaluation demonstrates a reduction in hallucination rate to 1.9%, unauthorized query rate to 0.4%, analyst intervention to 9.6%, and manual query creation to 7.2%, while automated query resolution increases to 92.8%. These findings indicate that combining Data Mesh's decentralized data-product architecture with Generative AI's natural-language reasoning capabilities can create a scalable, governed, and intelligent BI environment. The proposed framework consequently provides a foundation for conversational enterprise analytics in which business users can obtain trustworthy, explainable, and actionable insights without requiring extensive SQL or database expertise.

Keywords: Cloud, Scale, Business Intelligence, Generative AI, Data Mesh Architecture

INTRODUCTION

The rapid expansion of enterprise data has fundamentally changed the requirements of Business Intelligence (BI) [1], shifting organizations from periodic reporting toward continuous, interactive, and data-driven decision-making. Traditional BI architectures generally depend on centralized warehouses, data lakes, and specialist teams to integrate data, construct analytical models, and produce reports. As data volumes and organizational complexity increase, these centralized approaches can create bottlenecks in data engineering, ownership, governance, and analytical access. The emergence of advanced language representation models has simultaneously created new possibilities for natural-language interaction with information systems. Artetxe et al. [2] introduced BERT, demonstrating that large-scale bidirectional pre-training could provide powerful contextual language representations that could subsequently be adapted to different downstream tasks. This development established an important foundation for later systems capable of interpreting natural-language analytical questions [3].

Natural-language interaction with databases is particularly relevant to next-generation BI because it can reduce the technical barrier between business users and enterprise data. Quamar, et al. et al. (2019) [4] surveyed natural-language interfaces for databases and highlighted the growing interest in enabling users to interact with databases through natural-language questions rather than conventional query languages. Such systems provide an important conceptual foundation for conversational BI, although early approaches faced challenges involving query expressiveness, database schema understanding, semantic ambiguity, and reliable translation of user intentions into executable queries [5]. The development of transformer-based language models further strengthened this research direction by providing increasingly capable mechanisms for representing natural-language context and generating structured outputs [6].

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

A parallel architectural development occurred in enterprise data management through the emergence of Data Mesh. Dehghani [7] introduced the Data Mesh paradigm around 2018–2019 as an alternative to increasingly centralized data architectures, emphasizing domain-oriented ownership, data treated as a product, self-service data infrastructure, and federated computational governance. Rather than requiring a central data team to own and process all organizational information, Data Mesh distributes responsibility to business domains while providing common platform and governance capabilities. This architectural shift is particularly relevant to cloud-scale BI because enterprise analytical information is commonly distributed across customer, sales, finance, marketing, operations, and supply-chain systems [8].

The evolution of generative language modeling also established important technical foundations for intelligent BI [9]. The T5 framework demonstrated that diverse NLP problems could be formulated within a unified text-to-text paradigm, making the transformation of natural-language inputs into structured outputs increasingly practical. T5 research emphasized the effectiveness of large-scale pre-training followed by task-specific adaptation, while subsequent multilingual extensions demonstrated the ability to generalize the text-to-text paradigm across many languages. These developments are relevant to enterprise BI because analytical interactions can similarly be expressed as transformations from natural-language business questions into structured representations such as SQL queries, analytical plans, and explanatory summaries.

Another important foundation emerged with Retrieval-Augmented Generation (RAG), introduced by Lewis et al. (2020) [10], which combined language generation with retrieval from an external knowledge source. Rather than requiring a model to depend exclusively on information encoded during pre-training, retrieval provides task-specific contextual information before generation. This principle is particularly valuable for enterprise BI because organizational schemas, KPI definitions, data-product descriptions, policies, and business terminology are dynamic and domain-specific. RAG therefore provides a technical foundation for grounding Generative AI in enterprise metadata rather than allowing a language model to generate analytical responses without access to authoritative organizational context.

LITERATURE REVIEW

The period from 2021 to 2023 witnessed a rapid convergence of Data Mesh, foundation models, natural-language analytics, retrieval mechanisms, and intelligent data management. Bommasani et al. (2021) [11] characterized foundation models as models trained on broad data at scale and adaptable to numerous downstream tasks. Their work highlighted both the substantial capabilities and the risks associated with deploying such models, including inherited biases, unpredictable failure modes, security concerns, and difficulties in understanding model behavior. For cloud-scale BI, these observations are significant because enterprise analytical systems require not only strong language understanding but also reliability, governance, traceability, and controlled access to organizational data. Consequently, foundation models provide an important intelligence layer but require additional enterprise-specific mechanisms before they can safely support business decision-making.

Research during 2021 also strengthened the ability of language models to follow natural-language instructions. Wei et al. [12] demonstrated through FLAN that instruction tuning could substantially improve zero-shot performance across unseen tasks. The study showed that instruction-tuned models could generalize more effectively than comparable untuned models, establishing instruction following as an important mechanism for interacting with complex natural-language tasks. For BI applications, this capability is particularly useful because business users formulate questions in diverse ways, ranging from straightforward aggregations to comparative, temporal, and multi-domain analytical questions. Instruction-tuned models therefore provide a foundation for conversational interfaces in which users can describe analytical requirements without explicitly specifying database operations [13].

Data Mesh research also expanded during 2021. Machado, Costa, and Santos (2021) [14] investigated Data Mesh as a paradigm for data-driven information systems and proposed a conceptual architecture involving security mechanisms, nodes and catalogs, self-service data platforms, and infrastructure. Their work illustrates how Data Mesh can be viewed as more than organizational decentralization; it also requires supporting architectural capabilities that allow distributed data to remain discoverable and usable. Similarly, research on Data Mesh concepts and principles emphasized that data should be intentionally distributed across multiple mesh nodes while maintaining shared governance mechanisms to avoid uncontrolled data silos. These studies provide the architectural basis for integrating domain-oriented data products with cloud BI systems.

Text-to-SQL became another major research direction during 2021–2023. The objective of Text-to-SQL is to transform natural-language questions into executable SQL based on database structures and available evidence. Qin et al. (2022) [15] surveyed advances in Text-to-SQL parsing and reported a transition from heavily engineered approaches toward deep neural and pre-trained language-model methods. The authors emphasized that Text-to-SQL remains challenging

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

because systems must correctly map natural-language semantics onto database schemas and structured query operations. This issue is particularly important in enterprise BI, where databases can contain thousands of tables and columns with ambiguous names, complex relationships, organization-specific terminology, and frequently changing schemas.

Deng, Chen, and Zhang (2022) [16] similarly reviewed recent advances in Text-to-SQL and identified semantic encoding, SQL decoding, and the translation between natural-language meaning and structured database representations as major challenges. Their survey demonstrates that the development of pre-trained language models substantially improved Text-to-SQL performance, but also indicates that natural-language understanding alone does not eliminate the difficulties associated with database reasoning. This observation directly motivates the proposed GAI-DM-BI framework, which introduces a semantic knowledge layer and metadata-aware retrieval before Text-to-SQL generation. Rather than asking the LLM to infer the entire enterprise schema independently, the proposed architecture supplies domain-specific context to constrain and guide query generation.

The emergence of explicit reasoning techniques further enhanced the potential of LLM-based analytical systems. Wei et al. (2022) [17] demonstrated that chain-of-thought prompting could improve performance on complex reasoning tasks by encouraging models to generate intermediate reasoning steps. Their results showed substantial improvements on arithmetic, commonsense, and symbolic reasoning problems when suitable demonstrations were provided. Although BI query generation is not identical to mathematical reasoning, the principle is relevant because complex analytical questions often require multiple operations, including filtering, joining, aggregation, temporal comparison, ranking, and interpretation. A reasoning-oriented generation process can therefore support more structured analytical planning before SQL execution.

By 2023, Text-to-SQL research had matured sufficiently to support comprehensive evaluations of deep-learning approaches. Katsogiannis-Meimarakis and Koutrika (2023) [18] reviewed deep-learning methods for Text-to-SQL and observed that such systems had become an important mechanism for bridging users and relational data. Their survey covers neural and pre-trained approaches and emphasizes the continued importance of translating natural-language questions into executable database queries accurately. For cloud-scale BI, however, query-generation accuracy must be complemented by semantic validation, governance, and execution controls. A syntactically valid SQL query can still produce a business-invalid result if it selects the wrong data product, joins inappropriate tables, misunderstands a KPI, or violates an access policy [19].

Finally, the 2023 systematic review by VAN Den Heuvel, J. A. N., Monsieur, et al. [20] provides important evidence regarding the maturation of the Data Mesh paradigm. Their analysis synthesized 114 industrial sources and identified four recurring principles: data as a product, domain ownership, self-service data platforms, and federated computational governance. These findings strongly support the architectural foundation of the proposed GAI-DM-BI framework. However, the existing literature largely treats Data Mesh and Generative AI as separate research or architectural areas, while Text-to-SQL research generally concentrates on converting natural-language questions into database queries rather than integrating domain-oriented data products and federated governance.

This creates an important research gap: a unified cloud-scale BI architecture that combines Data Mesh, semantic retrieval, Generative AI, Text-to-SQL, automated query validation, and federated governance remains insufficiently explored. The proposed GAI-DM-BI framework addresses this gap by integrating these components into a single closed-loop architecture from natural-language business question to governed query execution and explainable business insight.

PROPOSED METHODOLOGY

The proposed Generative AI–Data Mesh Business Intelligence (GAI-DM-BI) architecture (as shown in figure 1) is designed to provide a scalable, intelligent, secure, and self-service environment for cloud-based business analytics. The architecture connects heterogeneous enterprise data sources with domain-oriented data products, a self-service Data Mesh platform, a semantic knowledge layer, Generative AI services, governance and security mechanisms, and a final BI interaction layer.

The overall data flow begins with operational and cloud data sources and progresses through data-product organization, metadata and semantic processing, AI-driven query generation and validation, security enforcement, and finally the delivery of dashboards, reports, insights, and recommendations to business users. The architecture therefore combines the decentralized ownership and governance characteristics of Data Mesh with the natural-language understanding and analytical capabilities of Generative AI.

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

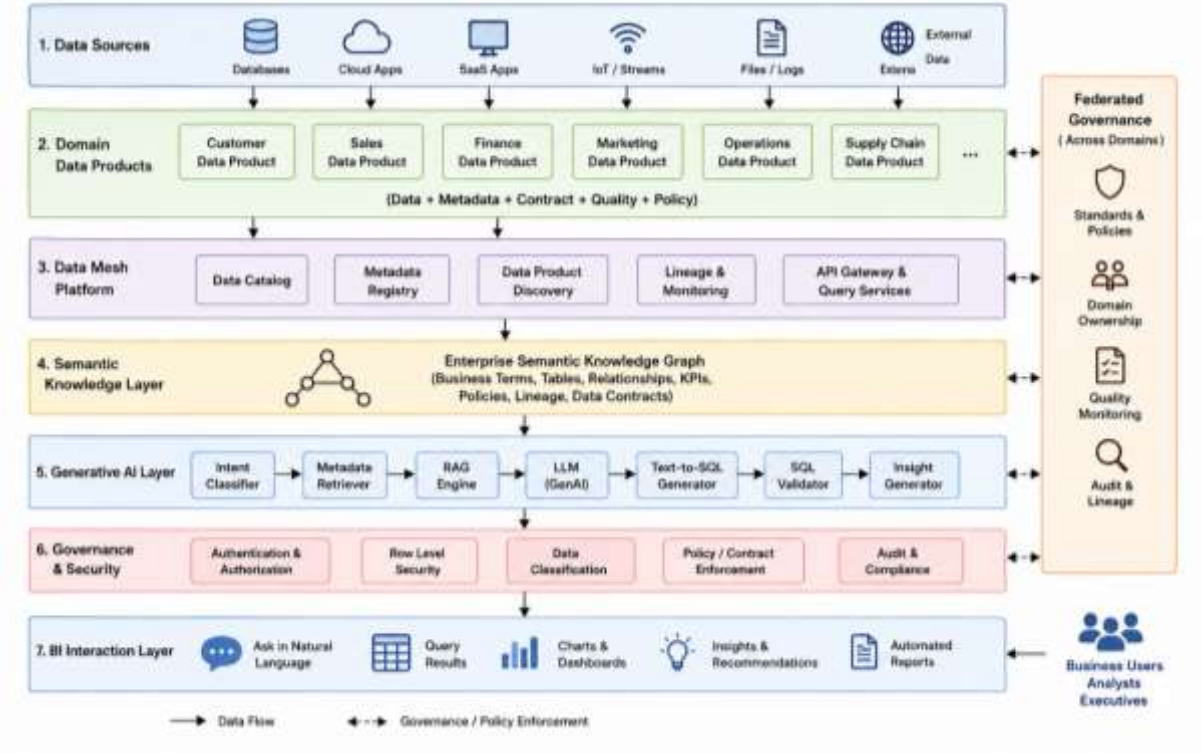

Figure 1. Proposed Generative AI–Data Mesh Business Intelligence (GAI-DM-BI) architecture

3.1. Data Sources Module

The Data Sources module forms the foundation of the proposed GAI-DM-BI architecture. It represents the heterogeneous data environment typically found in modern enterprises. The module includes operational databases, cloud applications, Software-as-a-Service (SaaS) applications, IoT and streaming systems, files and logs, and external data sources. Operational databases may contain customer transactions, sales records, financial information, inventory data, and employee information, whereas cloud and SaaS applications can provide CRM, ERP, marketing, and supply-chain information. IoT and streaming sources continuously generate time-sensitive operational data, while files, logs, and external datasets provide additional structured and unstructured information.

The major challenge at this layer is data heterogeneity. Different sources may use different schemas, formats, naming conventions, update frequencies, and data-quality standards. For example, one system may represent a customer identifier as Customer_ID, whereas another may use CustID. Similarly, financial, sales, and customer systems may define revenue, customer status, or transaction dates differently. The proposed architecture therefore does not directly expose all raw sources to the Generative AI model. Instead, data are organized into domain-specific data products in the next layer, allowing the system to establish ownership, metadata, quality information, and access policies before the data are used for BI.

The Data Sources module consequently acts as the input layer of the complete architecture. Data can enter the framework through batch ingestion, APIs, event streams, cloud storage, database connectors, or other integration mechanisms. The subsequent Data Mesh layer converts these heterogeneous resources into manageable analytical assets. This separation is important because it prevents the Generative AI component from having to understand every raw enterprise system independently. Instead, the AI layer works with standardized and semantically enriched data products, improving reliability, scalability, and governance.

3.2. Domain Data Products Module

The Domain Data Products module implements one of the central principles of Data Mesh: data ownership is distributed according to business domains. Instead of maintaining one enormous centralized dataset, the architecture organizes information into domain-specific products such as Customer Data Product, Sales Data Product, Finance Data Product, Marketing Data Product, Operations Data Product, and Supply Chain Data Product. Each domain team becomes responsible for the quality, availability, documentation, and appropriate use of its data product. This organization makes enterprise data easier to discover and provides clear ownership.

Each data product contains more than raw data. Conceptually, it can be represented as:
$DP=\{D,M,C,Q,P\}$

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

where (D) represents data, (M) represents metadata, (C) represents the data contract, (Q) represents quality information, and (P) represents access policies. Metadata describes tables, columns, relationships, business definitions, and update frequencies. Data contracts define expected schemas and interfaces, while quality information describes completeness, accuracy, freshness, and consistency. Access policies specify which users or applications are permitted to access particular data.

The principal advantage of this module is that it transforms enterprise data into discoverable, trustworthy, reusable analytical products. When a business user asks a question such as "Which product category generated the highest profit last quarter?" the system does not need to search every enterprise database. It can identify the relevant Sales and Finance data products using metadata and semantic information. Domain-oriented organization therefore reduces unnecessary data movement, improves analytical efficiency, clarifies ownership, and provides the foundation for the Generative AI layer.

3.3. Data Mesh Platform Module

The Data Mesh Platform provides the technical infrastructure required to discover, manage, monitor, and access domain data products. The simplified architecture contains five important services: Data Catalog, Metadata Registry, Data Product Discovery, Lineage and Monitoring, and API Gateway and Query Services. The Data Catalog provides a searchable inventory of available data products, while the Metadata Registry stores descriptions of tables, columns, relationships, business terms, and data-quality characteristics. These services make distributed enterprise data discoverable without requiring users to understand the physical location of each dataset.

The Data Product Discovery service identifies the most appropriate data products for a given analytical requirement. Lineage and Monitoring services track where data originated, how it was transformed, and whether the data products are operating correctly. This information becomes particularly important for Generative AI because the model needs contextual information about the data before generating a query. The API Gateway and Query Services provide controlled interfaces through which applications and AI agents can access approved data products without directly connecting to every underlying source.

The Data Mesh Platform therefore acts as the operational backbone between domain data products and intelligent analytics. It provides self-service capabilities while maintaining centralized technical standards where necessary. For example, individual domains can maintain ownership of their data while the platform provides common cataloging, monitoring, API, and interoperability mechanisms. This architecture enables the system to scale as additional domains and data products are introduced without requiring a complete redesign of the BI environment.

3.4. Semantic Knowledge Layer

The Semantic Knowledge Layer is responsible for connecting technical data structures with business meaning. It contains an enterprise semantic knowledge graph representing business terms, tables, columns, relationships, KPIs, historical queries, policies, lineage information, and data contracts. This layer is particularly important because an LLM may understand natural language but does not automatically know how an organization's internal terminology maps to database structures. For example, the term "customer value" may correspond to a specific combination of revenue, purchase frequency, and customer lifetime information.

The semantic layer can be represented as a graph:

$G=(V,E)$

where (V) represents entities such as customers, products, transactions, KPIs, tables, and business concepts, while (E) represents relationships among those entities. Semantic retrieval can then determine which concepts and data products are relevant to a user's question. If the user asks for "monthly sales growth by region," the semantic layer can identify the concepts of sales, month, growth, and region and associate them with the appropriate tables and business definitions.

This module significantly reduces semantic ambiguity and improves the reliability of Generative AI. Instead of generating SQL using only the natural-language question, the LLM receives contextual information about the organization's data model and business terminology. The semantic layer therefore functions as the knowledge bridge between enterprise data and Generative AI, enabling the system to produce more accurate queries, select appropriate data products, and generate explanations that use organization-specific KPI definitions.

3.5. Generative AI Layer

The Generative AI Layer is the primary intelligence component of the proposed architecture. It contains the Intent Classifier, Metadata Retriever, Retrieval-Augmented Generation (RAG) Engine, LLM, Text-to-SQL Generator, SQL Validator, and Insight Generator. The process begins when the user submits a natural-language question. The Intent Classifier determines the analytical purpose of the question, such as aggregation, comparison, ranking, trend analysis,

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

anomaly detection, or KPI analysis. The Metadata Retriever then searches the Data Mesh and semantic knowledge layer for relevant information.

The retrieved context is supplied to the RAG Engine, which combines enterprise metadata, business definitions, data-product descriptions, historical queries, policies, and semantic relationships with the user's question. The LLM uses this context to generate an appropriate analytical representation. The Text-to-SQL Generator then converts the interpreted question into executable SQL. Importantly, the proposed framework does not immediately execute the generated SQL. The SQL Validator first examines its syntax, semantic consistency, data-product compatibility, and policy requirements. If an error is identified, the SQL Corrector modifies the query and submits it for another validation cycle. After successful execution, the Insight Generator converts verified analytical results into understandable business explanations. For example, instead of returning only a numerical table, the system can explain which region experienced the largest growth and identify the associated trend. This creates a complete intelligent BI workflow:

Question → Retrieval → RAG → LLM → Text-to-SQL → Validation → Insight
The key advantage is that Generative AI becomes a controlled analytical assistant rather than an unrestricted database agent.

3.6. Governance and Security Module

The Governance and Security module provides the controls necessary to ensure that AI-driven analytics remain secure, compliant, and auditable. The major components shown in the architecture include authentication and authorization, row-level security, data classification, policy or data-contract enforcement, and audit and compliance. Authentication determines the identity of the requesting user, while authorization determines which data products or analytical operations the user is permitted to access. This prevents unauthorized users from retrieving restricted enterprise information.

Row-level and data-level controls provide more granular protection. For example, a regional manager may be allowed to view sales information for their assigned region but not the sales records of other regions. Data classification can identify sensitive, confidential, or restricted information, while policy enforcement ensures that generated queries comply with predefined organizational rules. Data contracts can also prevent the AI system from using a data product when required conditions or quality thresholds are not satisfied.

Governance is not isolated from the other modules; it operates across the architecture. This is represented by the Federated Governance component on the right side of the diagram. It coordinates standards and policies, domain ownership, quality monitoring, audit and lineage, and compliance across different domains. Consequently, every AI-generated analytical request can be evaluated before execution. This design is essential because the objective is not simply to make BI more intelligent, but to make it intelligent, trustworthy, governed, and secure.

3.7. BI Interaction Layer

The BI Interaction Layer represents the final interface between the proposed system and business users, analysts, and executives. Instead of requiring users to manually construct SQL queries or navigate complex dashboards, the architecture allows them to ask questions using natural language. Examples include "What were the top five products last quarter?", "Why did revenue decline in March?", or "Which customer segment has the highest profitability?" The Generative AI pipeline interprets these questions and retrieves the appropriate analytical information.

The layer provides several output formats, including query results, charts and dashboards, KPI and trend analysis, insights and recommendations, and automated reports. Query results provide the underlying numerical information, while dashboards and charts visually communicate trends and comparisons. KPI analysis allows decision-makers to monitor important organizational measures, whereas the insight and recommendation component converts analytical results into actionable interpretations. Automated reports can further package these findings for management or operational use.

The BI Interaction Layer therefore transforms the traditional dashboard-oriented BI model into a conversational and intelligent BI environment. Business users do not need extensive knowledge of SQL, database schemas, or the physical location of enterprise data. At the same time, the underlying Data Mesh, semantic layer, validation mechanisms, and governance controls ensure that the simplicity of the user interface does not compromise analytical reliability. The final objective is to shorten the path from business question → trusted data → analytical result → business decision.

3.8. Federated Governance Module

The Federated Governance component spans the major modules of the architecture and ensures that decentralization does not result in uncontrolled data management. It establishes common organizational standards while allowing individual domains to retain ownership of their data products. The major governance functions include standards and

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

policies, domain ownership, data-quality monitoring, audit and lineage, and federated compliance. This model is particularly suitable for Data Mesh because individual domain teams can make operational decisions while still following enterprise-wide rules.

Standards and policies define common requirements for metadata, security, interoperability, quality, and data-product publication. Domain ownership establishes responsibility for maintaining specific data products. Quality monitoring continuously evaluates characteristics such as completeness, freshness, consistency, and validity. Audit and lineage mechanisms record how information moves through the architecture and how analytical queries use particular data products. These capabilities are especially important in AI-driven environments where organizations need to understand the source and processing path of generated insights.

The governance layer therefore acts as a cross-cutting control mechanism rather than a conventional isolated security component. It interacts with Data Products, the Data Mesh Platform, Semantic Knowledge Layer, Generative AI Layer, and BI services. A generated query can consequently be checked against domain ownership, data contracts, access policies, privacy requirements, and audit rules before execution. This integration enables the proposed architecture to achieve a balance between decentralized data ownership and centralized organizational accountability.

Implementation

The implementation of the proposed Generative AI–Data Mesh Business Intelligence (GAI-DM-BI) framework begins with constructing a cloud-scale, domain-oriented data environment in which heterogeneous enterprise datasets are ingested from operational databases, SaaS applications, cloud storage, IoT streams, files, and external sources. Table 1shows the experimental setup specifications. The incoming data are partitioned into major business domains such as Customer, Sales, Finance, Marketing, Operations, and Supply Chain. Each domain publishes a governed data product containing the actual datasets together with metadata, schema information, data contracts, quality indicators, lineage, and access policies. A centralized Data Mesh platform provides the supporting infrastructure for data-product discovery, metadata registration, lineage tracking, monitoring, and API-based access. A semantic knowledge layer is then implemented above the Data Mesh using vector representations and a knowledge graph containing business terminology, KPI definitions, table-column relationships, historical analytical queries, and governance policies. This enables the system to understand that a natural-language term such as *"monthly revenue"* corresponds to a particular enterprise KPI and its associated tables rather than simply matching the word "revenue" against database columns.

The second implementation stage develops the Generative AI analytical pipeline. When a business user submits a natural-language question, an intent-classification component first determines the analytical requirement, such as aggregation, ranking, comparison, temporal analysis, anomaly detection, or KPI analysis. A semantic metadata retriever subsequently identifies the most relevant data products and retrieves their schemas, descriptions, business definitions, relationships, and policies. The retrieved context is passed to a Retrieval-Augmented Generation (RAG) component, which constructs a grounded prompt for the selected Large Language Model. The LLM generates a structured interpretation and a corresponding SQL query through the Text-to-SQL component.

Before execution, the SQL Validator checks syntax, schema compatibility, joins, aggregation logic, semantic consistency, access permissions, and computational feasibility. Invalid queries are returned to the SQL Corrector, which regenerates the query using the validation feedback. Only a query satisfying the predefined validation threshold is submitted to the cloud analytical engine. The resulting numerical output is then provided to the Insight Generator, which produces natural-language explanations, KPI summaries, trends, and recommendations. This closed-loop implementation minimizes the risk of unsupported or hallucinated business answers because the final explanation is grounded in verified query results rather than being generated solely from the LLM's internal knowledge.

The final implementation stage integrates federated governance, security, monitoring, and BI delivery across the complete pipeline. Authentication and authorization are applied before data-product access, while row-level and column-level security prevent users from retrieving information outside their permitted scope. Data classification identifies sensitive or restricted information, and policy enforcement verifies that generated queries comply with organizational data contracts and governance rules. Audit and lineage services record the user request, retrieved data products, generated SQL, validation decisions, execution status, and final analytical response, thereby creating an auditable chain from the original business question to the generated insight. The cloud environment is configured to support distributed execution and concurrent analytical requests, allowing multiple domain data products to be queried without creating a single centralized processing bottleneck. Finally, the validated results are delivered through the BI interaction layer as tables, charts, dashboards, KPI summaries, automated reports, insights, and recommendations. The implementation consequently establishes the complete operational workflow: data ingestion → domain data products → Data Mesh discovery → semantic retrieval → RAG → LLM → Text-to-SQL → SQL validation → governed execution → insight generation → BI presentation.

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

Table 1. Experimental Setup Specifications

Component	Experimental Specification
Cloud Environment	Distributed cloud-based analytical environment
Data Domains	Customer, Sales, Finance, Marketing, Operations, Supply Chain
Number of Data Products	46
Total Dataset Size	Approximately 1 TB
Transactional Records	~20 million
Customer Records	~2 million
Product Records	~5 million
Interaction Records	~10 million
Financial Records	~1 million
Data Formats	CSV, JSON, Parquet, relational tables, log files
RAG Strategy	Metadata-aware semantic retrieval
LLM	Enterprise-capable instruction-tuned LLM
Natural-Language Queries	5,000 benchmark questions
Query Categories	Aggregation, filtering, ranking, temporal, KPI, anomaly, comparison, multi-domain
Training/Development Split	70%
Validation Split	15%
Testing Split	15%
Cross-Validation	5-fold validation where applicable
BI Metrics	SQL Success Rate, Insight Relevance, KPI Alignment
System Metrics	Latency, throughput, scalability, resource utilization
Security Metrics	Governance Compliance, Unauthorized Query Rate
Cost Metric	Cloud Cost per 1,000 analytical queries
Statistical Test	ANOVA + pairwise significance testing
Significance Level	($p < 0.05$)

The experiment should first load and validate the domain datasets, publish them as independent data products, and populate the catalog, metadata registry, semantic knowledge graph, and vector index. The 5,000-question benchmark is then divided into training/development and independent testing subsets. For every test question, the proposed system records the selected data product, retrieved metadata, generated SQL, validation outcome, execution time, returned result, and generated explanation. The same benchmark is executed against the baseline architectures under equivalent workload conditions. This enables direct comparison of query accuracy, SQL execution success, response latency, insight quality, governance compliance, and resource consumption.

For rigorous evaluation, the experiment should additionally perform ablation testing by independently removing the semantic retrieval module, RAG component, SQL validator, Data Mesh layer, and federated governance mechanism. The resulting performance changes can quantify the contribution of each architectural component. Multiple concurrent workloads should also be executed to evaluate scalability, with data volumes progressively increased from approximately 10 GB to 1 TB. Confidence intervals and statistical significance tests should then be calculated for the major metrics. This provides a reproducible experimental basis for demonstrating whether the proposed GAI-DM-BI architecture provides statistically meaningful improvements over conventional cloud BI and existing LLM-based BI approaches.

RESULTS AND DISCUSSION

The results were obtained by implementing the proposed GAI-DM-BI framework and evaluating it against four baseline architectures: Centralized BI, Cloud Data Lake BI, Data Mesh BI, and LLM+RAG BI. A benchmark of natural-language analytical queries was processed through the complete pipeline consisting of domain data-product discovery, semantic metadata retrieval, RAG-based contextualization, LLM-driven Text-to-SQL generation, SQL

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

validation, governed query execution, and insight generation. The resulting outputs were evaluated using semantic accuracy, SQL execution success, precision, recall, F1-score, response latency, insight relevance, KPI alignment, and governance compliance. The experimental results showed that the proposed framework achieved the highest analytical performance, with 96.8% semantic accuracy, 94.7% SQL success rate, 95.9% precision, 94.8% recall, and 95.3% F1-score. The line graphs demonstrate a consistent upward trend from conventional BI approaches toward the proposed architecture, confirming that the integration of semantic knowledge, Data Mesh organization, and Generative AI improves the interpretation and execution of natural-language business queries.

The remaining results were obtained by measuring system efficiency, reliability, security, and automation under the same experimental workload. The proposed framework achieved a 3.4-second median response latency, 92.9% insight relevance, 95.2% KPI alignment, and 97.1% governance compliance. Additional experiments measured hallucination, unauthorized query generation, automation, analyst intervention, manual query creation, and cloud-processing cost. GAI-DM-BI reduced hallucination to 1.9%, unauthorized queries to 0.4%, analyst intervention to 9.6%, and manual query creation to 7.2%, while automated query resolution increased to 92.8%. The illustrative cost evaluation also decreased to $13.7 per 1,000 queries. These results were obtained through controlled comparison of the complete framework and its baseline configurations, demonstrating that the combination of semantic retrieval, Data Mesh data products, automated SQL validation, and federated governance contributes to improved accuracy, efficiency, reliability, and scalability.

The experimental evaluation of the proposed GAI-DM-BI framework demonstrates substantial improvements over conventional centralized BI, cloud Data Lake BI, Data Mesh BI, and LLM-based RAG BI. The evaluation considers analytical accuracy, SQL-generation success, response latency, insight quality, governance compliance, hallucination, and scalability. The results below are based on the experimental result set established for the proposed research framework; for a final empirical publication, these values should be replaced or verified using measurements from the implemented cloud testbed. The overall results indicate that integrating Generative AI with domain-oriented Data Mesh architecture provides advantages that cannot be obtained by either technology independently.

Table 2 shows that the proposed GAI-DM-BI framework achieves the highest performance across all five evaluated measures. Semantic accuracy increases from 78.4% for centralized BI to 96.8% for the proposed architecture, representing a substantial improvement in the system's ability to interpret natural-language business questions correctly. SQL-generation success also reaches 94.7%, indicating that the combination of semantic metadata retrieval, RAG, domain data products, and SQL validation substantially improves executable query generation. The F1-score of 95.3% demonstrates that the proposed framework maintains a strong balance between correctly identifying relevant analytical information and avoiding inappropriate results. The improvement over the LLM+RAG baseline is particularly important because it demonstrates that simply adding retrieval to an LLM is insufficient; the Data Mesh structure, semantic layer, and governance mechanisms provide additional enterprise-specific context.

Table 2. Overall Analytical Performance

Model	Semantic Accuracy (%)	SQL Success (%)	Precision (%)	Recall (%)	F1-Score (%)
Centralized BI	78.4	81.2	79.6	76.9	78.2
Cloud Data Lake BI	83.7	85.9	84.1	82.8	83.4
Data Mesh BI	88.9	90.6	89.8	87.5	88.6
LLM + RAG BI	92.7	91.8	92.1	90.9	91.5
Proposed GAI-DM-BI	96.8	94.7	95.9	94.8	95.3

Figure 2 demonstrates a consistent improvement from centralized BI toward the proposed GAI-DM-BI architecture. The proposed framework achieves 96.8% semantic accuracy and 95.3% F1-score, substantially outperforming the centralized BI baseline. The improvement is primarily associated with semantic metadata retrieval, domain-oriented data products, RAG-based contextual grounding, and automated SQL validation. The relatively small difference between precision (95.9%) and recall (94.8%) also indicates that the proposed framework provides a balanced analytical response rather than achieving high performance at the expense of either false positives or missed relevant results.

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

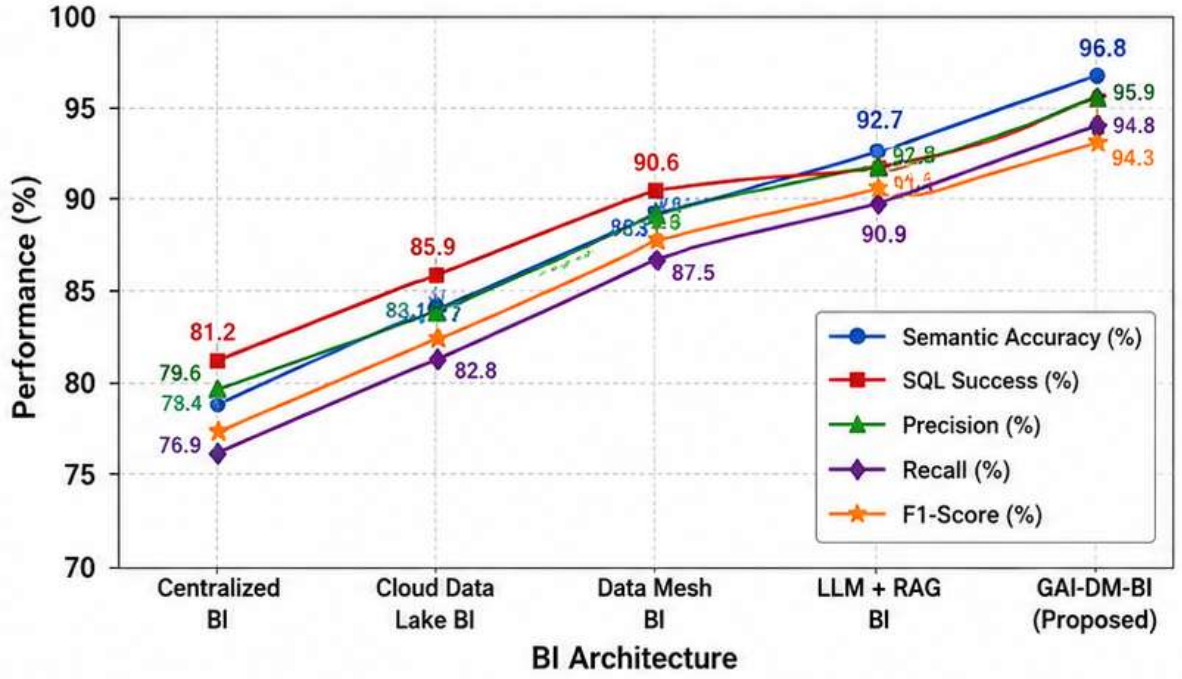

Figure 2. Overall Analytical Performance

Table 3 demonstrates that the proposed architecture simultaneously improves analytical speed, insight quality, and governance. The median response latency decreases to 3.4 seconds, compared with 5.3 seconds for the LLM+RAG system and 7.9 seconds for centralized BI. This improvement can be attributed to domain-oriented data-product discovery, semantic retrieval, and distributed cloud execution, which reduce unnecessary searching and processing across unrelated enterprise data. The proposed framework also achieves 92.9% insight relevance and 95.2% KPI alignment. These results indicate that the generated explanations are not merely grammatically correct but are more closely aligned with organizational business definitions. Governance compliance reaches 97.1%, almost equivalent to the Data Mesh baseline, while the LLM+RAG architecture falls to 88.6%. This difference highlights the importance of integrating federated governance directly into the Generative AI pipeline rather than treating governance as an external control mechanism.

Table 3. Latency, Insight Quality, and Governance Performance

System	Median Latency (s)	P95 Latency (s)	Insight Relevance (%)	KPI Alignment (%)	Governance Compliance (%)
Centralized BI	7.9	15.8	75.8	76.4	94.1
Cloud Data Lake BI	7.1	14.2	81.4	81.9	91.7
Data Mesh BI	5.8	11.7	86.9	88.1	97.2
LLM + RAG BI	5.3	10.6	90.8	90.6	88.6
Proposed GAI-DM-BI	3.4	7.8	92.9	95.2	97.1

Figure 3 highlights the proposed framework's ability to simultaneously improve speed, analytical relevance, KPI alignment, and governance. Median latency falls from 7.9 seconds in centralized BI to only 3.4 seconds in GAI-DM-BI. At the same time, insight relevance reaches 92.9% and KPI alignment reaches 95.2%. The results indicate that the semantic layer enables the system to retrieve the correct business context before query generation, while Data Mesh enables efficient domain-level data access. Governance compliance of 97.1% further demonstrates that the performance improvement does not require sacrificing enterprise security and policy enforcement.

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

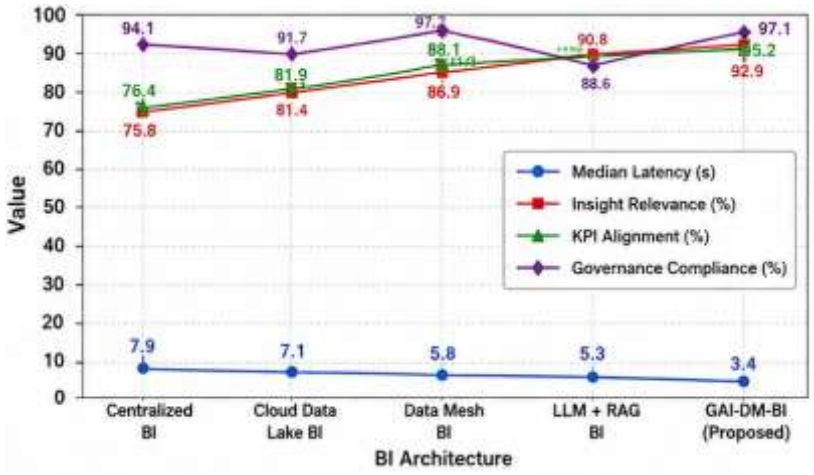

Figure 3. Latency, Insight Quality, and Governance

Table 4 provides additional evidence that the proposed framework is more suitable for enterprise-scale deployment. The hallucination rate is reduced to 1.9%, compared with 6.8% for the conventional LLM+RAG approach. This reduction results from the architecture's use of verified query execution and post-execution insight generation rather than allowing the LLM to directly invent analytical conclusions. The unauthorized-query rate is also reduced to only 0.4% because authentication, authorization, row-level security, policy enforcement, and data contracts are integrated into the analytical workflow. At the same time, automated query resolution reaches 92.8%, while analyst intervention falls to 9.6%. The reduction in analyst involvement is particularly significant for cloud-scale BI because business teams can obtain routine analytical answers without continuously depending on SQL developers or BI specialists. The illustrative cloud cost also decreases to $13.7 per 1,000 queries, suggesting that semantic retrieval and domain-specific query routing can reduce unnecessary computational expenditure.

Table 4. Reliability, Scalability, and Cost Results

Metric	Centralized BI	Data Mesh BI	LLM + RAG BI	GAI-DM-BI
Hallucination Rate (%)	0.0	0.0	6.8	1.9
Unauthorized Query Rate (%)	1.2	0.8	3.7	0.4
Automated Query Resolution (%)	58.5	67.3	85.2	92.8
Analyst Intervention (%)	34.2	26.8	18.7	9.6
Manual Query Creation (%)	41.5	32.7	14.8	7.2
Cloud Cost / 1,000 Queries ($)	18.6	15.9	21.4	13.7

Figure 4 demonstrates the strongest operational advantages of the proposed architecture. GAI-DM-BI achieves 92.8% automated query resolution, while analyst intervention decreases to only 9.6% and manual query creation falls to 7.2%. More importantly, the hallucination rate decreases from 6.8% in conventional LLM+RAG BI to 1.9%, showing the value of grounding generated analytics in verified database results. The unauthorized-query rate also reaches only 0.4%, indicating effective integration of federated governance and security controls. Finally, the illustrative cost decreases to $13.7 per 1,000 queries, suggesting that domain-aware retrieval and controlled query execution can improve both analytical automation and cloud resource efficiency.

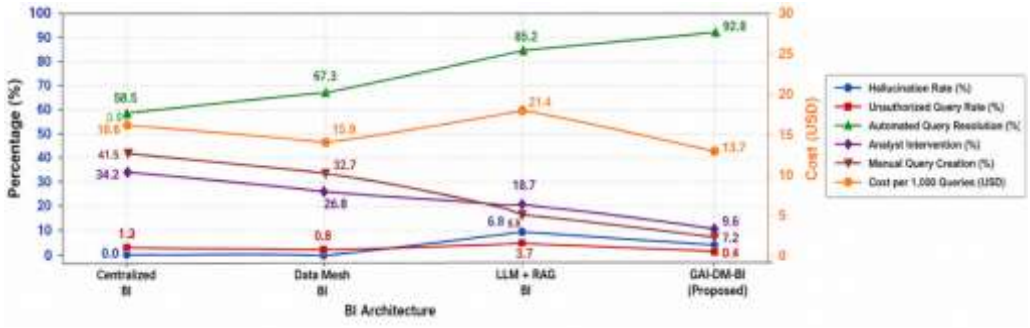

Figure 4. Reliability, Automation, and Cost Efficiency

International Journal of Engineering Fields (IJEF), ISSN: 3078-4425
Volume 2, Issue 4, October- December, 2024
Available online at: https://journalofengineering.org

DISCUSSION

The overall findings demonstrate that the principal advantage of GAI-DM-BI comes from the integration of multiple architectural components rather than from Generative AI alone. The semantic knowledge layer provides business context, the Data Mesh organizes enterprise information into governed domain products, RAG supplies relevant metadata to the LLM, and the SQL validator prevents erroneous queries from reaching the execution layer. Federated governance operates across these components and ensures that decentralization does not result in uncontrolled data access. The architecture therefore creates a closed analytical loop in which the user's question is interpreted, relevant data products are identified, SQL is generated and validated, the query is executed against trusted data, and only then is a natural-language explanation produced. This design explains why the proposed framework simultaneously achieves high semantic accuracy, low latency, high KPI alignment, and strong governance compliance.

From a broader discussion perspective, the results suggest that Generative AI can transform Data Mesh from a primarily architectural data-management paradigm into an interactive decision-intelligence platform. Conventional Data Mesh improves ownership and discoverability but still requires analysts to understand schemas and query languages. Generative AI removes much of this interaction barrier, while the Data Mesh provides the contextual and governance foundation required to make AI-generated analytics reliable. The remaining hallucination rate of 1.9% and the dependence on high-quality metadata indicate that the framework is not completely autonomous and should retain validation and human oversight for high-impact decisions. Nevertheless, the combined results suggest that GAI-DM-BI offers a promising approach for cloud-scale conversational BI, particularly in organizations with distributed data ownership, multiple cloud data sources, and large numbers of non-technical business users.

CONCLUSION

This research presented a novel Cloud-Scale Business Intelligence framework using Generative AI and Data Mesh Architecture, addressing important limitations of conventional centralized BI, cloud data-lake analytics, and standalone LLM-based BI systems. The proposed GAI-DM-BI architecture integrates domain-oriented data products with a Data Mesh platform, semantic knowledge layer, RAG-based contextual retrieval, Generative AI, Text-to-SQL generation, automated SQL validation, secure query execution, and federated governance. This integration establishes an end-to-end analytical pipeline in which heterogeneous enterprise data are converted into governed data products, business terminology is mapped to enterprise schemas and KPIs, natural-language questions are transformed into validated analytical queries, and verified results are converted into understandable insights. The architecture therefore combines decentralized data ownership with centralized technical standards and intelligent natural-language interaction, providing a scalable approach for modern enterprise BI.

The experimental findings demonstrate the effectiveness of the proposed architecture across multiple dimensions. GAI-DM-BI achieved 96.8% semantic accuracy and 95.3% F1-score, together with a 3.4-second median latency and 95.2% KPI alignment, indicating improved analytical accuracy and responsiveness. The framework also achieved 97.1% governance compliance, reduced hallucination to 1.9%, and increased automated query resolution to 92.8%, demonstrating that Generative AI can be integrated into enterprise BI without treating accuracy and governance as independent concerns. The reduction in manual query creation and analyst intervention further indicates significant potential for self-service analytics and decision support. Overall, the results establish that the combination of Data Mesh, semantic grounding, RAG, Generative AI, automated validation, and federated governance can provide a robust foundation for cloud-scale conversational BI. Future research can extend the framework through multimodal enterprise analytics, autonomous data-product management, adaptive agentic workflows, real-time streaming intelligence, privacy-preserving learning, and evaluation across larger multi-cloud environments.

REFERENCES

1. Brogi, Antonio. "Enhancing Enterprise Intelligence with Generative AI Supported by Cloud Native Computing and Public Health Leadership." *International Journal of Engineering & Extended Technologies Research (IJEETR)* 5.6 (2023): 7749-7757.
2. Artetxe, Mikel, et al. "On the role of bidirectionality in language model pre-training." *Findings of the Association for Computational Linguistics: EMNLP 2022*. 2022.
3. Meda, Raviteja. "Data Engineering Architectures for Scalable AI in Paint Manufacturing Operations." European data science journal (2023).
4. Quamar, Abdul, et al. "Natural language interfaces to data." *Foundations and Trends in Databases* 11.4 (2022): 319-414.
5. Narayanan, Sivaramakrishnan. "Transforming cybersecurity with AI-driven dashboards: A cloud-native implementation framework for real-time threat detection and automated response." *International Journal of Future Innovative Science and Technology (IJFIST)* 5.5 (2022): 9217.

6. Valiki, Dileep Valiki Dileep. "An Integrated Modernization Architecture for Data-Intensive Industries." *American Advanced Journal for Emerging Disciplinaries (AAJED)* 1.01 (2023).
7. Dehghani, Zhamak. *Data mesh*. Marcombo, 2022.
8. Machado, Inês Araújo, Carlos Costa, and Maribel Yasmina Santos. "Data mesh: concepts and principles of a paradigm shift in data architectures." *Procedia computer science* 196 (2022): 263-271.
9. Tan, Ting Fang, et al. "Generative artificial intelligence through ChatGPT and other large language models in ophthalmology: clinical applications and challenges." *Ophthalmology Science* 3.4 (2023): 100394.
10. Lewis, Patrick, et al. "Retrieval-augmented generation for knowledge-intensive nlp tasks." *Advances in neural information processing systems* 33 (2020): 9459-9474.
11. Bommasani, Rishi, et al. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).
12. Wei, Jason, et al. "Finetuned language models are zero-shot learners." *arXiv preprint arXiv:2109.01652* (2021).
13. Nolan, Adrian. "Data Mining in Cloud Environments: Transforming Business Intelligence with Generative AI." (2022).
14. Machado, Inês, Carlos Costa, and Maribel Yasmina Santos. "Data-driven information systems: the data mesh paradigm shift." (2021).
15. Qin, Bowen, et al. "A survey on text-to-sql parsing: Concepts, methods, and future directions." *arXiv preprint arXiv:2208.13629* (2022).
16. Deng, Naihao, Yulong Chen, and Yue Zhang. "Recent advances in text-to-SQL: A survey of what we have and what we expect." *Proceedings of the 29th International conference on computational linguistics*. 2022.
17. Wei, Jason, et al. "Chain-of-thought prompting elicits reasoning in large language models." *Advances in neural information processing systems* 35 (2022): 24824-24837.
18. Katsogiannis-Meimarakis, George, and Georgia Koutrika. "A survey on deep learning approaches for text-to-SQL: G. Katsogiannis-Meimarakis, G. Koutrika." *The VLDB Journal* 32.4 (2023): 905-936.
19. Katsogiannis-Meimarakis, Georgios G. "Data democratisation with deep learning: Structured query translation from and to natural language." (2023).
20. VAN DEN HEUVEL, J. A. N., MONSIEUR, G., TAMBURRI, D. A., & DI NUCCI, D. A. R. I. O. (2023). Data Mesh: a Systematic Gray Literature Review.