INTERNATIONAL JOURNAL OF INFORMATION TECHNOLOGY RESEARCH STUDIES (IJITRS)
(Open Access, Double-Blind Peer Reviewed Journal)
ISSN Online: 3139-3209

Self-Service Business Intelligence and Analytics in Digital Transformation

Meena Jose Komban

Assistant Professor, Department of Computer Science, YuvakshatraInstitute of Management Studies (YIMS), Mundur, Kerala, India

Article information

Received: 20th April 2026
Received in revised form: 22nd May 2026
Accepted: 24th June 2026
Available online: 30th July 2026

Volume: 2
Issue: 3
DOI: https://doi.org/10.63090/IJITRS/3139.3209.0034

Abstract

The digital-transformation era has changed how organizations generate, govern, and consume analytical insight. This paper analyzes the shift from centralized, IT-centric reporting toward self-service business intelligence (BI), in which business users author governed analyses without continuous engineering intervention. The discussion traces the architectural move from on-premises data warehouses to cloud data warehouses and lakehouse platforms that separate compute from storage and unify structured and semi-structured data under open table formats. A layered reference architecture is proposed. It comprises integration, storage and processing, augmented analytics, semantic and governance, and consumption layers. The paper then examines augmented analytics and the use of large language models for natural-language querying and natural-language-to-SQL translation. It pairs these capabilities with the governance controls that make self-service trustworthy at scale: semantic metric layers, row-level security, data catalogs, and lineage. A structured comparison of four leading platforms, namely Microsoft Power BI, Tableau, Google Looker, and Qlik, is presented across authoring, semantic modeling, cloud scalability, augmented analytics, and governance dimensions using illustrative capability scores. A five-level maturity model is introduced to help organizations benchmark their analytical capability and sequence investment. The analysis concludes that durable value comes not from tool selection alone but from coupling self-service freedom with a governed semantic foundation that treats data as a managed product. The metrics reported here are illustrative and synthesized from public industry sources, not from a single deployed study.

Keywords:- Self-Service Business Intelligence, Digital Transformation, Cloud Data Warehouse, Lakehouse, Augmented Analytics, Natural Language to SQL, Data Governance, Semantic Layer, Analytics Maturity.

.

I. INTRODUCTION

Digital transformation has recast data. What was once a passive by product of operations is now a strategic asset that shapes competitive position across nearly every sector. As organizations digitize customer journeys, supply chains, and internal processes, the volume, velocity, and variety of available data have grown well beyond what centrally produced reports can serve [1]. A coherent transformation strategy must specify not only which technologies to adopt but also how decision-making itself becomes data-driven, and it must avoid common traps such as technology-first initiatives disconnected from measurable outcomes [2]. Business intelligence, the set of technologies and practices for collecting, integrating, analyzing, and presenting business information, sits at the center of this shift [3].

For much of its history, BI was an IT-centric discipline. Specialists modeled data into enterprise warehouses, and a central team produced reports in response to formal requests [4], [5]. The model delivered consistency and control, but it created a structural bottleneck. Business users waited days or weeks for new reports. The backlog of analytical requests grew faster than central teams could clear it. A persistent gap opened between the questions decision-makers needed answered and the speed at which the organization could answer them [6].

Self-service BI emerged as a response to this bottleneck. By equipping business analysts with governed datasets, intuitive visual authoring tools, and curated semantic models, organizations let domain experts explore data and build their own analyses without continuous engineering support [7]. Three forces accelerated the transition: the migration of analytical workloads to elastic cloud platforms, the maturation of augmented analytics that automates parts of insight generation, and, most recently, large language models (LLMs) that let users query data in plain language [8]. Self-service freedom without governance, however, reintroduces the inconsistency that centralized BI was built to prevent. The central technical problem of the modern era is to reconcile autonomy with trust.

This paper offers a technical analysis of self-service BI in the digital-transformation era. Its contributions are fourfold. First, it synthesizes the architectural evolution from data warehouses to cloud lakehouses. Second, it proposes a layered reference architecture that positions augmented analytics and governance as first-class layers. Third, it presents a structured, multi-dimensional comparison of four leading platforms, Power BI, Tableau, Looker, and Qlik, drawing on a recent technical vendor analysis [9]. Fourth, it introduces a five-level maturity model with associated governance and adoption guidance. All quantitative figures are illustrative and synthesized from public industry sources; the paper does not report a single deployed empirical study.

II. EVOLUTION OF BI AND RELATED WORK

A. From Decision Support to Enterprise Data Warehousing

The intellectual roots of BI lie in decision-support systems and, later, in the data-warehousing movement of the 1990s. Inmon described the corporate information factory, in which a normalized, subject-oriented, integrated, time-variant, and non-volatile warehouse acts as the single source of truth feeding downstream data marts [4]. Kimball offered a complementary, bottom-up philosophy centered on dimensional modeling, where conformed dimensions and fact tables organized as star schemas optimize query performance and readability for business users [5]. Both methodologies remain foundational. The dimensional model in particular still shapes how semantic layers and self-service datasets are designed today.

Davenport and colleagues later reframed analytics as a source of strategic differentiation. They argued that organizations could compete on analytics by embedding fact-based decisions into their core processes and culture [10]. That perspective shifted the conversation from infrastructure to value. It also foreshadowed the later emphasis on democratizing access to data, so that analytical capability is not confined to a specialist elite.

B. The Rise of Self-Service and Visual Analytics

The 2010s saw the ascent of visual, self-service analytics tools that lowered the technical barrier to data exploration. Interactive visualization, grounded in principles set out by Few and others, made patterns and outliers visible to non-technical users, while drag-and-drop authoring removed the dependence on hand-written queries [7], [11]. Independent industry evaluations document this transition. The Gartner Magic Quadrant for Analytics and Business Intelligence Platforms, in particular, repeatedly identifies ease of use, governed self-service, and augmented capabilities as the primary axes of differentiation among vendors [12].

C. Cloud Data Warehouses and the Lakehouse

A decisive architectural change came with the move to the cloud. Cloud-native data warehouses introduced an architecture that separates compute from storage, letting each scale independently and enabling elastic, pay-per-use processing over very large datasets [13]. The data-lake paradigm, in parallel, offered inexpensive object storage for raw, semi-structured, and unstructured data [14]. Early data lakes, however, lacked the transactional guarantees and schema management of warehouses. The lakehouse architecture was proposed to unify the two worlds. It implements warehouse-grade management features, including ACID transactions, schema enforcement, and time travel, directly on low-cost object storage through open table formats [15]. Open formats such as Delta Lake and Apache Iceberg provide the transactional metadata layer that makes this convergence practical, separating the storage of data from the engines that query it [16], [17]. For self-service BI, the lakehouse matters because it supports governed, performant analytics over a single copy of data, with no cost or latency penalty for replicating that data into a separate warehouse.

III. SELF-SERVICE BI REFERENCE ARCHITECTURE

A dependable self-service capability rests on a layered architecture. Each layer has one clear responsibility and a governed contract with the layers above and below it. Figure 1 presents the proposed reference architecture. It comprises six layers: source systems, integration, storage and processing, augmented analytics, semantic and governance, and consumption. The design principle is consistent throughout: push freedom upward to business users while concentrating control in the semantic and governance layer.

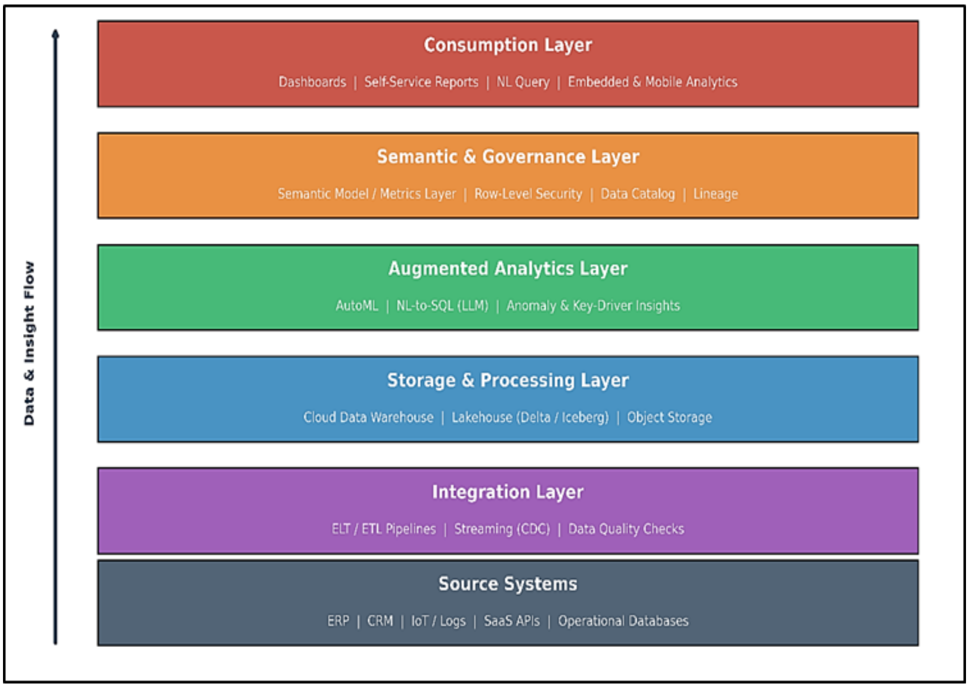

Fig. 1: Proposed layered reference architecture for governed self-service BI, in which a semantic and governance layer mediates between cloud storage and business consumption.

A. Integration, Storage, and Processing

The integration layer ingests data from operational sources: enterprise resource planning, customer relationship management, software-as-a-service applications, logs, and event streams. It uses batch extract-load-transform (ELT) pipelines and change-data-capture streaming [18]. The contemporary ELT pattern loads raw data into the cloud platform first and transforms it in place, exploiting elastic compute and keeping an auditable raw history [13]. The storage and processing layer is realized as a cloud data warehouse or a lakehouse on open table formats, supplying the transactional guarantees and elastic scale that downstream analytics depend on [15], [16]. Data-quality and validation checks belong here. Placed at this layer, they stop defects from propagating into self-service datasets, where the defects would be far more costly to detect and correct.

B. The Semantic and Governance Layer

The semantic layer is the architectural keystone of trustworthy self-service. It encodes business definitions, including metrics, dimensions, hierarchies, and relationships, as governed, reusable objects. A term such as "active customer" or "net revenue" then resolves to one agreed calculation, no matter who builds the report. By centralizing metric logic, the semantic layer prevents the proliferation of subtly inconsistent definitions that erodes trust in self-service environments [6], [7]. The governance controls sit alongside it: role-based and row-level security, so users see only the data they are entitled to; a data catalog that makes certified datasets discoverable; and lineage that traces every figure back to its sources for audit and impact analysis [19]. This layer is what lets the consumption layer above it stay genuinely open without becoming chaotic.

C. The Consumption Layer

The consumption layer is where business value is realized. It hosts interactive dashboards, self-service report authoring, natural-language query interfaces, and analytics embedded into operational applications and mobile experiences. Every artifact in this layer draws on certified semantic objects. As a result, analyses produced independently by different teams stay mutually consistent. Embedding analytics directly into the workflows where decisions are made, rather than confining them to a standalone BI portal, is a defining trait of mature, transformation-era deployments [3], [11].

IV. COMPARISON OF SELF-SERVICE BI PLATFORMS

Four platforms dominate enterprise self-service BI: Microsoft Power BI, Tableau (Salesforce), Google Looker, and Qlik. Each is consistently positioned among the leaders in independent market evaluations, yet each embodies a distinct architectural philosophy [12]. The comparison that follows synthesizes a recent comprehensive technical analysis of enterprise BI platforms [9] with vendor documentation and public evaluations. Capability scores in Figure 2 and Table 1 are illustrative. They convey relative emphasis rather than exact rankings, which shift with every product release.

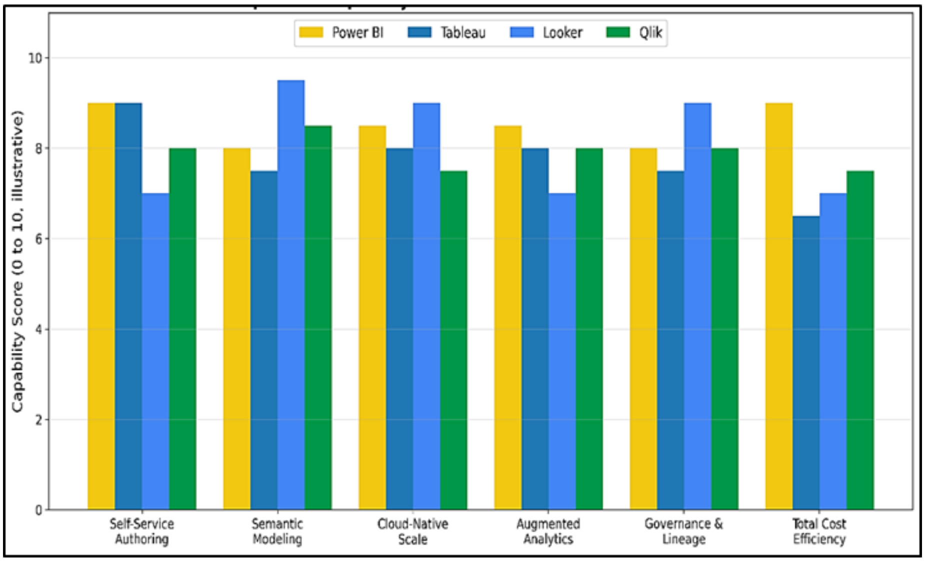

Fig. 2: Illustrative comparative capability assessment of Power BI, Tableau, Looker, and Qlik across six self-service dimensions (scores 0 to 10).

Power BI pairs strong self-service authoring with deep integration into the Microsoft ecosystem and an aggressive cost position. For enterprises already invested in that ecosystem, it is often the default choice [9]. Tableau is widely admired for the depth and fluidity of its visual-analytics experience, and it excels where exploratory, visually driven analysis is paramount [11]. Looker is distinguished by its code-defined semantic model, which expresses metrics as version-controlled, reusable definitions and aligns naturally with lakehouse-centric, in-database analytics. Qlik differentiates itself through an associative in-memory engine that lets users explore relationships across all data without predefined drill paths. Table 1 summarizes the comparison across the dimensions most relevant to governed self-service.

Table 1. Illustrative Feature Comparison of Leading Self-Service BI Platforms

Dimension	Power BI	Tableau	Looker	Qlik
Self-service authoring	Excellent	Excellent	Good	Very good
Semantic / metrics layer	Very good	Good	Excellent	Very good
Cloud-native scalability	Very good	Very good	Excellent	Good
Augmented analytics	Very good	Very good	Good	Very good
Governance & lineage	Very good	Good	Excellent	Very good
Core engine	VertiPaq in-memory	Hyper in-memory	In-database LookML	Associative in-memory
Typical strength	Ecosystem + cost	Visual exploration	Governed semantics	Associative discovery

No single platform is uniformly superior. The appropriate choice depends on existing technology investments, the relative priority placed on visual exploration versus governed semantics, and the target deployment model. One trend cuts across all four: each is increasingly architected to query cloud warehouses and lakehouses directly. That reflects the broader shift toward a shared, governed data foundation beneath whichever consumption tool an organization selects [9], [15].

V. AUGMENTED ANALYTICS AND LARGE LANGUAGE MODELS

Augmented analytics applies machine learning and natural-language technologies to automate parts of data preparation, insight discovery, and explanation that once required a data scientist. Its capabilities include automatic detection of anomalies and trends, key-driver analysis that surfaces the factors most associated with an outcome, and automated machine learning that fits and ranks candidate models [20]. By lowering the analytical skill threshold, augmented analytics extends self-service beyond power users to a wider population of decision-makers [8].

A. Natural-Language Querying and NL-to-SQL

The most consequential recent development is the application of large language models to the natural-language interface. Natural-language-to-SQL translation lets a user pose a question in plain language, for example, "what was quarter-over-quarter revenue growth for the western region?", and have the system generate and execute the corresponding query [8]. Research on semantic parsing and text-to-SQL, exemplified by large cross-domain benchmarks, established the task and its evaluation. Modern LLMs have since sharply improved generation accuracy on realistic schemas [21], [22]. When an LLM is grounded in the governed semantic layer rather than left to interpret raw tables, its generated queries inherit certified metric definitions and security rules. That grounding substantially reduces the risk of plausible but wrong answers.

B. Reliability, Grounding, and Limitations

Progress has been rapid, yet LLM-driven analytics carries well-documented risks. Models can hallucinate columns or join conditions that do not exist. They can misread an ambiguous business term. They can silently apply an incorrect aggregation. Three practical mitigations keep a human in the loop: grounding generation in the semantic layer, constraining outputs to the catalog of certified metrics, and returning the generated SQL for inspection [19], [22]. The aim is not to replace the analyst. It is to compress the distance between a business question and a trustworthy, governed answer. Seen this way, augmented analytics and self-service governance are complementary, not competing, concerns.

VI. GOVERNANCE, CHALLENGES, AND MATURITY MODEL

A. Governance and Data-as-a-Product

Governance is the discipline that makes self-service sustainable. It reaches beyond access control to the certification of trusted datasets, stewardship roles with clear accountability, and policies for metric definition and change management [19]. A growing organizational pattern treats curated datasets as managed products, each with documented owners, service expectations, and quality guarantees. Consumers can then rely on such a dataset as they would on any other product. This data-as-a-product orientation pushes responsibility for quality toward the domains that understand the data best, while a federated governance framework keeps definitions and controls globally consistent.

B. Adoption Challenges

Many obstacles to self-service adoption are organizational rather than technical. Poor data quality undermines trust in any analysis. Inconsistent metric definitions generate conflicting numbers that drain confidence. Uneven data literacy limits how far self-service can spread. Without a semantic layer, an unchecked proliferation of redundant reports can recreate the very fragmentation that self-service was meant to resolve [6], [7]. Figure 3 summarizes two things: the trajectory of the wider BI and analytics market, and the benefits that organizations most often associate with self-service adoption. These figures are illustrative syntheses of public industry sources, not measurements from a specific deployment.

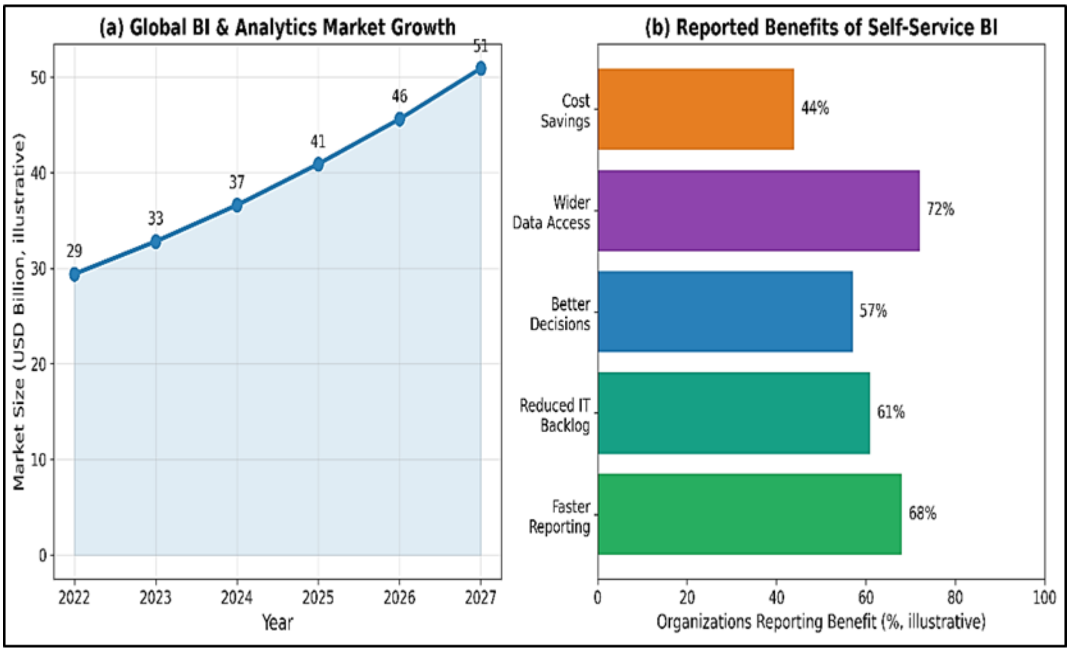

Fig. 3: (a) Illustrative growth trajectory of the global BI and analytics market and (b) frequently reported benefits of self-service BI adoption.

C. A Five-Level Maturity Model

To help organizations benchmark capability and sequence investment, Figure 4 presents a five-level maturity model. At Level 1 (Initial), analysis is ad-hoc and spreadsheet-driven, with no single source of truth. Level 2 (Managed) adds a central warehouse and standardized, IT-produced reports. Level 3 (Self-Service) establishes governed datasets and a semantic layer that empower business authoring. Level 4 (Augmented) brings automated insight, natural-language querying, and embedded analytics. Level 5 (Pervasive) reflects a genuine data culture, in which governed, real-time analytics is woven into everyday decisions across the organization. Progression is cumulative. Each level builds on the governance foundation of the one below it.

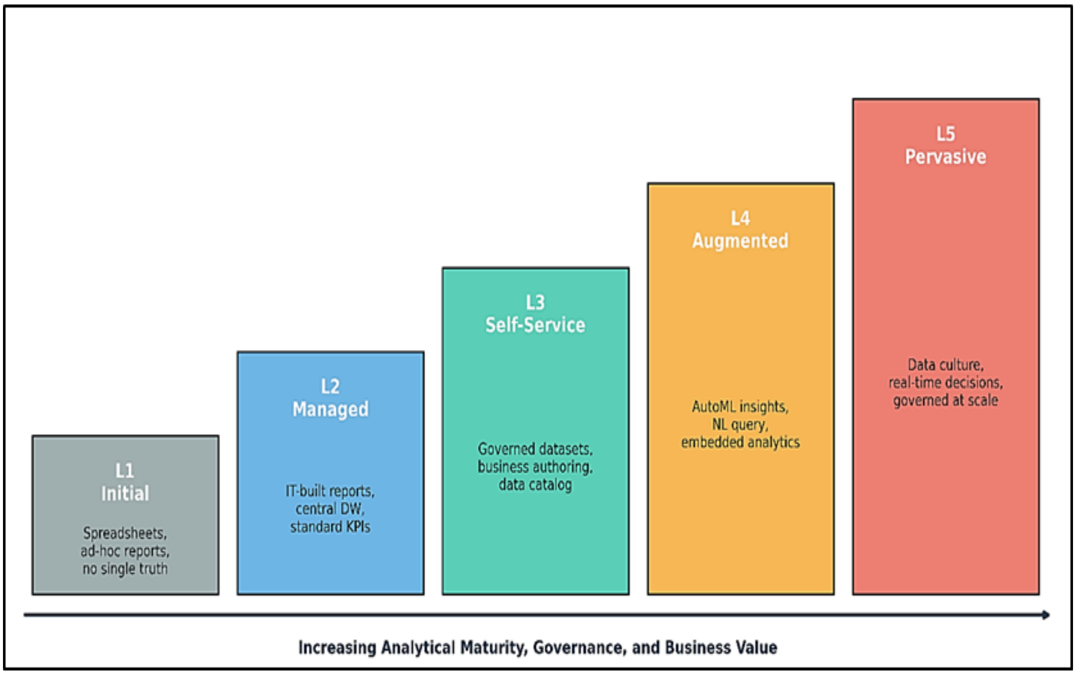

Fig. 4: Five-level self-service BI maturity model, from ad-hoc reporting to pervasive, governed analytics.

Table 2. Characteristics of the Self-Service BI Maturity Levels

Level	Stage	Data Foundation	Primary Authors	Key Risk
L1	Initial	Spreadsheets / silos	Individuals	No single truth
L2	Managed	Central warehouse	IT / BI team	Report backlog
L3	Self-Service	Governed semantic layer	Business analysts	Definition sprawl
L4	Augmented	Lakehouse + ML/LLM	Analysts + automation	Ungrounded AI output
L5	Pervasive	Federated data products	Whole organization	Governance at scale

The maturity model also clarifies sequencing. Deploying augmented analytics (Level 4) before a governed semantic layer (Level 3) tends to amplify inconsistency rather than insight, because automated and LLM-generated outputs inherit whatever definitional ambiguity sits beneath them. A disciplined path secures a governed foundation first, then layers automation on top [2], [19].

VII. CONCLUSION

Self-service business intelligence is a defining capability of the digital-transformation era. It shifts analytical authorship from a central IT function to the business users closest to each decision. This paper traced the architectural arc from enterprise data warehouses to cloud lakehouses, proposed a layered reference architecture that gives the semantic and governance layer a first-class role, compared the four leading platforms (Power BI, Tableau, Looker, and Qlik), and examined how augmented analytics and large language models are reshaping the analytical interface through natural-language querying and NL-to-SQL translation.

The recurring lesson is simple. Self-service freedom and governance are complementary, not opposing, forces. Durable value comes not from selecting a single tool but from establishing a governed semantic foundation, treating curated data as a managed product, and grounding automation in certified definitions. The maturity model offered here gives organizations a practical way to benchmark capability and sequence investment, so they advance toward pervasive, trustworthy analytics. Future work should validate the proposed architecture and maturity model empirically across deployments and quantify the reliability gains obtained by grounding LLM-based querying in the semantic layer. The metrics presented in this paper are illustrative syntheses of public sources and are not drawn from a single deployed study [2], [9], [12].

REFERENCES

[1] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, 2015.

[2] S. Charly, "Digital Transformation Strategy: Where to Start and What to Avoid," Int. J. Inf. Technol. Res. Stud. (IJITRS), vol. 1, no. 3, pp. 180–189, Oct. 2025, https://doi.org/10.5281/zenodo.17510445 .

[3] H. Chen, R. H. L. Chiang, and V. C. Storey, "Business intelligence and analytics: From big data to big impact," MIS Quart., vol. 36, no. 4, pp. 1165–1188, Dec. 2012.

[4] W. H. Inmon, *Building the Data Warehouse*, 4th ed. Indianapolis, IN, USA: Wiley, 2005.

[5] R. Kimball and M. Ross, *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling*, 3rd ed. Indianapolis, IN, USA: Wiley, 2013.

[6] B. Evelson, "Topic Overview: Business Intelligence," Forrester Research, Cambridge, MA, USA, 2008.

[7] C. Imhoff and C. White, "Self-Service Business Intelligence: Empowering Users to Generate Insights," TDWI Best Practices Report, The Data Warehousing Institute, Renton, WA, USA, 2011.

[8] R. Sallam, C. Howson, *et al*., "Augmented Analytics Is the Future of Data and Analytics," Gartner Research, Stamford, CT, USA, 2017.

[9] K. Abraham, "Business Intelligence Tools Comparison: Tableau, Power BI, and Alternatives: A Comprehensive Technical Analysis of Enterprise BI Platforms," Int. J. Inf. Technol. Res. Stud. (IJITRS), vol. 2, no. 1, pp. 34–41, Jan. 2026, https://doi.org/10.5281/zenodo.18490659 .

[10] T. H. Davenport and J. G. Harris, *Competing on Analytics: The New Science of Winning*. Boston, MA, USA: Harvard Business Review Press, 2007.

[11] S. Few, *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Burlingame, CA, USA: Analytics Press, 2009.

[12] Gartner, "Magic Quadrant for Analytics and Business Intelligence Platforms," Gartner Research, Stamford, CT, USA, 2024.

[13] B. Dageville *et al*., "The Snowflake elastic data warehouse," in *Proc. 2016 ACM SIGMOD Int. Conf. Management of Data*, San Francisco, CA, USA, 2016, pp. 215–226.

[14] P. Russom, "Data Lakes: Purposes, Practices, Patterns, and Platforms," TDWI Best Practices Report, The Data Warehousing Institute, Renton, WA, USA, 2017.

[15] M. Armbrust, A. Ghodsi, R. Xin, and M. Zaharia, "Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics," in *Proc. 11th Conf. Innovative Data Systems Research (CIDR)*, 2021.

[16] M. Armbrust *et al*., "Delta Lake: High-performance ACID table storage over cloud object stores," *Proc. VLDB Endowment*, vol. 13, no. 12, pp. 3411–3424, Aug. 2020.

[17] Apache Software Foundation, "Apache Iceberg: A high-performance open table format for huge analytic datasets," *Apache Iceberg Documentation*, 2023. [Online]. Available: https://iceberg.apache.org

[18] P. Vassiliadis, "A survey of extract-transform-load technology," *Int. J. Data Warehousing Mining*, vol. 5, no. 3, pp. 1–27, 2009.

[19] J. Ladley, *Data Governance: How to Design, Deploy, and Sustain an Effective Data Governance Program*, 2nd ed. Cambridge, MA, USA: Academic Press, 2019.

[20] G. Shmueli and O. R. Koppius, "Predictive analytics in information systems research," *MIS Quart.*, vol. 35, no. 3, pp. 553–572, 2011.

[21] T. Yu *et al*., "Spider: A large-scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task," in *Proc. 2018 Conf. Empirical Methods in Natural Language Processing (EMNLP)*, Brussels, Belgium, 2018, pp. 3911–3921.

[22] J. Li *et al*., "Can LLM already serve as a database interface? A big bench for large-scale database grounded text-to-SQLs (BIRD)," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023.