EMPOWERING AI AND ADVANCED ANALYTICS THROUGH DOMAIN-CENTRIC DECENTRALIZED DATA ARCHITECTURES

Meethun Panda [1] and Soumyodeep Mukherjee [2]

[1] Associate Partner, Bain & Company, Dubai, UAE
[2] Associate Director, Genmab, Avenel - NJ, USA

ABSTRACT

As organizations aim to become data-driven, they face a persistent paradox: centralization enables control and consistency, while decentralization fosters scalability and innovation. This "Data Platform Unification Paradox" creates a dynamic tension that challenges the design and management of modern data platforms. This paper introduces Data Mesh as a solution to this paradox, combining domain-driven decentralization with federated governance to balance these competing demands.

The study demonstrates that Data Mesh not only addresses scalability and agility challenges but also enables organizations to enhance data quality, governance, and collaboration. It highlights the transformative impact of Data Mesh in fostering domain-centric innovation, streamlining data ownership, and supporting advanced analytics, including AI and generative AI applications. Furthermore, the findings suggest that emerging technologies like quantum databases and multi-agent frameworks powered by Large Language Models (LLMs) require robust decentralized architectures to achieve their full potential. These insights provide actionable pathways for organizations aiming to modernize their data ecosystems while navigating the complexities of centralization and decentralization.

KEYWORDS

Data Mesh, Data Governance, Centralization, Decentralization, Data Paradox, AI, Quantum Databases, LLM Multi-Agent Systems, Distributed Data Platforms

1. INTRODUCTION

The rapid proliferation of data sources and the increasing demand for real-time analytics have reshaped the landscape of data platform design. Traditional approaches such as data warehouses and data lakes offer centralized control but struggle to scale with the growing complexity and heterogeneity of modern data ecosystems. Conversely, decentralized architectures promise scalability and agility but often lack consistency and governance, creating operational challenges. This paradox of centralization versus decentralization, termed here as the Data Platform Unification Paradox, highlights the competing priorities in data management. Centralization ensures control, standardization, and governance, but risks creating bottlenecks and limiting flexibility. Decentralization empowers domain teams and fosters innovation but can lead to fragmented data silos and governance gaps.

David C. Wyld et al. (Eds): ADCO, CMCA, NeCoM, NLPI, CSITEC, BDBS, DMML, SP – 2025
pp. 75-85, 2025. CS & IT - CSCP 2025 DOI: 10.5121/csit.2025.150506

This paper proposes Data Mesh as a solution to this paradox. By decentralizing data ownership while implementing a federated governance model, Data Mesh achieves a balance between scalability and control, enabling organizations to accelerate analytics and AI innovations.

2. KEY CHALLENGES FOR MODERN DATA ARCHITECTURES

As organizations increasingly embark on data-driven journeys to unlock business value from the data assets, it becomes necessary to embrace architectural changes required to maintain and enhance existing system landscape. Technology leaders have spent decades figuring out the best data-centric architecture patterns tailored to their organization needs. However, the ever-changing nature of data with proliferation of sources and users makes it challenging to settle on one model that supports real-time, data-driven decision making.

Additionally, a lack of clarity around governance models - especially regarding data ownership - has made it difficult for organizations to manage data efficiently as it moves across systems.

Organizations must address several challenges to navigate the complexities of the Data Platform Unification Paradox:

1. **Architectural Strategy**: Defining an organization's data architecture strategy amidst the growing complexity of data sources and use cases
2. **Governance Models**: Defining governance frameworks that align with organizational goals.
3. **Ownership and Collaboration**: Ensuring clarity in data ownership and fostering effective collaboration.
4. **Regulatory Compliance**: Adapting to evolving regulatory requirements without sacrificing agility.
5. **Data Silos**: Breaking down barriers to ensure seamless data access across domains.

As the organizations scale, these challenges are further compounded by evolving regulatory needs, mergers and acquisitions, and new models of customer service delivery. For today's technology leaders, addressing these dimensions is critical to delivering scalable, data-driven use cases quickly and efficiently, and they must evaluate their data maturity regularly

These challenges underscore the need for a paradigm shift in data architecture, emphasizing both technical scalability and organizational alignment.

The next section introduces **Data Mesh** as the decentralized architectural framework designed to enable organizations to achieve scalability, agility, and governance over centralized monolithic architectures that have existed historically

3. DATA MESH: RESOLVING THE DATA PARADOX

Data Mesh introduces domain-oriented architecture where data is treated as a product. It decentralizes data ownership, empowering domain teams to manage the lifecycle of their data while adhering to global governance standards.

Data Mesh emphasizes sharing data domain as a product for different use case needs from BI/reporting to advanced analytics via a diverse set of data accessibility methods. It empowers domain teams to manage the entire life cycle of their data. This includes localizing changes to the data domains and improving data quality to support agility and scale.

In a logical view, it represents a mesh of interconnected nodes - each node representing the data domain as a product. Without bringing data into a common monolithic data platform, it localizes business data domains in each business unit who owns data and who understands the particular data domain at best. Data domains focus on where data resides in an enterprise and specify data ownership. Data owners are given ultimate responsibility for all data quality decisions and key quality indicators including data quality improvement - a necessary ingredient for AI use cases.

Key Principles of Data Mesh:

1. **Domain Ownership**: Each domain team is responsible for its data products, ensuring data quality and contextual relevance.
2. **Data as a Product**: Data is packaged with clear documentation, discoverability, and APIs for easy consumption.
3. **Federated Governance**: Governance policies are centrally defined but allow domains autonomy for local adaptations.
4. **Infrastructure as a Platform**: A shared, self-service infrastructure enables domain teams to focus on innovation rather than operational complexity.

3.1. Stages for Data Mesh Implementation

The implementation of Data Mesh can be broken into key stages, each addressing specific technical and organizational challenges:

Table 1. Phases of Implementing Data Mesh in a Data Ecosystem

SL No.	Stage	Definition	Challenge	Data Mesh Consideration
1.	**Data Acquisition**	Capturing raw data from diverse sources	Ensuring data accuracy and breaking silos	Decentralized domain teams capture and manage data
2.	**Data Storage**	Organizing collected data	Defining taxonomy and ownership	Domain teams optimize storage within governance boundaries
3.	**Data Transformation**	Preparing data for analytics	Scaling preparation processes	Domain teams adopt self-service transformation tools
4.	**Data Analytics**	Generating actionable insights	Producing relevant and scalable insights	Domain teams adopt self-service transformation tools
5.	**Data Consumption**	Sharing and using data products	Enabling easy discoverability and access	Data marketplaces promote reusable data products

4. EVOLUTION OF DATA PLATFORM

The evolution of data platforms reflects the tension between centralization and decentralization:

1. Data Warehouses: Highly centralized, designed for reporting and structured analytics but limited in scalability for advanced use cases.
2. Data Lakes: Centralized yet flexible, capable of handling unstructured data but prone to governance and quality issues.
3. Data Mesh: A decentralized approach addressing the limitations of its predecessors by localizing ownership and governance at the domain level.

Challenges in Legacy Approaches

Traditional architecture relies on heavy ETL pipelines and central repositories, creating inefficiencies in data synchronization, quality, and governance. Data Mesh overcomes these challenges by fostering peer-to-peer data sharing through APIs and event-driven architectures, reducing the need for duplicative pipelines.

Deep Dive

Let's take a closer look at the data platform landscapes which have evolved in the last few decades catering to the end user consumption - Data Warehouse approach, Data Lake approach and Data Mesh based approach

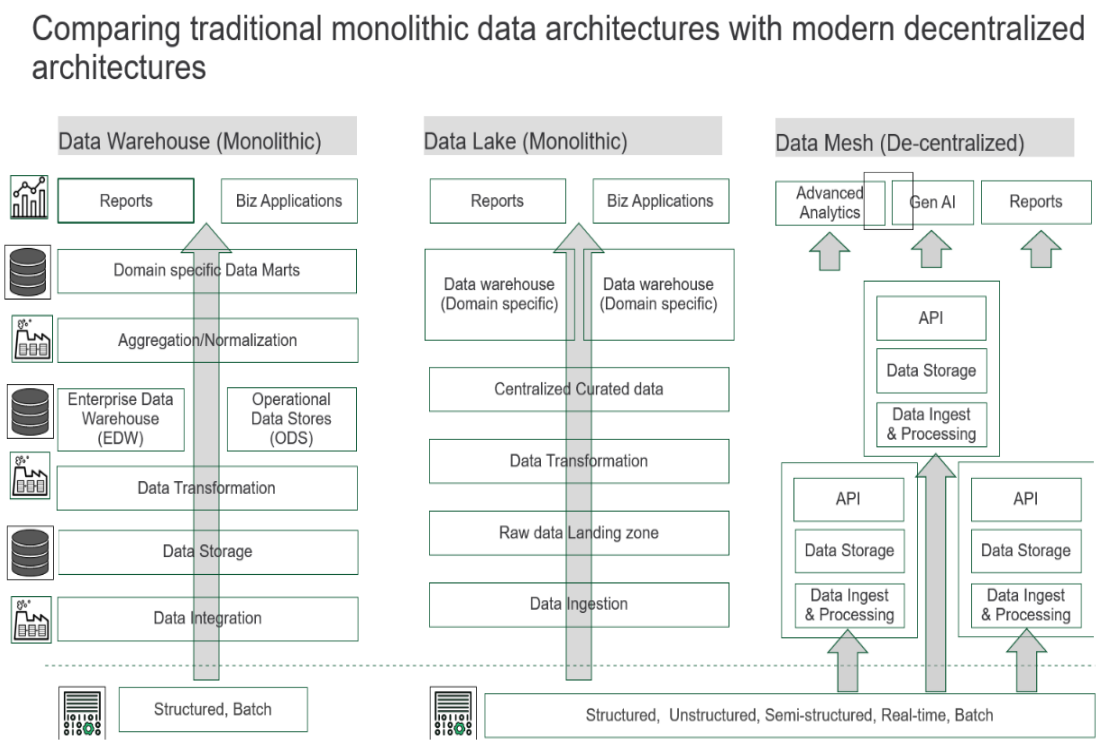

Figure 1. Data platform evolution

Data warehouse or Data Lake approach to build a data platform is monolithic in nature and limits agility and sustainability. In short, data warehouse-based architecture is achieved primarily by means of facts and related dimensions. But the principle behind the scene ranges from extracting data from operational sources, transforming into a dimensional schema, loading into the warehouse, and serving visualization/reporting users or analysts.

Data lake architecture evolved mitigating gaps in warehouse principles and enabling AI use cases in addition to supporting massive distributed processing of data. Data lake architecture principal ranges from extracting data from operational sources, loading into a scalable enterprise data repository in a particular format, processing raw data (cleanse, enrich, aggregate) based on the data models, and serving to a variety of end users with a diverse set of needs. It supports convergence of warehouse and lake including the support for streaming.

5. IMPACT ON ORG STRUCTURE AND COLLABORATION FRAME WORK

When we look at the Data lake based platform, there are typically 3 actors : Data engineers acquiring and processing the data based on end user consumption, Data consumers (e.g.,Data scientist, BI/Reporting developers) building their use cases such as advanced analytics model, and platform engineer maintaining the Dev Sec Ops.

What we miss here in this approach is the true interpretation of data and its ownership. Data engineers will bring just the data from the raw sources with limited understanding of data and curate for data scientists' needs. As the data is curated, data consumers can not claim that they own this data. The source system owners often think that they lose control of ownership of the data as it is acquired to another system.

Furthermore, in a data lake-based approach, we center around building too many ETL(Extract-Transform-Load) data pipelines, copying data from various data sources into data lake and curating further for end user consumption (Data scientists, BI/Reporting users, analysts etc.). The problem comes from having limited insights on data quality and maintainability of execution of data pipelines for reprocessing the already processed data making sure data at multiple places are synchronized.

The better approach would be to leverage a shared infrastructure which enables data discovery and consumption from data origins. Data Mesh allows peer-to-peer interaction among the data consumers by means of API or micro services or Kafka without copying data over to a central data lake and limiting the massive number of data pipelines built. For instance, personalization analytics use cases can directly consume required data domains (e.g., customer, loyalty) from various sources instead of bringing those data daily into a lake. This makes faster development of use cases with a higher degree of data quality insights including frequency, recency etc. Access to the data at scale no matter where the data resides is key to continuously deliver advanced analytics use cases

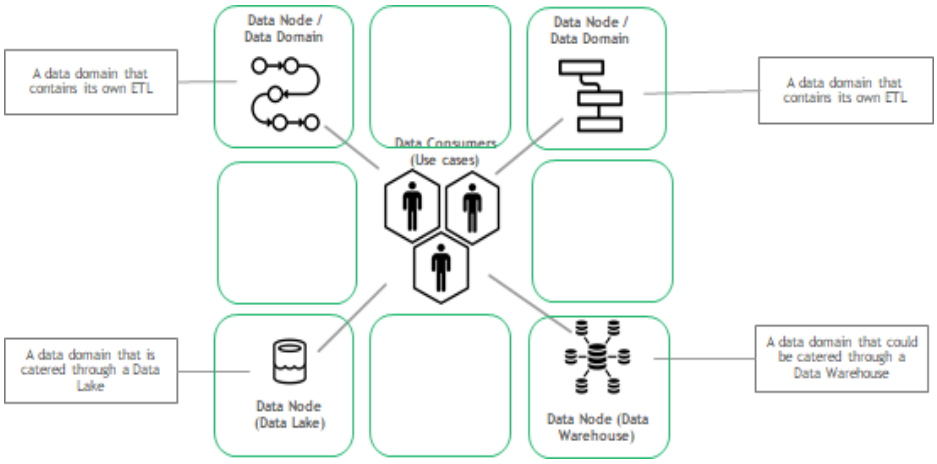

Figure 2. Data sharing across domains to enable various Use-cases

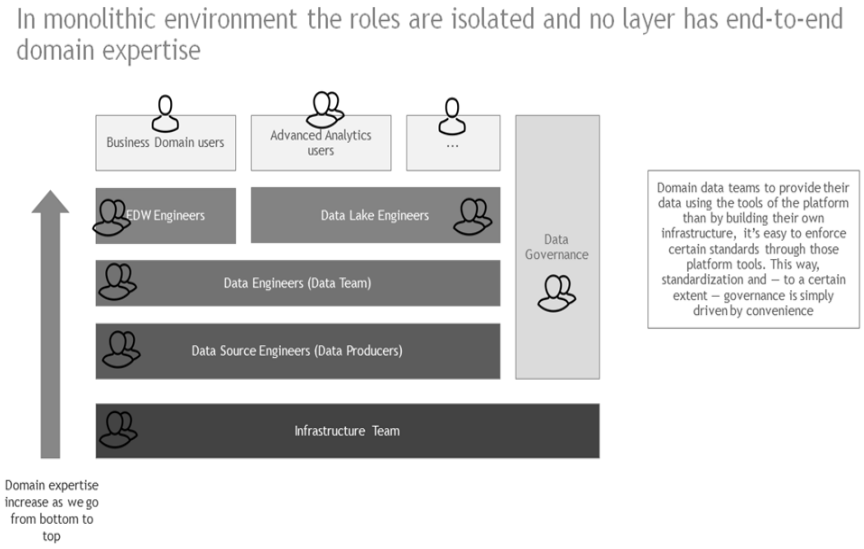

Figure 3. Org structure for traditional monolithic data platform

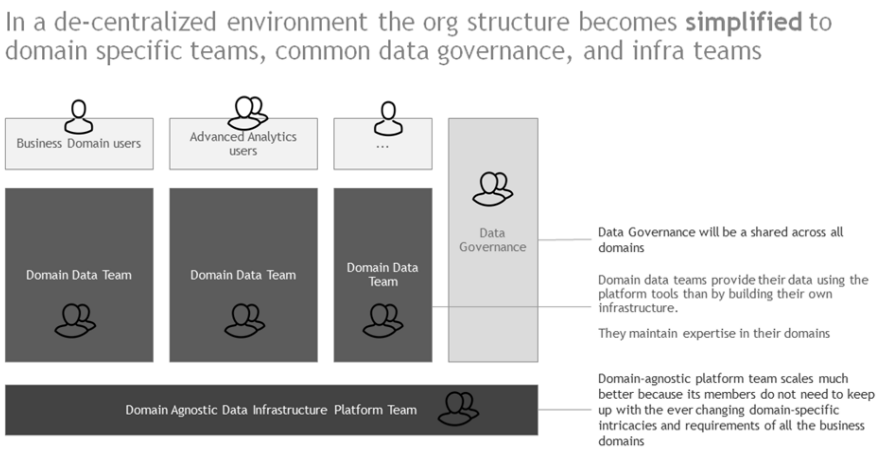

Figure 4. Collaboration framework in the data mesh paradigm

Roles of Key Teams in the Data Ecosystem within the Data Mesh Framework –

- **Data Governance Team (Federated):** Global team defining gov. policies & standards while providing domains team autonomy to extend based on local rules
- **Domain agnostic Platform Team (Centralized)**: Unified platform team providing self service capabilities required by data products
- **Data domain teams (Decentralized)**: Team owns data products in given domain and are responsible for design, dev. and operation of the given data product(s)

6. INDUSTRY APPLICATIONS

While the concept of data mesh is promising, it is also important to look at the feasibility of implementation, cost effectiveness, and roadmap with a promise of faster time to market.

Below is an illustration primarily focusing on data mesh architecture for a multi-brand restaurant company with key data domains such as customer, transaction, menu, store etc. are required to enable advanced analytics or BI reporting use cases

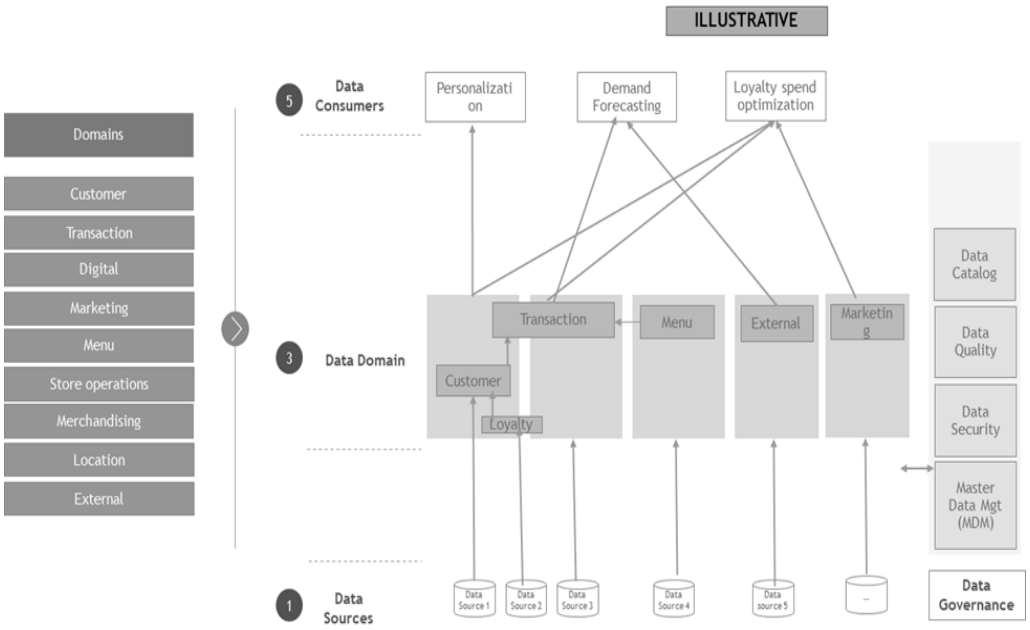

Figure 5. Illustrative Data Mesh Reference for multi-brand restaurant

Above demonstration of a Data Mesh implementation for a food and beverage industry supports data domain as a product exposed for other use cases and users including data ownership and transformation under the responsibility of each domain's teams. Well-designed data products in Data mesh can enable GenAI use cases at scale supporting various types of use cases - Summarization, Generation, Interaction, Knowledge management etc. This is possible by leveraging the data mesh paradigm

- **Domain-specific ownership of data** fosters a sense of responsibility for data quality, and ultimately leads to better training data for AI models
- **Data marketplace is facilitating the identification of available data**, and ultimately leads to faster time to market in building new AI use cases

- **Data mesh improves collaboration between data scientists & domain experts**, and ultimately leads to more relevant AI use cases
- **Data mesh enforces data governance and controls access authorized data** (for end users & AI models), and ultimately mitigates security risks & compliance

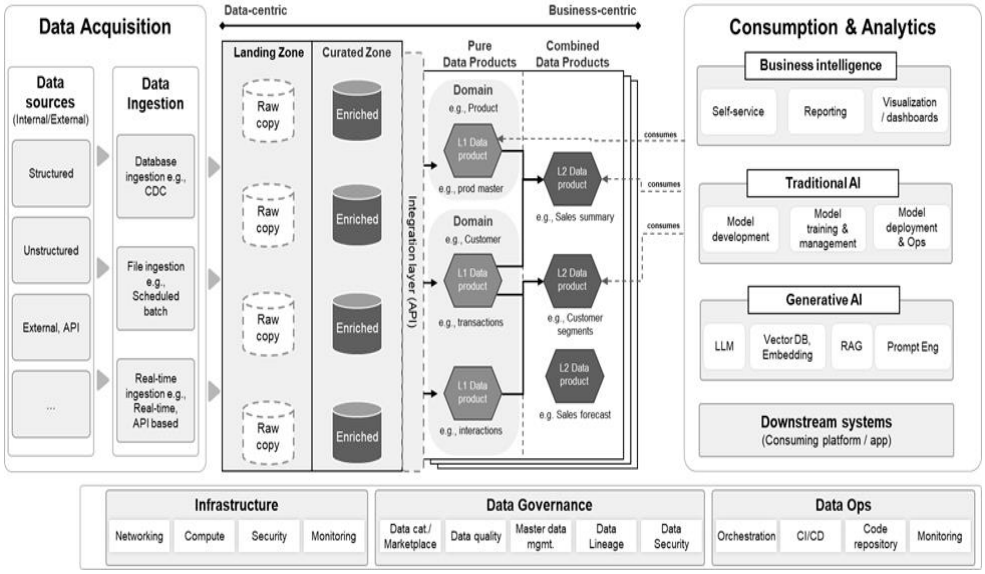

Figure 6. Generic Reference data architecture in the data mesh paradigm

7. PROS & CONS - DATA PLATFORM UNIFICATION PARADOX

The evolution of data platforms has been shaped by a recurring struggle: balancing centralization and decentralization. This tension, termed the **Data Platform Unification Paradox**, is a fundamental challenge for organizations seeking to derive value from their data assets. While Data mesh has its advantages and helps overcome certain demerits of Centralized/Monolithic data platform approach however it is not a silver bullet for analytics platform and there are still scenarios where a centralized approach might shine and hence it's important to have a balanced view when seeking to resolve data platform challenges. This section reviews key merits and demerits of both the approaches.

7.1. Centralization: Strengths and Limitations

Centralized data platforms, such as data warehouses and data lakes, provide a unified repository for managing and analyzing data. They are particularly effective for certain types of analytics:

- **Self-Service Analytics**: Promotes data democratization by making data accessible to a wide range of users.
- **Exploratory Analytics**: Enables business users to explore large datasets without requiring deep domain expertise.

Strengths:

- **Consistency**: Centralized platforms maintain uniform governance, security, and compliance policies across the organization.
- **Efficiency**: Centralized data processing reduces redundancy in data pipelines.
- **Control**: Simplifies regulatory compliance and enforces data quality standards.

Limitations:

- **Limited Scalability**: Struggles to accommodate the complexity and diversity of modern data ecosystems.
- **Inflexibility**: Adapting centralized platforms to specific domain needs can be slow and resource intensive.
- **Ownership Gaps**: Centralized ownership can dilute accountability for data quality at the source.

7.2. Decentralization: Opportunities and Challenges

Decentralized data platforms, as exemplified by domain-driven architectures like Data Mesh, emphasize local ownership and flexibility. These platforms are well-suited for:

- **Domain-Centric Data Products**: Supporting advanced analytics tailored to specific business domains.
- **Domain-Centric AI Models**: Enabling AI models trained on data with deep domain knowledge.
- **Business Intelligence and Descriptive Analytics**: Facilitating focused, business-specific insights for targeted decision-making.

Opportunities:

- **Scalability**: Decentralized platforms adapt easily to the growing diversity of data sources and use cases.
- **Agility**: Domain teams can innovate and develop analytics solutions without bottlenecks.
- **Improved Quality**: Localized ownership fosters accountability for data quality and relevance.

Challenges:

- **Fragmentation**: Inconsistent standards and governance across domains can hinder interoperability.
- **Duplication**: Risks of redundant efforts and data silos if not governed properly.
- **Complexity**: Requires significant coordination between domains and central governance teams,

7.3. The Need for Balance

The **Data Platform Unification Paradox** illustrates the duality of these approaches. While centralization excels in democratizing data and fostering exploratory insights, decentralization shines in creating domain-focused analytics and actionable intelligence for specific business areas.

To harness the best of both worlds, organizations must:

- Balance global governance with local autonomy.
- Scale analytics and AI use cases without sacrificing control and consistency.
- Foster domain-specific innovation while ensuring enterprise-wide compliance and data quality.

The next section introduces **Data Mesh** [1] as the decentralized architectural framework designed to resolve this paradox, enabling organizations to achieve scalability, agility, and governance since centralized monolithic architectures have existed historically

8. EMERGING TECHNOLOGIES AND FUTURE DIRECTIONS

The landscape of analytics is poised for a revolution with the advent of quantum databases and multi-agent systems powered by LLMs. These technologies promise unprecedented computational power and the ability to process vast and complex datasets in real-time. However, their adoption requires robust data platforms capable of supporting:

- **Quantum Databases**: Offering massively parallel processing and quantum state storage, requiring seamless integration with existing data infrastructures.[14]
- **LLM Multi-Agent Frameworks**: Enabling collaborative decision-making across intelligent agents, demanding scalable, high-quality, and real-time data access. [15]

To fully leverage these advancements, organizations must adopt architectures like Data Mesh that ensure flexibility, scalability, and governance at scale.

8.1. Future Research Opportunities

The intersection of domain-centric data sharing and decentralized architectures presents numerous avenues for future research. One promising direction involves exploring how advancements in quantum databases and multi-agent systems powered by large language models (LLMs) can be seamlessly integrated into Data Mesh frameworks to enable real-time, intelligent decision-making. Additionally, the development of adaptive governance models that dynamically balance centralization and decentralization as organizational needs evolve is a critical area of investigation. Further studies could also focus on leveraging Data Mesh principles to support emerging use cases in generative AI, edge computing, and real-time analytics, paving the way for more resilient, scalable, and innovative data ecosystems.

9. SUMMARY AND CONCLUSION

Although Data Mesh adoption is still in its early stages, it represents a transformative shift for organizations striving to become truly data driven. By addressing the limitations of monolithic platforms and decentralizing data ownership, Data Mesh enables faster, more scalable, and innovative use cases across BI, ML, and generative AI applications

The Data Platform Unification Paradox illustrates the inherent tension between centralization and decentralization in modern data architectures. Data Mesh provides a balanced solution, combining domain-level ownership with federated governance to support scalable analytics and innovation. As emerging technologies like quantum databases and LLM multi-agent frameworks gain traction, the need for robust and adaptive data platforms will only intensify.

Technology leaders must embrace this paradigm shift, aligning their strategies to deliver not just analytics and AI value today but also prepare for the quantum and agentic future of data-driven enterprises.

REFERENCES

[1] Zhamak Dehghani, (2020), "How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh," ThoughtWorks.

[2] Fowler, M., (2003), "Patterns of Enterprise Application Architecture," Addison-Wesley.

[3] Kiran, B., Vohra, D., & Sengupta, S., (2019), "Data Lake for Enterprises: Leveraging Data Lakes for Advanced Analytics," Springer.

[4] G. Piatetsky-Shapiro, (2019), "The Evolution of Data Warehousing and Big Data," KDnuggets.

[5] Z. Li & J. Zhang, (2020), "Federated Data Governance: Balancing Local Autonomy and Global Standards," IEEE Transactions on Data Engineering, Vol. 15, No. 3, pp. 234-245.

[6] Srivastava, J., (2021), "Quantum Databases: Advancing Beyond Classical Data Storage," Journal of Quantum Computing, Vol. 7, No. 2, pp. 95-110.

[7] T. Nguyen & L. Johnson, (2020), "AI-Driven Data Architectures for Business Intelligence," Data Science Quarterly, Vol. 6, No. 4, pp. 88-101.

[8] Zhamak Dehghani, (2022), "Data Mesh: Delivering Data-Driven Value at Scale," O'Reilly Media.

[9] Y. Chen, S. Wang, & R. Patel, (2018), "Decentralized Data Platforms and the Role of Blockchain," ACM Transactions on Information Systems, Vol. 36, No. 5, pp. 423-437.

[10] H. J. Watson, (2018), "Big Data Analytics: Concepts and Techniques," Communications of the ACM, Vol. 61, No. 2, pp. 22-25.

[11] D. Laney, (2012), "The Emerging Role of Data Governance in Modern Organizations," Gartner Research Report.

[12] B. Stonebraker, (2016), "The Case for Data Warehouses in an Era of Data Lakes," IEEE Data Engineering Bulletin, Vol. 39, No. 2, pp. 3-7.

[13] A. Gawande, T. Shroff, & L. Peters, (2019), "Domain-Centric AI Models: Enabling Innovation through Localized Data Architectures," Journal of AI Research, Vol. 12, No. 1, pp. 55-70.

[14] Soumyodeep Mukherjee and Meethun Panda, "General-Purpose Quantum Databases: Revolutionizing Data Storage and Processing", International Journal of Data Engineering (IJDE), Volume (9), Issue (1), 2024 (ISSN: 2180-1274)

[15] Soumyodeep Mukherjee, "The Rise of Multi-Agent LLMs: Insights from Agent Smith and the Challenges of Distributed Data Processing in AI Systems", International Journal of Artificial Intelligence and Expert Systems (IJAE), Volume (13), Issue (1), 2024 (ISSN: 2180-124X)

AUTHORS

Meethun Panda, Associate Partner at Bain & Company is a thought leader having deep expertise in technology, cloud, Data, AI, LLM, and Quantum computing. He brings 15+ years of experience across technology realms leading and delivering large-scale data and analytics transformations. One of the leading Data/AI consultants in North America by CDO Magazine. Meethun's key focus is to drive Tech/AI strategy and large-scale transformation cases for fortune 500 clients.

Soumyodeep Mukherjee, Associate Director of Commercial Data Engineering at Genmab (an international biotech company specializing in antibody research for cancer and other serious diseases) is a seasoned data professional with over 14 years of experience in data engineering, architecture, and strategy. Currently steering commercial data initiatives at Genmab, Soumyodeep's key focus is on crafting innovative data and analytics strategies to drive commercialization efforts. Previously, he served as a Project Leader at BCG.X and a Data Specialist at McKinsey & Company, where he led teams in implementing robust, end-to-end data solutions across healthcare, insurance, and retail sectors. His expertise includes deploying machine learning models and leveraging Generative AI to streamline data management and enhance organizational efficiency.