ISSN 2277-2685
IJESR/June. 2025/ Vol-15/Issue-3s/367-375
Nandari Sai Nikitha *et. al.*, / International Journal of Engineering & Science Research

Data Mess to Mesh: Strategic Lessons from the Field

A V S Radhika[1] , Nandari Sai Nikitha[2], Ballure Suprathika[3]

[1]Assistant Professor, Department of CSE, Bhoj Reddy Engineering College for Women.

[2,3]B. Tech Students, Department of CSE, Bhoj Reddy Engineering College for Women.

Abstract: As the significance of data and artificial intelligence escalates, firms endeavor to adopt a more data-driven approach. Nonetheless, existing data infrastructures are not inherently structured to accommodate the magnitude and breadth of data and analytics applications. Indeed, current designs often do not provide the anticipated value associated with them. Data mesh is a socio-technical, decentralized, and distributed framework for business data management. The notion of data mesh remains fresh and is devoid of empirical inputs from the field. An grasp of the driving elements for implementing data mesh, the related problems, tactics for execution, its commercial implications, and possible archetypes is lacking. To rectify this deficiency, we do 15 semi-structured interviews with industry specialists. Our findings indicate that organizations encounter challenges in transitioning to federated data governance linked to the data mesh concept, the redistribution of responsibility for the development, provision, and maintenance of data products, and the understanding of the overarching concept. We propose many implementation tactics for enterprises, including the establishment of a cross-domain steering unit, monitoring data product use, achieving early fast wins, and favoring small, specialized teams that prioritize data products. While we recognize that firms must tailor implementation techniques to their own requirements, we also identify two archetypes that provide more detailed recommendations. Our results consolidate views from industry experts and provide academics and professionals with first instructions for the effective implementation of data mesh.

Keyword TERMS Big data, data governance, data mesh, management information systems.

Introduction

As data volume escalates, firms endeavor to adopt a more data-driven approach to surpass competitors. The International Data Corporation (IDC) predicts that data volume will more than double from 2022 to 2026, with private entities driving this expansion.

Nonetheless, in the swiftly changing realm of data management, the shortcomings of conventional centralized data architectures reliant on data warehouses and data lakes are more evident. These systems find it difficult to manage the growing amount and diversity of data, presenting considerable issues for central IT departments The production of data in a more dispersed fashion and at an elevated volume burdens these departments, resulting in extended response times for data requests .

This delay is a significant impediment affecting data consumers' access to pertinent data and diminishing the organization's overall agility and responsiveness in a data-driven world.

The increasing diversity of data adds another degree of complication. Central IT often lacks the requisite understanding of specialized areas necessary for the successful management of this diversity. The deficiency of domain-specific knowledge hinders the precise and effective management of data, resulting in discrepancies between data availability and the real requirements of various organizational units.

Furthermore, the centralized method of data administration generates substantial issues pertaining to data ownership within the wider domain of data governance [3]. Without explicit data ownership, accountability for data quality and upkeep becomes unclear, resulting in possible complications with data integrity and quality. These obstacles together hinder unobstructed access to high-quality data, undermine data integrity, and prolong the time to market and value realization for data-driven projects [7].

As a result, the potential scope and efficacy of data and artificial intelligence (AI) applications are constrained, impeding the organization's transformation into a completely data-driven entity.

LITERATURE REVIEW

Organizations must constantly reevaluate and modify their data strategies, structures, and management systems to get value from the ever expanding volume of data in order to maintain competitiveness in the sector [14]. Historically, several terminology have arisen concerning similar ideas, including but not limited to "data warehouse," "data lake," and more recently, "data lakehouse," "data mesh," and "data fabric." This foundational part elucidates these terminologies, their fundamental principles, and interrelations. Generally, data warehouses and data lakes emphasize data management, whereas data lakehouses, data mesh, and data fabric pertain to data structures. Data management systems and architectures vary in their degree of abstraction. A data architecture may include and coordinate several data management systems [16].

Data warehouses are specialized databases that aggregate structured data from many sources and primarily function as centralized repositories for processed data [17]. They routinely retain data for business intelligence and reporting objectives, hence avoiding the preservation of data for prospective analysis [18].

Data lakes can consume data more rapidly, accommodate larger quantities, and store various data kinds. Unlike data warehouses, data lakes also retain unprocessed data for future analysis and potential commercial applications. Consequently, they have significant relevance for machine learning (ML) applications. The newly developed architecture integrates the adaptable storage capabilities of data lakes with the analytical framework of data warehouses, providing a scalable solution for the management and analysis of various data kinds.

This hybrid paradigm improves data accessibility and analytics, catering to the changing requirements of big data management .

Consistent with the relevant literature, we delineate and define the three concepts as follows:

Definition 1 (Data Warehouse): A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant repository of data that facilitates management decision-making.

Definition 2 (Data Lake): A data lake is a centralized repository system for the storing, processing, and analysis of unstructured, semi-structured, or structured raw data in its original format.

Definition 3 (Data Lakehouse): A data lakehouse is a data management system that utilizes low-cost, readily available storage for structured, semi-structured, and unstructured data, while still offering standard analytical database management system features and performance capabilities .

A pertinent but distinct phrase is data fabric. Data fabric is a technological framework that integrates

heterogeneous data from several sources, enabling companies to oversee and manage data irrespective of its location, while ensuring proper data governance and cataloging. Consequently, its main emphasis is on the integration of various data management systems, including data warehouses, data lakes, and data lakehouses. Data fabric employs extensive metadata [23] and a data virtualization layer [22] to facilitate access for users throughout the enterprise. The use of metadata is essential for accessing, finding, and comprehending data, as well as for automating data integration, engineering, and governance processes. This encompasses centralized administration of data access, privacy, and compliance issues.

Definition 4 (Data Fabric): A data fabric is a novel data management architecture aimed at achieving flexible, reusable, and enhanced data integration pipelines, services, and semantics.

Conversely, data mesh is a socio-technical framework that encompasses architectural elements. It further encompasses social and organizational elements such as decentralization and ownership. A data mesh, akin to a data fabric, often comprises several data management systems augmented by an integration and governance layer, coupled with a decentralized organizational framework. Reference [3] indicates that data mesh comprises four fundamental ideas enabling firms to handle data at scale. Initially, domain-specific decentralized data ownership: individual domains own the data they generate and use their subject expertise to enhance data quality.

Domains are defined as organization-specific demarcations of the relevant competitive boundaries of the organizations. Thus, domain knowledge refers to an individual's skill in a particular sector or area acquired via experience, study, or training. The production department serves as the authority over all production-related data due to their extensive experience and ability to comprehend intricate technical linkages inherent in the data. Secondly, data is seen as a product, with comprehensive accountability throughout its lifecycle. Data products are offered, including metadata, accessible options such as APIs, and the actual data. It is analogous to a software product that needs supplementary services, such security updates or manuals. Moreover, data products conform to the following usability attributes: discoverable, addressable, comprehensible, reliable, accessible, interoperable, useful, and secure [3]. The third concept, self-service data platform, delineates a specialized data platform that offers high-level abstract architecture for various domains, facilitating a high degree of autonomy within those areas. This is essential for domains to prevent the duplication of technological efforts and concentrate on the development of superior data products. The fourth and final principle: federated data governance delineates the governance framework for data products. Domain data product owners and pertinent stakeholders jointly establish uniform standards and norms, which are to be automatically implemented within each domain, to guarantee the interoperability of data products.

The significance of this matter is paramount, since data products provide the most value when integrated. The four principles together empower enterprises to transcend the constraints of centralized data systems, facilitating a more data-driven approach. The architectural idea is shown in Figure 1.

The implementation of a data mesh comprises three essential phases: exploration and bootstrapping,

expansion and scaling, and extraction and sustainability. Initially, certain domains function as both data suppliers and consumers, developing core methods and combining data into cohesive products. During the expansion and scaling phase of the mesh, a greater number of domains integrate, standardizing technological and organizational frameworks to facilitate swift scaling and the integration of older systems. In the extract and maintain phase, domains attain autonomous data ownership, concentrating on the optimization and refinement of data product distribution and use, so culminating in a mature, integrated data ecosystem. Each step builds upon its predecessor to improve scalability and integration across the company.

We like to emphasize that data mesh has architectural elements, as previously described, however fundamentally represents a socio-technical idea.

Consequently, drawing from reference [3], we delineate data mesh as follows.

Definition 5 (Data Mesh): Data mesh is a socio-technical, decentralized, and distributed framework for business data management.

In view of the existing ambiguity about the separation between data mesh and data fabric, we elucidate their similarities and distinctions below.

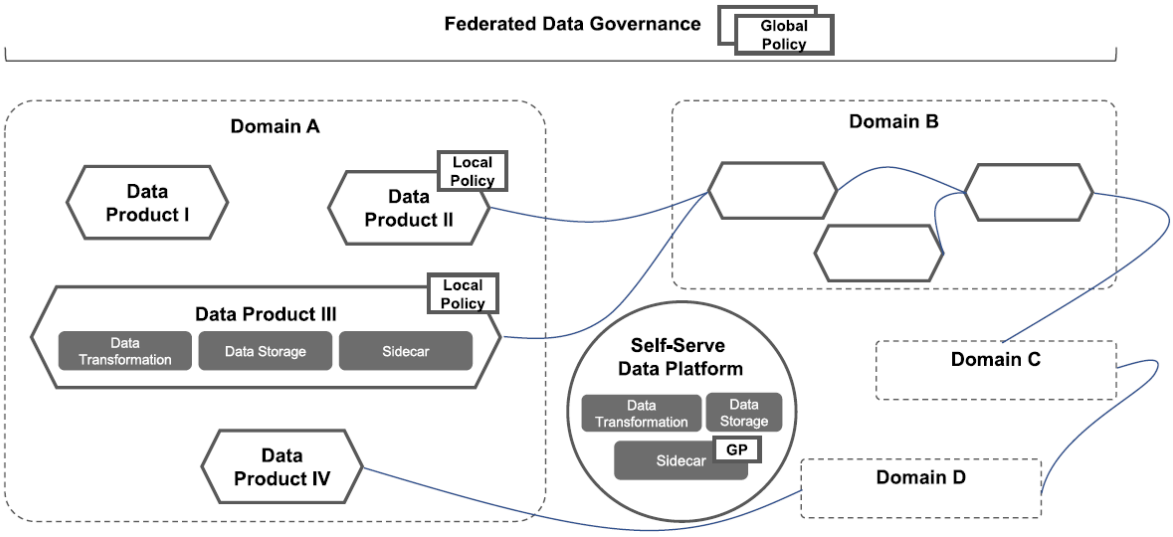

FIGURE 1. Conceptual overview of a data mesh based on the four key principles: 1) domain-oriented decentralized data ownership, 2) data as a product, 3) self-serve data platform, and 4) federated data governance. The figure shows different levels of granularity (high on the left and low on the right).[1]

METHODOLOGY

To provide a thorough understanding of the motivating reasons, obstacles, and implementation techniques associated with data mesh adoptions, 15 semi-structured expert interviews were performed from November 2022 to January 2023. The semi-structured interview approach is selected for its capacity to reconcile the rigidity of closed queries with the adaptability of open-ended inquiries. This equilibrium is essential for investigating intricate and innovative subjects, such as data mesh, while facilitating the emergence of new concepts and themes throughout the interview process [28]. In accordance with the methodology presented in

reference [29], an interview guideline is used to organize the interviews about the specified themes. The guideline was originally designed to encapsulate the fundamental elements of our study inquiry. We evaluated the efficacy of our strategy via pilot testing, refining the guideline accordingly. During the interview process, we refined the guideline to include themes introduced by the interviewees, ensuring a thorough examination of the subject. The primary version of the interview guideline was created after the fourth interview. We use a purposive sample strategy to interview partners from various sectors in order to thoroughly examine data mesh qualities by integrating multiple viewpoints and applications.

We use expert sampling by identifying specialists by their LinkedIn job titles and activity, particularly their posts and comments.

Furthermore, we actively engage with important players on LinkedIn who were instrumental in the success narratives of publicly accessible data mesh.

Candidates needed a minimum of one year of experience in data mesh and five years in data and AI to qualify for the interview process. In one case, an interviewee had just six months of experience in data mesh; yet, this individual was included owing to their substantial expertise in the closely related field of distributed data structures. Although one year may seem to be a minimal need, it signifies substantial proficiency in the nascent field of data mesh, which was presented in 2019. 3

Our survey includes participation from firms of diverse sizes and varied degrees of expertise on the subject.

Table 1 presents an overview of the interviewees and their attributes. Figure 2 presents an overview of the main themes addressed throughout the interviews.

Each interview underwent an initial independent analysis with open coding, which was then enhanced by axial coding to synthesize and facilitate inferences across interviews. Ultimately, axial codes were organized into themes using selective coding. The procedure was conducted 15 times, with each interview signifying one iteration. This method enabled the extraction of essential findings and permitted their integration into a more comprehensive framework.

In each iteration, we did a comprehensive check of the automatically generated transcripts to ensure content correctness. Text fragments were later paraphrased and reduced to provide a clearer perspective. Subsequently, we began the first code iteration at the interview level, using the paraphrased chunks. Subsequently, the codes were examined to ensure they appropriately represented the substance of the interviews. In the second step, interview-level codes were included into the comprehensive framework. To achieve this objective, we first categorized codes into the following primary themes: theoretical knowledge, case description, motivating factors, obstacles, implementation techniques, effects, preparedness, viewpoint, and archetypes to facilitate deductions according to the specified study topic. Figure 3 illustrates the principal themes together with their relative and absolute distributions in a pie chart.

During each cycle, we developed and refined sub-codes within each topic to aggregate analogous remarks from several interviews into axial codes [33]. Upon completing the axial coding, we refined each subgroup to derive chosen themes including motivating elements, difficulties, implementation techniques, effects, and archetypes, which are

detailed in the following section. We identified 717 (sub)-codes from 15 respondents, resulting from 48 hours of coding effort. Figure 4 illustrates the total quantity of coded segments among the respondents.

RESULTS

This section synthesizes results from the interviews.

We provide insights into interviewees' theoretical comprehension, their motivating drivers for adopting a data mesh, the hurdles they encounter, and the implementation methods they devise. Moreover, we concentrate on the effects seen by respondents. Ultimately, we provide two kinds of companies that implement the data mesh notion. We concentrate only on elements that are critically pertinent to data mesh applications. Nonetheless, many issues and implementation tactics are not exclusive to data mesh but pertain to the broader subject of change management and technological integration.

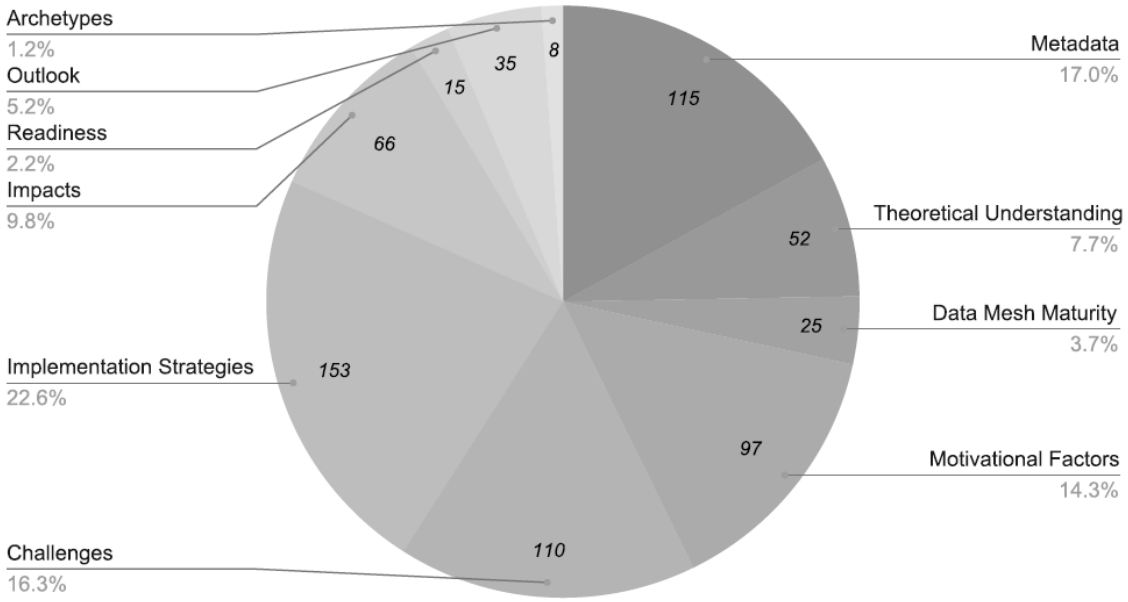

FIGURE 2. *Pie-chart of interview themes. Themes are sorted clockwise according to the interview guideline. 39 codes of the archive theme are omitted.*

CHALLENGES

Organizations encounter many problems with the implementation of data mesh. We primarily concentrate on difficulties that are distinctive and highly pertinent to data mesh-related subjects. The transition from centralized to federated data governance is seen as the primary difficulty for professionals.

Interviewees argue that the federated method complicates tasks and duties that were once controlled centrally. While they emphasize the significance of federated data governance to formulate rules based on domain requirements, interviewees underscore constraints linked to automated implementation, particularly in relation to security, regulatory, and privacy issues (A, E − 2, N, O). O observes that personnel in the area lack awareness about which data are safeguarded and controlled. N cautions that managers may "find

themselves at risk of incarceration" due to non-compliance with data regulations.

CONCLUSION

This study offers a thorough examination of the implementation of the data mesh idea across many sectors, informed by insights from 15 semi-structured expert interviews. The research fills a need in the current literature by providing real, empirical insights into the motives, obstacles, implementation techniques, and effects of data mesh implementations, which have mostly remained theoretical so far owing to the concept's novelty. The results indicate that motives for implementing a data mesh include the aspiration to diminish bottlenecks, use domain expertise, enhance data ownership, and dismantle data silos, all directed towards improving data accessibility and quality. These motives correspond well with the theoretical advantages proposed by Dehghani's basic theory on data mesh [3].

The shift to a data mesh architecture presents many issues. This encompasses the intricacies of transitioning to federated governance, overseeing the obligations associated with decentralized data ownership, guaranteeing superior metadata quality, and confronting the organizational opposition that may arise from substantial alterations in data management techniques. The study suggests various implementation strategies to address these challenges, including the formation of cross-domain units, enhancing and monitoring domain initiatives, achieving rapid successes, encouraging intentional adoption, enforcing committed ownership, and acknowledging the function of data stewards.

The first outcomes of data mesh deployments are encouraging, including increased accessibility and velocity of data retrieval, higher data quality, diminished redundancies, and a general advancement towards a more data-centric organization. These results validate the capacity of data mesh to enhance corporate data management procedures considerably.

The research delineates two initial organizational archetypes—startups and scaleups, alongside existing organizations—that gain from customized strategies for data mesh deployment. This distinction aids in comprehending how data mesh may be tailored to meet the unique requirements and attributes of various enterprises.

This study offers a comprehensive review of real-world experiences using data mesh, enhancing both academic research and practical applications in data management.

It establishes a foundation for further research and assists businesses in effectively preparing for the difficulties and possibilities associated with implementing a data mesh architecture. Subsequent study should persist in examining these topics using a complementary quantitative approach as data mesh evolves and its popularity increases.

LIMITATIONS & FUTURE RESEARCH

To address our study issue, we do 15 semistructured expert interviews. The qualitative aspect of our study leads to restricted quantitative validity. We substantiate the qualitative method due to the originality of the study issue. Future study should quantitatively examine the results, for instance, using surveys.

Concerning our sample methodology, we recognize a possible bias, as respondents may portray their data mesh implementations as more effective than they really are, particularly in their efforts to publicly establish themselves as leaders in the field

of data mesh. To tackle this difficulty, we have been forthright in conveying our exclusive emphasis on research endeavors. Additionally, we anonymize respondent data to promote candid discussions about both the favorable and unfavorable elements of their experiences inside their businesses. This methodology aims to reduce prejudice and promote a more precise and sophisticated comprehension of data mesh implementations.
Our primary offering consists of industry knowledge for experts implementing a data mesh. Nonetheless, we recognize that these techniques are only relevant to a limited extent based on the specific circumstances of each business. Consequently, pros must modify their technique accordingly—integrating just pertinent elements.
The proposed organizational archetypes provide a first step toward developing more detailed rules for companies based on distinct traits. Nonetheless, we recognize that the industry insights overall and the archetypes in particular are devoid of quantitative substantiation. This presents a significant possibility for future research, since this exploratory qualitative method may be enhanced by a quantitative investigation. In this regard, researchers may examine data mesh inside small and medium-sized enterprises to enhance the framework of archetypes. Furthermore, further research should thoroughly explore the technology implementation of the data mesh notion. This may include the meticulous design of technical data products and the integration of data warehouses, data lakes, or blob storage to actualize the data mesh idea. Furthermore, potential data mesh topologies and factors regarding optimal data mesh node sizes may be examined in greater depth.

REFERENCES

1. Jan Bode 1, Niklas Kühl 2, Dominik Kreuzberger1, And Carsten Holtmann, Toward Avoiding The Data Mess: Industry Insights From Data Mesh Implementations, Ieee Access, *Digital Object Identifier 10.1109/Access.2024.3417291*
2. Z. Dehghani,'' *Data Mesh: Delivering Data-Driven Value at Scale*. Sebastopol, CA, USA: O'Reilly Media, 2022.
3. I. A. Machado, C. Costa, and M. Y. Santos, ''Data mesh: Concepts and principles of a paradigm shift in data architectures,'' *Proc. Comput. Sci.*, vol. 196, pp. 263–271, Jan. 2022.
4. Z. Dehghani, ''How to move beyond a monolithic data lake to a distributed data mesh,'' Thoughtworks, Chicago, IL, USA, May 2019.
5. K. Vestues, G. K. Hanssen, M. Mikalsen, T. A. Buan, and K. Conboy, *Agile Data Management in NAV: A Case Study*, vol. 445. Cham, Switzerland: Springer, 2022, pp. 220–235.
6. V. K. Butte and S. Butte, ''Enterprise data strategy: A decentralized data mesh approach,'' in *Proc. Int. Conf. Data Anal. Bus. Ind. (ICDABI)*, 2022, pp. 62–66.
7. P. Awasthi and J. George, ''A case for data democratization,'' in *Proc. AMCIS*, Aug. 2020, pp. 1–23.
8. I. A. Machado, C. Costa, and M. Y. Santos, *Advancing Data Architectures With Data Mesh Implementations*, vol. 452. Atlanta, Georgia: Association for Information Systems, 2022, pp. 10–18.

9. N. J. Podlesny, A. V. D. M. Kayem, and C. Meinel, *CoK: A Survey of Privacy Challenges in Relation to Data Meshes*, vol. 13426. Cham, Switzerland: Springer, 2022, pp. 85–102.