Research Article

Impact of Data Mesh Architecture for Enterprise Data Lakes

Girish Ganachari*

*Corresponding author: Girish Ganachari, USA, E-mail: girish.gie@gmail.com

A B S T R A C T

The appearance of data mesh architecture brings changes with which the enterprises use data lakes. This type of architecture involves the decentralization of data ownership and governance since it is scalable, agile, of better quality, and consistent. It also helps to create necessary cooperation and innovation, and, at the same time, control costs and usage of resources. Delegating the management of data to specialized domain teams allows organizations to act faster in response to business requirements, enhance data accuracy, and thus promote the concept of constant enhancement. This paper covers the effects of the data mesh architecture on enterprise data lakes to understand how it is beneficial and how it can be implemented.

Keywords: Data Mesh, Data Ownership, Data Governance, At scale, Exclusive/Inclusive adaptability, Data Excellence, Cross-functional cooperation, Innovation, Cost optimization, Resource utilization

1. Introduction

In the current society where everything is analyzed and processed into data, businesses have the enormous challenge of handling large quantities of data from various sources[1]. The existing architecture of data lakes, with the focused management systems, tends to fail in meeting the growing needs for data availability, data quality, and flexibility in utilizing the data. This paper focuses on data mesh architecture in light of enterprise data lakes researching aspects of scalability, elasticity, constructiveness, quality, synergy, creativity, cost, and resource usage. Thus, based on the analysis of these aspects, one should be able to develop a clear overall picture of how ideas and practices associated with data mesh may redefine the concept of enterprise data management and how it can contribute to achieving strategic goals.

2. Decentralized Data Ownership and Governance

Data ownership and sovereignty, a prospect of the data mesh architecture, are revolutionary concepts changing how enterprises approach data[3]. In prior data lake designs, the control of the data is centrally located and it leads to excessive delays in decision-making and finger-pointing. Data mesh manages to tackle these problems by breaking down data ownership and making it central to domain teams, which are the most knowledgeable about the data in question[4]. It allows teams to operate data-like products and is given full control over the treatment of the data focusing on quality, safety, and availability.
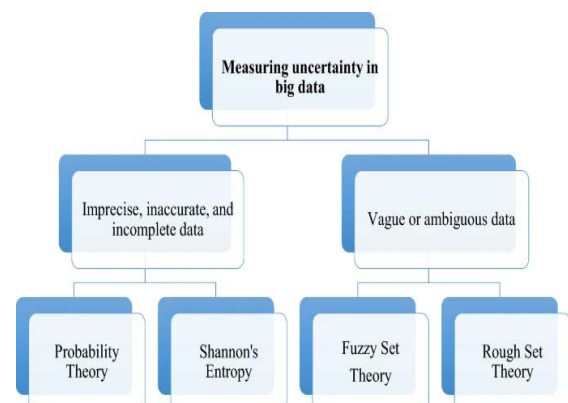

Figure 1: Uncertainty in big data.
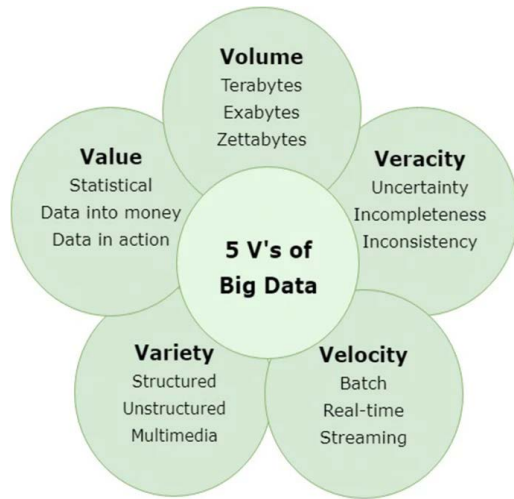(Source: Hariri et al. 2019)[2]

Figure 1.2: 5V's of Big data.
(Source: Hariri et al. 2019)[2]

It is also meaningful in terms of the domain-specific expertise that each of the domain teams possesses[5]. They can then apply the governance practices for their domain, which correspond to the requirements and goals of the domain. This leads to better outgoing data control since the teams that actively work with the data are the ones who set the policies and standards. Moreover, decentralization of operations results in a sense of responsibility, which is amplified by the fact that every team doing the processing of data is motivated to do so in the most accurate and efficient way possible.
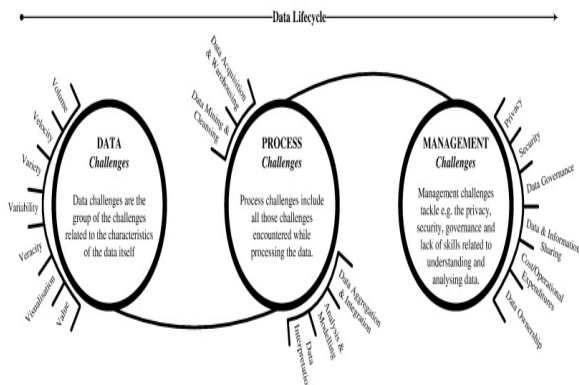

Figure 1.3: Lifecycle of Data.
(Source: Sivarajah et al. 2017)[1]

It increases the credibility and utilisation of the overall data and allows for a swift response to compliance and regulatory needs. This decreases the likelihood of data leakage and increases overall compliance with data privacy laws. The decentralization of data assets and their management improves the adaptability, responsibility, and overall responsiveness of the systems.

3. Improved Scalability and Agility

Scalability and agility are the major advantages that one can expect when using a data mesh architecture[6].

Centralized data lakes present in more conventional system architectures pose challenges of scalability because of the size, heterogeneity, and speed characteristics of data in today's business environments[6]. Such centralized systems may turn into a problem as these are limiting factors that do not allow to process and analyze data effectively. On the other hand, data mesh architecture deploys data management by dispersing them across domain levels where each domain manages its

data pipeline, data storage, and processing capabilities. What this brings is that the domain teams can now handle their data operations based on their needs without influences from the structures of a centralized approach.

Figure 1.4: Transforming Data Strategy with Data Mesh Architecture.
(Source: Loganathan, 2024)[27]

The received data mesh architecture implies the independence that leads to faster reactions to shifts in business requirements[7]. With the help of domain teams, it is possible to speed up the creation, implementation, and fine-tuning of data solutions that in turn decrease time on bringing new data products and services to the market.
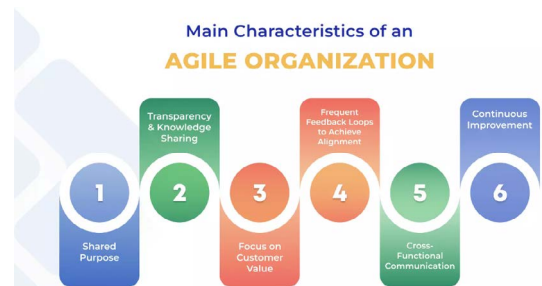

Figure 1.5: Main characteristics of an agile organisation.
(Source: businessmap, 2024)[28]

This agility is very important in today's dynamic business environment where an organization is able to quickly adapt to shifts in phenomena and this may translate to means to wealth. Also, decentralized scalability serves as a guarantee for avoiding the concentration of performance issues in a single domain, since each of the domains may allocate its resources according to workload and need[8]. In this connection, data mesh architecture greatly improves the data operation adaptability and extendibility of an organization and enables data evolution for the sake of future development[9].
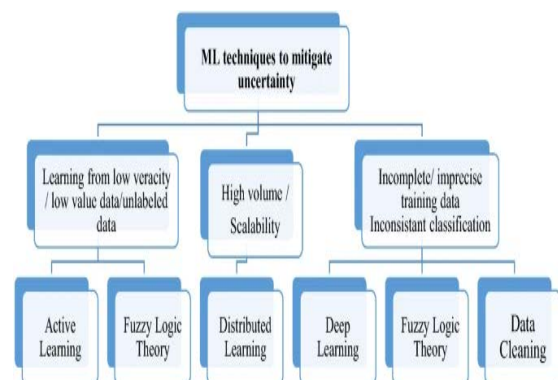

Figure 1.6: Techniques of handling uncertainty in big data.
(Source: Hariri et al. 2019)[2]

4. Enhanced data quality and consistency

Integration of multiple source systems improves data

quality and decreases variations as major benefits of data mesh architecture[10]. Data is approached as a product hence every domain team is held accountable for the entire pipeline process of the product they are delivering in the form of data[11]. Such a product-oriented approach guarantees that the information is collected, verified, updated, and enriched to meet the highest quality requirements. Domain teams are capable of applying specific quality assurance methods and validation procedures relating to their field and the type of data utilized.
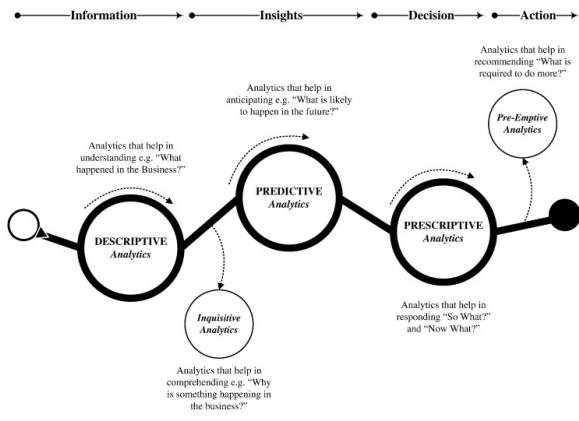

Figure 1.7: Classification of big data analytic methods. (Source: Sivarajah et al. 2017)[1]

Data mesh also promotes a consistently stable data environment through decentralization of the data management process[12]. Domain teams set down guidelines and specifications for formats, naming conventions, and processing of their data[24]. Such consistency is imperative in order to allow for proper data connection and coordination, which in turn allows for the improvement in Enterprise Business Intelligence and reporting. In addition, accountability within the context of a data mesh architecture guarantees the prevention of data quality problems as such concerns are promptly addressed in a data mesh environment[13]. This minimizes mistakes and discrepancies that can negatively impact business decisions.

The movement of data as a product in the context of data mesh produces better, more accurate, and standardized data[14]. This not only makes the results derived from data analysis accurate and more reliable but also optimizes the data operations in the entire firm. Ultimately, the fact that data thickens enhances its quality and makes it more consistent leading to improved decision-making, which in turn results in enhanced business outcomes and strategic success.

5. Increased Collaboration and Innovation

The use of a data mesh is useful as it enhances data working and creativity in a firm through distribution, instead of consolidating data management to a singular department or project[15]. In typical big data architectures based on data lakes, communication and centralized control points tend to inhibit cross-team interaction. They often do not have visibility into or direct access to the data residing in other areas of the data lake. Data mesh eliminated these problems by allowing the domain-oriented teams to have control and access to share their data more easily[16]. As a result, each team works with the product approach to produce data-related products that are ready for consumption and compatible with the other domains.

This decentralization helps for better cooperation of the teams, as they share and use each other's data products to get a deeper insight and to develop new ideas. When domain teams exchange case data and solution experiences, one team is already reaping the results on which another is working[17]. So, the process is cumulative, presenting a multiplying effect toward quicker advancement of new solutions and service creation. The social structure created by data mesh also insists on learning and adaptation because everyone is always improving the data products and processes that they are delivering[18]. This openness, and cross-fertilization of ideas results in fresh strategies for data management and analysis, creating competitiveness and organizational capacity to respond to changes and opportunities in the market faster.

6. Cost Efficiency and Resource Optimization

Data mesh architecture is quite cost-efficient and optimizes resource utilization[19].
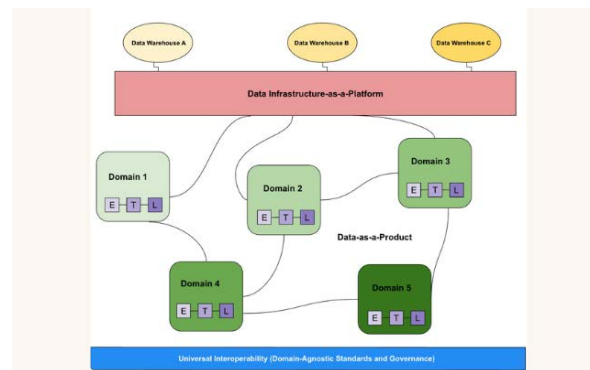

Figure 1.8: Data mesh structure. (Source: Moses, 2020)[26]

In the traditional centralized data lakes the cost of managing a single large-scale infrastructure can further lead to an internal sprawl of resources to handle various data requirements[20]. This structure also has problems since resources are located centrally and become wasteful in one place while scarce in others such as in human resources. Data mesh on the other hand solves the issues by decentralizing data processing and storage duties in the domain teams thus enabling each to align its capacities to the requirements it has[21].
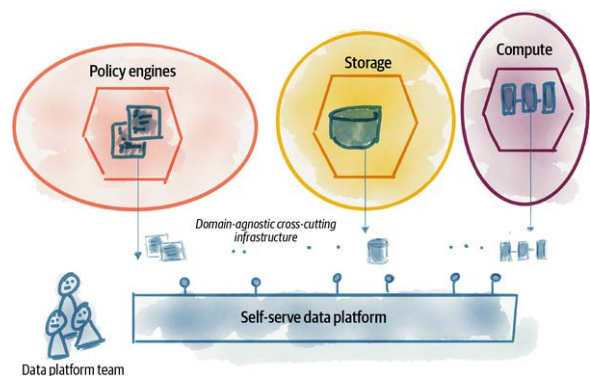

Figure 1.9: Data mesh usage. (Source: Maeda, 2024)[22]

This weak-centralized model virtually decreases the usage of a centralized IT enforcement thereby optimally utilizing computational capacity, storage space and other infrastructural attributes[22]. Domain teams can add resources or reduce them depending on the workload which helps to prevent over-allocating resources like in centralization. Also, a selective focus on resource allocation to match specific domain requirements makes it possible for every team to perform the best it can

without spending much[23]. It not only decreases the operating expenses but also proves advantageous for the efficient handling of data.

Independent management of data resources by the domain teams results in quick problem-solving and little or no system unavailability thus increasing efficiency in costs[24]. In this respect, data mesh architecture helps to release substantial overhead involved in centralized coordination and allows for a more accurate allocation of resources[25]. This cost efficiency also frees up resources for investment in other value-added activities, which are more strategic to the corporation's growth and development.

7. Conclusion

In conclusion, data mesh as an architecture provides a solid basis for changing the data lake concept in enterprises and has strengths in various perspectives of data management. Data mesh architecture can be considered as the modern approach to enterprise data management as it differs from the centralized data lake approach. It decentralizes the identities and governance of data to promote the cross-functional data teams optimizing and experimenting with their data pathways from the ground, thus maximising scalability, flexibility, and data excellence. Also, due to the integration of numerous teams in data mesh, work becomes more innovative and helps organizations manage data more effectively and be ready for changes in the market.

8. References

1. Sivarajah U, Kamal MM, Irani Z, Weerakkody V. Critical analysis of Big Data challenges and analytical methods. J Business Research 2017;70: 263-286.

2. Hariri RH, Fredericks EM, Bowers KM. Uncertainty in Big Data analytics: survey, opportunities, and Challenges. J Big Data 2019;6: 1-16.

3. Machado I, Costa C, Santos MY. Data-Driven information systems: The data mesh paradigm shift. International Conference on Information Systems Development 2021.

4. Joshi D, Pratik S, Rao M. Data Governance in data mesh infrastructures: The saxo bank case study. ICEB 2021 Proceedings 2021.

5. MacLeod M. What makes interdisciplinarity difficult? Some consequences of domain specificity in interdisciplinary practice. Synthese 2016;195: 697-720.

6. Armbrust M, Ghodsi A, Xin R, Zaharia M. Lakehouse: A new generation of open platforms that unify data warehousing and advanced analytics. 11th Annual Conference on Innovative Data Systems Research 2021.

7. Feng D, Tsolakis C, Chernikov AN, Chrisochoides NP. Scalable 3D hybrid parallel Delaunay image-to-mesh conversion algorithm for distributed shared memory architectures. Computer-Aided Design 2017;85: 10-19.

8. Sai AR, Buckley J, Fitzgerald B, Gear AL. Taxonomy of centralization in public blockchain systems: A systematic literature review. Information Processing & Management 2021;58: 102584.

9. Pfaff T, Fortunato M, Sanchez-Gonzalez A, Battaglia PW. Learning Mesh-Based Simulation with graph networks. arXiv 2021.

10. GokulRaj N, Umapathy K. 5G Wireless Mesh Network 802.11s Load Balancing Architecture for 802.11 Bgn Radio-PCI Interface. Procedia Computer Science 2016;87: 252-257.

11. Ismail A, Truong H-L, Kastner W. Manufacturing process data analysis pipelines: a requirements analysis and surve., J Big Data 2019;6.

12. Cilfone A, Davoli L, Belli L, Ferrari G. Wireless Mesh Networking: An IoT-Oriented Perspective Survey on Relevant Technologies. Future Internet 2019;11: 99.

13. Feng Y, Feng Y, You H, Zhao X, Gao Y. MeshNet: Mesh Neural Network for 3D Shape Representation. Proceedings of the AAAI Conference on Artificial Intelligence 2019;33: 8279-8286.

14. Chandramouli R, Butcher Z. Building secure microservices-based applications using service-mesh architecture. NIST Special Publication 2020.

15. Wolfengagen V, Ismailova L, Kosikov S, Dohrn J. Superimposing semantic mesh to prevent information processes entanglement. Procedia computer science 2020;169: 645-651.

16. Nguyen TT, Dahl VA, Bærentzen JA. Cache-mesh, a Dynamics data structure for performance optimization. Procedia Engineering 2017;203: 193-205.

17. Lehtinen TOA, Itkonen J, Lassenius C. Recurring opinions or productive improvements-what agile teams actually discuss in retrospectives. Empirical Software Engineering 2016;22: 2409-2452.

18. Prade LR, Uberti VA, Abaide AR, Pereira PRS, Figueiredo RM, Both CB. LoRa mesh architecture for automation of rural electricity distribution networks. Electronics Letters 2020;56: 739-741.

19. Darroudi SM, Gomez C. Bluetooth Low Energy Mesh Networks: A Survey. Sensors 2017;17: 1467.

20. Ramakrishnan R, Sridharan B, Douceur JR, et al. Azure Data Lake Store. Proceedings of the 2017 ACM International Conference on Management of Data 2017.

21. Lin K, Wang L, Liu Z. Mesh Graphormer. CVF 2021.

22. Maeda J. 4. Principle of the Self-Serve Data Platform - Data Mesh. oreilly 2024.

23. Gilani SSA, Qayyum A, Rais RNB, Bano M. SDNMesh: An SDN based routing architecture for wireless mesh networks. IEEE Access 2020;8: 136769-136781.

24. Sahni Y, Cao J, Zhang S, Yang L. Edge Mesh: A new paradigm to enable distributed intelligence in internet of things. IEEE Access 2017;5: 16441-16458.

25. He A, Feng G, Zhang J, Li P, Hei Y, Chen H. Click-Based asynchronous mesh network with bounded bundled data. ICPP '18: Proceedings of the 47th International Conference on Parallel Processing 2018.

26. Moses B. What is a data mesh-and how not to mesh it up. Monte Carlo Data 2020.

27. Loganathan P. Transforming your data strategy with data mesh architecture. 2024.

28. Businessmap. Organizational Agility: How to Adapt and React Quickly to Changes. Kanban Software for Agile Project Management 2024.