Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 44 No. 7
July 2023

Review on Data Mesh Architecture and its Impact

Ashish Dibouliya[1] (ashish.dibouliya@gmail.com)**, Dr. Varsha Jotwani**[2] (varsha.jotwani@gmail.com)
[1] Research Scholar (Computer Science & Technology), Rabindranath Tagore University, Bhopal M.P. India
[2] HoD (Computer Science & Technology), Rabindranath Tagore University, Bhopal M.P. India

Abstract— Data Mesh is a decentralized approach to data management that has gained significant attention in recent years. This review paper provides an in-depth analysis of Data Mesh and its architecture, focusing on its principles, components, and implications. The paper begins by introducing the concept of Data Mesh and its departure from traditional centralized data management approaches. It then explores the key principles of Data Mesh, including domain-oriented decentralized ownership, self-serve data infrastructure. The architecture of Data Mesh is dissected, discussing the various components such as data domains, data products, data infrastructure. The proliferation of data within organizations has led to a growing need for scalable, flexible, and efficient data management solutions. Traditional centralized approaches often struggle to keep pace with the increasing volume, variety, and velocity of data. In response to these challenges, the concept of Data Mesh has emerged as a promising paradigm for organizing and scaling data infrastructure. This paper provides a comprehensive review of Data Mesh and its architecture, highlighting its key principles, components, and implications for modern data management.

Keywords— Data Mesh, Data Mesh Architecture

1. Introduction

The era of data is here, and it's glorious. According to projections made by IDC (International Data Corporation), by 2025, the quantity of digital data produced by businesses and individuals would have increased by 61%, reaching 175 zettabytes. While improvements in the analytic data environment might help businesses get value from their enormous stores of raw data, challenges in the integration, management, and governance of data at scale threaten to derail data analytics efforts. Currently[15], a centralized data management team is responsible for collecting, processing, and managing all domain-specific data. This core group is becoming the stumbling block to liberating valuable domain data. The responsibility for the domain's data must be transferred from the central team to the domain teams. Applying domain-oriented, self-service design and product thinking, a data mesh is developing as a revolutionary decentralized solution to manage data at scale. In 2019, Zhamak Dehghani was the first to coin the phrase "data mesh." Google Trends' search volume for "Data Mesh" over the past five years is seen in Figure 1. The rising trend line shows the rising popularity of data mesh. Despite data mesh's widespread use, there is a dearth of scholarly resources dedicated to the issue. But as more and more businesses investigate data mesh, we recognized an opening to conduct a systematic assessment of the grey literature in order to properly characterize the idea and create a research agenda.

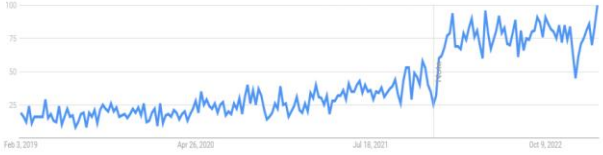


Figure: 1 Google trends for the word "Data Mesh"

To overcome the problems that might arise from using a centralized data architecture, Data Mesh proposes a decentralized method of data administration. It promotes delegating data ownership and management tasks to specialized groups or "domains" inside an organization so that subject matter experts may exercise autonomy over their data.

The collection, processing, and distribution of data to different parties within an organization have traditionally all been handled by a single central data team. While initially effective, this centralized method has become more inefficient as data volumes and complexity have grown. It can also make people too reliant on the central data team, which is bad for both flexibility and creativity.

Data Mesh suggests a change in perspective in which information is seen as a commodity and ownership of this commodity is shared among several departments or fields[16]. Each discipline will oversee its own data products, from gathering to storing to processing to delivering the information. The teams' domain specialists have a comprehensive grasp of the data and its context, allowing them to make more informed decisions and extract greater value. [3]

Organizations in today's data-driven world confront formidable obstacles in making effective use of the mountains of data they produce and amass. Centralized data management techniques have historically struggled to keep up with the exponential growth in data volume, variety, and velocity, which has resulted in problems including data silos, bottlenecks, and insufficient scalability. Considering these difficulties, a new paradigm for data management and Organization known as "Data Mesh" has evolved.

Data Mesh is a shift from the traditional, centralized data platforms used by many businesses. It promotes a decentralized strategy in which data ownership and

administration are assigned to certain groups or departments within an organization. Data Mesh frees specialists from needing to rely on a centralized data team or department, instead giving them the tools, they need to take charge of their own data and use it to their advantage. [5]

The core of Data Mesh is its architectural approach, which centers on a set of concepts and components that, when combined, allow for efficient data administration and application [17][18]. Data Mesh's design provides a foundation for managing and expanding an organization's data infrastructure in ways that foster teamwork, adaptability, and creativity.

This study aims to do just that by providing a thorough analysis of Data Mesh and its structure. The purpose of this article is to investigate [6] the fundamentals, components, implementation difficulties, and consequences of using Data Mesh in contemporary data management. Organizations and researchers may make more educated judgements about adopting Data Mesh and evaluating its potential advantages if they are familiar with its underlying architecture.

2. Why You May Need a Data Mesh

A data mesh is an innovative method of distributing and sharing information within a company. It's meant to solve problems that crop up when businesses double their data infrastructure and the amount of people using it. A data mesh might be useful for a few different reasons:

Data silos and fragmentation: Data silos and fragmentation are common issues that arise for expanding businesses. There may be inconsistencies in data quality and unnecessary duplication of work if different teams or business units used their own data systems and methods. A data mesh is a framework for connecting previously isolated data sources into a coherent whole**. [9]**

Agility and scalability: As data volume and diversity grow, traditional centralized data systems may become inefficient. In a data mesh, ownership and administration of data is shared between many groups or industries. When teams have more discretion over their own information, they can make decisions with more speed and flexibility.

Democratizing data access: In many companies, only a select group of data scientists or a central IT department has access to the company's data. This causes delays and hinders the efficiency of data-driven decision making. A data mesh promotes a self-serve approach in which data is viewed as a product and distributed to various groups via well-defined application programming interfaces (APIs), catalogues, and data markets. As a result, teams will have greater autonomy in terms of data access and use.

Data quality and ownership: Data quality and governance might be difficult to maintain using a conventional centralized method. Using a data mesh, you may delegate data quality oversight to certain groups or disciplines. The quality, correctness, and dependability of the data produced are the responsibility of the data product owners. This dispersed control increases transparency and enhanced data integrity.

Innovation and experimentation: Data meshes encourage a mindset that values exploration and discovery. Teams may try out new data sources, technologies, and analytical methods with the freedom that comes from decentralized data ownership. As a result, the company will benefit from increased chances for learning and discovery.

Collaboration and cross-functional teams: Collaborating across departments and forming specialized teams are both facilitated by a data mesh. Data engineers, data scientists, domain experts, and other stakeholders collaborate in these groups to find solutions to pressing business issues. A data mesh improves teamwork and problem-solving by connecting specialists from different fields.

Implementing a data mesh calls for forethought, organizational buy-in, and a mental shift. Despite its useful features, it may not be appropriate for your company or circumstance. Before adopting a data mesh strategy, it is crucial to evaluate your organization's unique requirements and objectives.

3. Standard Steps followed to setup Data Mesh Architecture

There is a specific order of operations that must be followed while establishing a data mesh architecture. The following are the typical procedures taken while establishing a data mesh, however the specifics may vary according on the needs of the organization:

Define the vision and goals: The first step in successfully deploying a data mesh inside your organization is to define its vision and objectives. In order to reach your goals, you must first pinpoint the problems you intend to fix and the obstacles you must overcome.

Assess current state: Think about how you're currently handling data, your current procedures, and your company's culture. Gain an awareness of the present data ecosystem, including data flows, data governance practices, and data source locations. Locate data islands and places where data sharing isn't well developed.

Identify domains and data products: Locate the many departments and functions inside your company that generate and utilize data. Set limits and establish roles

for each area. Determine whatever data products may be developed in each area to meet the needs of certain departments or teams. [12]

Form domain-oriented teams: Form interdisciplinary groups, or "squads," to work on certain projects. A team needs data engineers, data scientists, subject matter experts, and other stakeholders. The data products for their respective fields will be the responsibility of these groups.

Define data product ownership: Put people in charge of data products on each domain team. Owners of data products are liable for their goods' completeness, accessibility, and dependability. Their data products have contracts, SLAs, and application programming interfaces (APIs) defined by them.

Implement data infrastructure: Install the data backend that will sustain the data mesh design. Data storage, processing frameworks, integration tools, and governance processes all fall under this category. Consider using flexible and scalable cloud-based technology and cutting-edge data platforms. [15]

Establish data mesh governance: Develop the data mesh architecture's guiding concepts and policies. Set requirements for data integrity, confidentiality, and compliance. In order to provide self-service access to data products, it is necessary to implement systems for data discovery, cataloguing, and metadata management.

Enable self-serve access: Create data catalogues, data markets, or data portals so that users may find and acquire data goods independently. Each data product on these hubs should have comprehensive documentation, explanations, and usage instructions. Data consumers would benefit from the addition of data lineage and data quality measures to help them determine the data's trustworthiness.

Foster a data-driven culture: Encourage employees to make decisions based on facts. Foster an environment where teams work together, share information, and learn from one another. Make sure everyone has access to training and resources to improve their data literacy. Create iterative loops for refining data products and the data mesh itself.

Monitor and iterate: Keep an eye on data mesh metrics including consumption, impact, and performance. The design should be refined throughout time in response to user input and changing requirements for the data product. Make necessary improvements to the data mesh based on your ongoing evaluations of how well it's doing. [1]

It is important to keep in mind that a data mesh implementation is an iterative process. Realizing the full potential of a data mesh architecture calls for organizational dedication, cultural shift, and continuous cooperation.

4. When to use/not use Data Mesh

While a data mesh architecture can bring numerous benefits to organizations, it may not be the ideal solution for every situation. Here are some considerations for when to use or not use a data mesh**:**

Use Data Mesh:

Scalability: A data mesh can give a more scalable and agile solution to handling data if your organization is having trouble growing its data infrastructure due to increased data volume, diversity, or sources.

Data democratization: A data mesh may promote self-serve access to data products and build a culture of data responsibility, both of which are important if you want to empower multiple teams or business units to access and exploit data independently without relying on centralized data teams.

Domain-specific expertise: A data mesh empowers domain-oriented teams to control their data products, capitalizing on their experience and understanding of the topic, which is especially useful for Organizations with many domains or business activities that each have unique data needs.

Innovation and experimentation: A data mesh allows teams to try out new data sources, technologies, [5] and analytical methodologies inside their own domains, which is great for fostering a culture of innovation, exploration, and experimentation with data.

Do Not Use Data Mesh:

Small-scale data operations: The effort required to set up a data mesh may not be worth it if your company has a tiny data environment with few data sources and consumers. It's possible that a data architecture with less moving parts might be preferable.

Consistency in data governance across domains might be difficult to achieve in a data mesh if your company works in a highly regulated industry with stringent data governance rules. In these circumstances, centralization may be the best option.

Limited domain diversity: It's possible that a data mesh won't be an improvement over a more conventional, centralized data architecture for your business if you just operate in one domain or have a small number of domains with identical data needs.

Lack of organizational readiness: It may not be the best moment to explore a data mesh strategy if your organization lacks the buy-in, dedication, or resources to make the cultural and organizational adjustments essential for a data mesh deployment.

Whether or whether your company decides to implement a data mesh relies on its unique requirements, objectives, and environment. To assess if a data mesh is in line with your strategic objectives and can manage your data difficulties, it is important to analyze your organization's size, complexity, data environment, and cultural preparedness.

5. Difference between data mesh and data fabric

Data Mesh and Data Fabric are both architectural approaches aimed at addressing data management challenges within organizations, but they differ in their focus and implementation.

Data Mesh:

Data Mesh is geared at creating domain-specific teams that are accountable for their own data products to decentralize data ownership and management.

It stresses the need of viewing data as a product and allowing other departments or teams to access data on their own. [8]

Data Mesh promotes data democratization by putting the onus of responsibility for data quality, governance, and access in the hands of individual teams.

It encourages employees to work together, acquire specialized knowledge, and try new things in the workplace.

In most cases, a Data Mesh design will need the formation of interdisciplinary domain teams, the introduction of data catalogues or marketplaces, and the provision of well-defined data contracts and application programming interfaces for data products.

Data Fabric:

Data Fabric is an organization-wide strategy for managing data that emphasizes standardization and consolidation.

The result is supposed to be an uninterrupted data flow across different programmers and services.

To guarantee data consistency, quality, and accessibility, Data Fabric prioritizes data integration, data governance, and data orchestration.

Data virtualization, metadata management software, and integration platforms are common technologies used in this context. [7]

Rather of requiring data consumers to directly integrate with each data source, Data Fabric provides a unified data layer that hides the complexities of the underlying data sources.

In conclusion, Data Mesh facilitates data democratization by distributing data ownership and management to specialized teams. But Data Fabric brings everything together in one place, including data integration, governance, and orchestration. Organizational considerations, the nature of the data being managed, and the extent of decentralization or centralization sought all play a role in determining which of these two approaches to data management will be the best fit.

6. Why and Who needs Data Mesh

When businesses have special data management difficulties, the architectural solution known as "Data Mesh" might help. Data Mesh might be useful for the following people and situations.

Scalability: Data Mesh can help businesses who are having trouble expanding their data infrastructure because of rising data volume, diversity, or sources. More scalable and adaptable data operations are made possible by the decentralized model's distribution of data ownership and management across domain-oriented teams. [14]

Data Democratization: Data Mesh is useful for businesses that wish to provide separate departments or divisions access to and control over company data on their own. It encourages teams to access data products on their own, rather than waiting for help from a centralized data team.

Domain Expertise: Data Mesh is useful for companies that have many domains or business operations that need domain-specific skills in organizing and analyzing data. It empowers subject-matter experts to own the data products they create, capitalizing on their deep familiarity with the subject matter.

Innovation and Experimentation: Data Mesh is useful for businesses who wish to encourage a culture of data-driven innovation, discovery, and experimentation. It allows groups to try out different data sources, technologies, and analytical methodologies inside their own areas, increasing opportunities for growth and development.

Data-driven Culture: Data Mesh is useful for businesses that want to foster an environment where employees are encouraged to share information, work together, and learn from one another. Data ownership and responsibility are promoted, and cross-departmental cooperation is facilitated.

Data Mesh is not a cookie-cutter answer to every problem. It works well for companies who want to decentralize data ownership and management, have complicated data landscapes, and need domain-specific data knowledge. Before committing to a Data Mesh strategy, it is essential to evaluate the unique requirements, objectives, and Organizational preparedness.

7. Cost Constraints in Data Mesh

Implementing a Data Mesh architecture can involve costs, as it often requires investments in technology,

resources, and organizational changes. Here are some cost considerations to keep in mind when planning for a Data Mesh implementation**:**

Infrastructure Costs: Investments in cloud platforms, data storage, data processing, and networking may be needed to support Data Mesh. The storage, computation, and data transfer expenses that come with using cloud-based solutions are the price you pay for the benefits they provide. Based on data volume, processing needs, and expected expansion, businesses should calculate and budget for these infrastructure expenditures.

Technology Investments: Data catalogues, data integration platforms, data virtualization tools, and metadata management solutions are only few examples of data management technologies that may need to be implemented or upgraded in order to support a Data Mesh design. Don't forget to include in the price of training, customization, and regular maintenance, in addition to the price of licensing or subscription. [14]

Skills and Expertise: Developing and maintaining a Data Mesh architecture requires skilled resources. Hiring or upskilling domain-oriented teams, including data engineers, data scientists, and data product owners, can involve costs in terms of recruitment, training, and professional development programs.

Organizational Changes: Shifting to a Data Mesh model may require organizational changes and cultural adjustments. This could involve redefining roles and responsibilities, promoting collaboration and knowledge sharing, and establishing new processes and governance mechanisms. The costs associated with organizational change management, training programs, and communication efforts should be considered.

Data Governance and Security: Implementing a Data Mesh architecture requires robust data governance and security measures. Investing in data governance frameworks, privacy controls, compliance processes, and security technologies can add to the overall cost of implementing and maintaining a Data Mesh. [13]

Maintenance and Support: There will be a need for constant upkeep, support, and improvements after the Data Mesh design has been put into action. This entails keeping an eye on things, fine-tuning their performance, fixing any problems that crop up, and making any necessary upgrades or modifications. The continuing expenses of maintaining and managing the Data Mesh ecosystem should be factored into an organization's budget.

It's worth noting that Data Mesh expenses might differ widely from one business to the next, based on factors like size, complexity, and requirements. In order to make educated judgements regarding the expenditures necessary for a Data Mesh deployment, it is recommended to do a complete cost-benefit analysis, considering the possible long-term benefits of increased data accessibility, agility, and creativity.

8. Do/ Don’t in Data Mesh

When implementing a Data Mesh architecture, there are several best practices to consider (Do's) and some common pitfalls to avoid (Don'ts). Here are some Do's and Don'ts for Data Mesh**:**

Do's:

Foster a Culture of Collaboration**:** Encourage cross-functional collaboration, knowledge sharing, and learning across teams. Create an environment where different domains and teams work together to solve data challenges and drive innovation. [11]

Define Clear Data Contracts: Establish clear data contracts that define the responsibilities, expectations, and quality requirements for each data product. Clearly communicate the data contracts to data product owners and consumers to ensure a shared understanding.

Implement Self-Serve Data Access: Provide tools, platforms, and data catalogs that enable self-serve access to data products. Empower teams to discover, access, and utilize data independently, promoting data democratization and reducing dependencies on centralized data teams.

Emphasize Data Quality and Governance: Maintain a strong focus on data quality and governance. Encourage data product owners to take ownership of data quality, implement data validation mechanisms, and ensure adherence to data governance policies and standards.

Establish Data Observability and Monitoring**:** Implement mechanisms to monitor the performance, usage, and impact of data products. Collect relevant data metrics, establish monitoring dashboards, and enable feedback loops to continuously improve the quality and effectiveness of data products.

Don'ts:

Centralize Data Ownership: Avoid centralizing data ownership and management. The essence of Data Mesh is to distribute ownership and accountability to domain-oriented teams. Resisting the urge to centralize data control is crucial for a successful Data Mesh implementation.

Overlook Data Governance and Compliance: Don't neglect data governance, privacy, and compliance requirements. Ensure that appropriate measures are in place to safeguard sensitive data, adhere to relevant regulations, and maintain data privacy and security. [4]

Ignore Data Product Discovery: Don't underestimate the importance of data product discovery. Invest in

creating a comprehensive data catalog or marketplace that facilitates easy discovery and understanding of available data products across the organization.

Neglect Domain Expertise: Domain expertise is vital in Data Mesh. Don't overlook the importance of having domain-oriented teams with the right skills and knowledge. Encourage collaboration between domain experts, data engineers, and data scientists to drive effective data product development.

Forget Continuous Improvement: The success of a Data Mesh design depends on its ability to undergo constant refinement. Data Mesh should not be viewed as a one-and-done solution. Foster a mindset of continuous improvement and flexibility in response to changing market conditions and technology trends.

Keep in mind that a Data Mesh implementation calls for extensive preparation, buy-in from top management, and a shift in Organizational culture. It's crucial to analyze your company's unique requirements, consult with key players, and refine the Data Mesh design considering ongoing input and shifting priorities.

9. Standard Tech Stacks for implementing Data Mesh

Implementing a Data Mesh architecture involves selecting appropriate technologies that support the principles and requirements of Data Mesh. While there is no one-size-fits-all tech stack for Data Mesh, here are some common components and technologies that can be considered when **Implementing a Data Mesh:**

Data Integration: Data integration platforms play a crucial role in connecting and integrating data sources within a Data Mesh architecture. Technologies like Apache Kafka, Apache Nifi, or cloud-based solutions like AWS Glue or Google Cloud Dataflow can facilitate data movement, transformation, and streaming across systems.

Data Storage: Data Mesh often involves a distributed data storage approach. Depending on the specific requirements, technologies such as data lakes (e.g., Apache Hadoop, Amazon S3, Google Cloud Storage), data warehouses (e.g., Snowflake, Amazon Redshift, Google BigQuery), or NoSQL databases (e.g., Apache Cassandra, MongoDB) can be utilized for storing and managing data products.

Data Catalog and Metadata Management: Data catalogs and metadata management tools help organize and document data products and their associated metadata. Solutions like Collibra, Alation, Apache Atlas, or custom-built metadata repositories can provide a centralized source of information for discovering, understanding, and managing data assets.

Data Virtualization: Data virtualization platforms like Denodo or Cisco Data Virtualization can create a unified and abstracted view of data across different sources. They enable data consumers to access and analyze data from multiple systems without the need for direct integration, simplifying data access and reducing data duplication. [3]

Data Governance and Compliance: Data governance and compliance tools help enforce data policies, security measures, and compliance requirements within a Data Mesh architecture. Solutions such as Collibra, Informatica, or Apache Ranger can assist in establishing and managing data governance, access controls, and regulatory compliance.

Data Quality and Observability: Implementing data quality and observability measures is essential in a Data Mesh architecture. Tools like Apache Airflow, Prometheus, or custom monitoring solutions can help monitor data quality, capture data metrics, and provide insights into the performance and usage of data products.

Analytics and Data Science: Data analysis, machine learning, and advanced analytics can be performed on the data products within a Data Mesh using technologies like Apache Spark, Python libraries (e.g., Pandas, NumPy, scikit-learn), or cloud-based analytics platforms (e.g., AWS Athena, Google BigQuery ML).

The organization's infrastructure, current technology stack, talent capabilities, scalability needs, and budget are all key considerations when deciding which technologies to use. Data Mesh architecture evaluation and selection should be guided by the unique requirements and objectives of the implementing Organization.

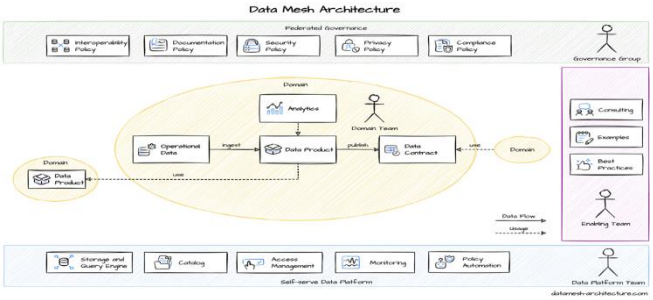
10. Standard Data Mesh Architecture

A data mesh is a distributed data architecture that gives data creators greater control by grouping data together according to business function, such as marketing, sales, customer support, and others. Producers are in a prime position to establish data governance principles by virtue of their familiarity with the domain data. This therefore opens the door for widespread self-service adoption inside an organization. Data lakes and data warehouses may still be used in conjunction with this federated strategy, even though they are less prone to the performance issues that plague centralized, monolithic systems. This simply indicates that their utility has evolved from a unified data hub to a smorgasbord of independent data stores. [2]

To expand and realize data management objectives, it is important to note that data mesh encourages the use of cloud native and cloud platform technologies. Microservices are often used as a comparison to clarify the concept's role in this context. However, not every company will benefit from a data mesh because the complexity of their enterprise data may not warrant its use. This distributed architecture is especially useful for scaling data needs across an organization. [11]

Domain teams can independently conduct cross-domain data analysis with the help of a decentralized data mesh

architecture. The domain, along with the accountable group and the operational and analytical data, is essential. In order to conduct their own analysis, members of the domain team consume operational data and construct analytical data models in the form of data products. It may also decide to offer data products accompanied by data contracts in order to meet the data demands of different domains.



Ref: https://www.datamesh-architecture.com/

Figure: 2 Basic Data Mesh Architecture

In a federated governance group, the domain team reaches consensus on global norms like interoperability, security, and documentation standards that help other domain teams learn how to find, analyze, and implement data mesh products. With the help of the data platform team's self-service, domain-agnostic data platform, domain teams may simply construct their own data products and conduct efficient analysis on their own. [6] When it comes to modelling analytical data, using the data platform, and creating and maintaining interoperable data products, a domain team will benefit greatly from the guidance of an enabling team.

Logical architecture: domain-oriented data and compute: The analytical data must be organized into domains, and we may construct such an architecture to encourage breakdown. In this design, the analytical data and operational capabilities of the domain are both exposed to the rest of the company via the domain's interface. The 'podcasts' domain, for instance, has both a set of operational APIs to "create a new podcast episode" and an analytical data endpoint to retrieve "all podcast episodes data over the last n> months." For domains to independently offer their analytical data and release the code that computes the data, the design must eliminate any connection or friction between them. The ability for domain teams to control the release and deployment of their operational and analytical data systems is crucial for a scalable architecture. Domain-oriented data ownership may be seen in action in the following scenario. The images are only symbols representing ideas, and they serve as examples. The goal isn't for them to be exhaustive.

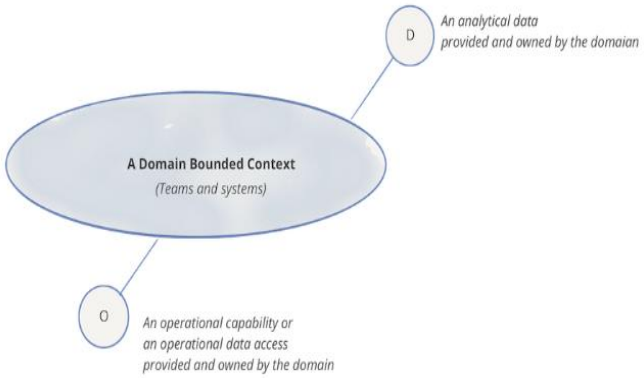


Figure: 3 Notation: domain, its analytical data and operational capabilities

Each domain may make available a single or many operational APIs and a single or multiple analytical data endpoints.
Each domain may rely on the analytical and operational data endpoints in other domains. To paint a picture of the demographics of podcast listeners, the 'podcasts' domain in the following example uses the analytical data of 'users updates' from the 'users' domain in order to create the 'Podcast listeners demographic' dataset.

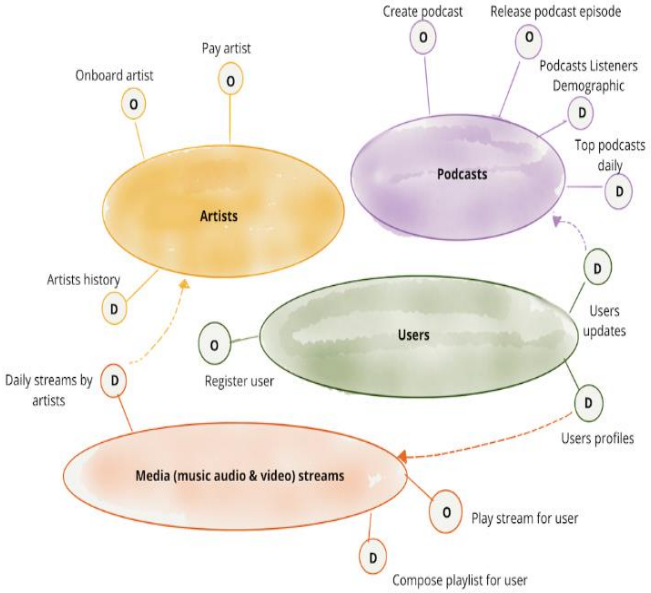


Figure: 4 Example of Domain Oriented Ownership of Analytical Data in Addition to Operational Capabilities

11. Benefits of a data mesh

Data democratization: With data mesh designs, non-technical users, not just data specialists, may access data from a variety of sources and build their own applications. Improved data discoverability and accessibility thanks to this domain-driven design cuts down on data silos and operational bottlenecks, allowing for quicker decision-making and freeing up technical

users to focus on activities that make greater use of their expertise. [15]

Cost efficiencies: This decentralized design abandons batch processing in favor of real-time data acquisition via cloud-based platforms and streaming pipelines. One further way in which cloud storage saves money is by enabling data teams to quickly deploy big clusters as required while paying for just the storage that is used. In other words, on a cloud data platform, you may simply buy more computing nodes if you need them to perform a job in a few hours rather than a few days. This also implies that storage costs can be seen more clearly, allowing engineering teams to allocate funds and resources more effectively.

Less technical debt: Technical debt increases with the complexity and necessity of teamwork inherent in a centralized data architecture. When more and more information is stored in one place, performance suffers. Data teams may better fulfil the needs of their data consumers and alleviate technical stresses on the storage system by dividing the data pipeline up by domain ownership. By providing APIs, they may make data more accessible and reduce the number of queries made to the server.

Interoperability: Data mesh models promote interoperability by having data owners pre-arrange how to standardize domain-agnostic data fields. By following these guidelines, cross-domain data connection may be performed efficiently while a domain team is still constructing their individual datasets. Field type, information, schema flags, and other similar categories are frequently standardized. Data consumers benefit from domain-wide consistency because it simplifies their interactions with APIs and allows them to build custom software tailored to their own business needs.

Security and compliance: By assisting in the enforcement of data standards for domain-agnostic data and access restrictions for sensitive data, data mesh architectures support better governance practices. This nature of this data ecosystem facilitates compliance with government rules like HIPPA limits by making data audits possible. Auditor visibility into which users are accessing which data and how often is embedded in a data mesh architecture through log and trace data.

12. Use cases of a data mesh

While distributed data mesh architectures are still gaining adoption, they're helping teams attain their goals of scalability for common big data use cases. These include:

Business intelligence dashboards: new initiatives frequently need specialized data views to be fully comprehended by teams. The increased accessibility of data provided by data mesh designs helps satisfy this requirement for customization and adaptability.

Automated virtual assistants: Chatbots are widely used by businesses as a supplement to human customer care representatives in contact centers. A distributed data architecture can make additional data assets available to these virtual agent systems in response to commonly requested inquiries that can touch on numerous datasets. [2]

Customer experience: Having access to customer information helps firms learn more about their consumers and create more tailored services for them. This is true in many different fields, from advertising to medicine.

Machine learning projects: Data scientists can save time by working with standardized domain-agnostic data as they piece together disparate datasets. This time can speed up the rate at which models enter a production setting, bringing the desired level of automation one step closer to reality.

13. Domain-Oriented Decentralized Ownership in Data Mesh

The idea of decentralized ownership of a domain is important to the design of a data mesh. Decentralized and focused on individual domains, data mesh is a method for creating and maintaining data infrastructure.
A centralized data architecture places ownership and management of data in the hands of a single group or department. As an organization expands and its data needs become more varied, this technique may become inefficient, slow, and unable to keep up.
A data mesh design, on the other hand, promotes data decentralization by encouraging domain teams and business units to take responsibility for their own data management. Data collection, storage, processing, and governance are now all the purview of the respective teams.
Each domain team retains control and responsibility for its own data assets under a decentralized ownership model that prioritizes domain specialization. When it comes to their data, they oversee determining its collection, storage, and usage within the broader data ecosystem.
Domain-oriented decentralized ownership in a data mesh architecture is based on the following tenets:

Data as a product: Each domain team should view its data assets as a valuable resource that it may share with other groups in the company, in accordance with the product mentality. The APIs, documentation, and quality criteria for these data items are all clearly established. [3]
Each domain team is given the autonomy to regulate its own data within the bounds of the stated Organizational rules, eliminating the need for a centralized governance

approach. They determine their own rules for data governance, data quality requirements, and security measures.

Data infrastructure as a platform: In a data mesh architecture, the data infrastructure is envisioned as a platform that gives domain teams access to self-service tools. Rather than relying solely on a centralized data team, teams may manage their data with the help of these technologies, frameworks, and services.

Data discovery and observability: To enable collaboration and data sharing across domains, there is a need for effective data discovery mechanisms. Metadata catalogs and data catalogs play a crucial role in facilitating the discovery and understanding of available data assets.

Data collaboration and community: A data mesh promotes collaboration and knowledge sharing among domain teams. Regular forums, communities of practice, and cross-functional teams can be established to foster collaboration and learning across domains.

By implementing domain-oriented decentralized ownership, organizations can achieve greater agility, scalability, and improved data autonomy. Each domain team has the flexibility to choose the most suitable tools, technologies, and processes for their specific needs, while still adhering to overall organizational data standards and guidelines.

14. Self-Serve Data Infrastructure in Data Mesh

Data mesh design relies heavily on self-service data infrastructure. As a result, domain teams are less reliant on a centralized data team for data-related activities and more capable of managing their own data assets. In a data mesh, domain teams may carry out a variety of data-related tasks thanks to the self-serve data infrastructure that is made available to them.

Here are some key aspects of self-serve data infrastructure in a data mesh:

Data storage and processing: With a self-service data architecture, domain teams may independently store and handle data. Technology such as distributed file systems, object storage, data lakes, and cloud-based storage services can help with this. The storage and processing infrastructures that a domain team use are completely up to them and their requirements.

Data pipelines and integration: The ability to construct and administer one's own data pipelines for data intake, transformation, and integration is a key feature of self-service data infrastructure. Tools and frameworks like Apache Kafka and Apache Airflow, as well as cloud-based data integration services, can help with this. Data pipeline procedures, source connections, and transformations may all be defined by individual domain teams to meet the needs of their unique domains. [9]

Data governance and security: Data governance and security features built into a self-service data infrastructure give domain teams the tools they need to safely and securely manage their data assets. This entails establishing parameters for data access, data privacy, and data retention. Domain teams might also put in place processes for monitoring and checking the quality of their data.

Data catalog and metadata management: A data catalogue or metadata management system is a common component of self-service data architecture, and it enables domain teams to catalogue and find their data assets. Metadata profiles, which include data schemas, data lineage, and usage standards, may be created by domain teams for their data products. Because of this, more people inside the company will be able to find and comprehend data.

Analytics and visualization: Data analytics and visualization activities may be carried out by domain teams with the help of self-service data infrastructure, which supplies them with the required tools and platforms. Data querying languages, exploratory tools, and integrations with well-liked analytics and visualization tools are all potential means of accomplishing this. With these tools at their disposal, domain teams can better understand their data and meet the analytics requirements unique to their field.

Monitoring and observability: Domain teams can monitor the status of their data assets in terms of availability, performance, and consumption thanks to the observability capabilities included into self-service data infrastructure. Data pipeline execution, data quality, and system performance may all be better understood with the use of logging, metric collecting, alerting systems, and dashboards.

The purpose of a data mesh's self-service data architecture is to lessen reliance on centralized data teams while encouraging domain teams to assume control of their data assets and autonomously make data-driven choices.

15. Data Mesh Implications for Modern Data Management

Data mesh has several implications for modern data management practices. Here are some key implications of data mesh on data management:

Decentralized data ownership: Domain teams or business divisions take responsibility for their own data in a data mesh. Data in their respective fields now fall within the purview of the respective teams. This requires moving away from a centralized data

management paradigm and towards one in which ownership, control, and decision-making power are dispersed throughout the Organization.

Domain-oriented data governance: In a data mesh, data governance is domain-oriented, with responsibility for enforcing regulations resting with individual domain teams. Standards for data quality, privacy laws, and other compliance measures also fall under this category. Decisions can be made by domain teams based on their unique information requirements and knowledge in their respective fields.

Data as a product mindset: Using data mesh encourages thinking of data as a commodity. This means that domain teams view their data assets as goods they sell to other teams in the company. They define APIs, documentation, and quality criteria for data products. This way of thinking motivates groups to put themselves in the shoes of data consumers and provide them with high-quality products.

Self-serve data infrastructure: When it comes to managing their data, domain teams may do so on their own with the help of the tools, platforms, and services made available by data mesh. Things like archiving, processing, integrating, analyzing, and visualizing data fall under this category. Without having to rely so much on a centralized data team, domain teams may gain more autonomy and control over the management of their data assets with the help of a self-service data infrastructure.

Data discovery and cataloging: The necessity of data finding and cataloguing is highlighted by data mesh. A data catalogue or metadata repository is kept by each domain team to record their data resources. This paves the way for other groups within the company to learn about and utilize the company's data resources. Data cataloguing is a vital tool for lowering data silos and increasing data sharing.

Collaboration and knowledge sharing: Collaborating and exchanging expertise across domain teams is facilitated by data mesh. Collaboration and knowledge sharing are encouraged by establishing cross-functional teams and holding regular forums for practitioners. This promotes an environment conducive to data sharing, wherein groups pool their knowledge and resources to make more informed decisions and innovate more rapidly.

Data observability and monitoring: The necessity of being able to observe and monitor data is emphasized by the data mesh. It is the responsibility of each domain team to keep an eye on the status of their data assets in terms of health, performance, and quality. Data quality checks, pipeline monitoring, and data dependability assurance are all part of this process. Observability ensures the reliability and availability of data by allowing teams to spot and fix problems rapidly. [14]

Data mesh, in its whole, revolutionizes the status quo of data management by dispersing data ownership, giving authority to domain teams, and encouraging a culture of data responsibility. Self-service features, domain-oriented governance, and a focus on products are all fostered, leading to more efficient and adaptable data management in today's businesses.

16. Example of real life implementation of Data Mesh

Spotify: Spotify is often cited as one of the pioneers of the data mesh approach. They implemented a decentralized data mesh model, where each squad (cross-functional teams) is responsible for managing their own data products and services. They emphasized data as a product and focused on enabling self-serve data infrastructure.

LinkedIn: LinkedIn implemented a federated data mesh model. They created a central data platform team responsible for providing common infrastructure and tools, while the data products and services were owned by individual teams. This approach allowed teams to have autonomy while ensuring data governance and consistency.

Zalando: Zalando, an e-commerce company, implemented a self-serve data infrastructure with a federated data mesh model. They established a centralized data platform team that built and maintained common data infrastructure components. They empowered individual product teams to take ownership of their data products and provided them with tools to manage and monitor their data.

ThoughtWorks: Thought Works, a global software consultancy, implemented a data mesh using the principles of domain-driven design. They emphasized domain-oriented decentralized data ownership, where each domain team was responsible for defining and managing their own data products. They focused on enabling cross-functional collaboration and providing data infrastructure support.

Netflix: Although Netflix does not explicitly follow the data mesh terminology, they have implemented similar principles. They embrace the concept of decentralized data ownership and treat data as a product. Different teams have ownership over specific datasets, and they provide self-serve tools and platforms for data discovery and consumption.

Conclusion

This paper concludes by summarizing the key insights and contributions of Data Mesh and its architecture. It emphasizes the significance of decentralization, domain-

oriented thinking, and self-serve data infrastructure in enabling organizations to effectively manage and leverage their data assets. The potential of Data Mesh to reshape data management practices and drive innovation is highlighted, providing a compelling argument for its adoption in the modern data landscape.

By providing a comprehensive review of Data Mesh and its architecture, this paper equips organizations and researchers with a deeper understanding of this emerging paradigm. It serves as a valuable resource for those interested in exploring the potential of Data Mesh to address the challenges of data management in the digital age.

References

[1] ABEL GOEDEGEBUURE, INDIKA KUMARA, STEFAN DRIESSEN "Data Mesh: a Systematic Gray Literature Review" arXiv:2304.01062v1 [cs.SE] 3 Apr 2023.

[2] Pushpender, Dr. Sohan Garg "A WIRELESS MESH NETWORK: ARCHITECTURE, ISSUES, APPLICATIONS" International Journal of Modern Computer Science and Applications 2013.

[3] Michele N. Lima, Helber W. da Silva, Aldri L. dos Santos "An Architecture for Survivable Mesh Networking" IEEE 2008.

[4] Sachin Lal Shrestha, Jinsung Lee, Anseok Lee, Kyunghan Lee "An Open Wireless Mesh Testbed Architecture with Data Collection and Software Distribution Platform" IEEE 2005.

[5] Syed Yasmeen Shahdad, Asfia Sabahath, Reshma Parveez "Architecture, Issues and Challenges of Wireless Mesh Network" International Conference on Communication and Signal Processing 2016

[6] Xin Fenga , Yuanzhen Wanga , Yanlin Weng "Compact Combinatorial Maps: a Volume Mesh Data Structure" Preprint submitted to Graphical Models 2012.

[7] S. Pena Serna, A. Stork and D. W. Fellner "Considerations toward a Dynamic Mesh Data Structure" SIGRAD 2011.

[8] L. Rodríguez and I. Navazo and Á. Vinacua "Data-driven Tetrahedral Mesh Subdivision" Ibero-American Symposium on Computer Graphics 2006.

[9] Tatiana Delgado Fernández "ARQUITECTURA DE REFERENCIA DE ECOSISTEMAS DE DATOS BASADA EN DATA MESH & DATA FABRIC" Cuban Journal of Public and Business Administration 2022.

[10] Zied Marrakchi, Hayder Mrabet, Christian Masson and Habib Mehrez "Efficient Mesh of Tree Interconnect for FPGA Architecture" IEEE 2008.

[11] Emerole Kelechi C, Nwadike Stanley, Nwogu Uchenna "Evaluation of Network Architecture and Its Implication on Connectivity and Data Security" International Journal on Recent and Innovation Trends in Computing and Communication 2014.

[12] Yuan Ye "A Data Extraction Algorithm for Huge Data Visualization Based on ComputationalMeshes" International Conference on Management of e-Commerce and e-Government 2014.

[13] Mohammad Nassiri, Fabrice Theoleyre, Martin Heusse "Molecular Architecture for Autonomic Wireless Mesh Networks" CoNEXT 2007.

[14] Mukhtiar Bano, Amir Qayyum, Rao Naveed Bin Rais "Soft-Mesh: A Robust Routing Architecture for Hybrid SDN and Wireless Mesh Networks" IEEE 2021.

[15] Ashish Dibouliya "Review on: Modern Data Warehouse & how is it accelerating digital transformation" International Journal of Advance Research, Ideas and Innovations in Technology 2023.

[16] Ashish Dibouliya, Dr. Varsha Jotwani "A REVIEW ON: A HYBRID SMART IOT BASED REAL TIME ENVIRONMENT MONITORING SYSTEM" A Review On: A Hybrid Smart IoT based Real Time Environment Monitoring System 2023.

[17] Authors: Ashish Dibouliya, Dr. Varsha Jotwani "The Transformational Impact of Modern Data Warehousing on the Banking System" IJRASET 2023.

[18] Ashish Dibouliya "Design of a Hybrid Authentication Technique for User and Device Authentication for an integrated IOT environment using Blockchain Authentication and Artificial Intelligence" IRJET 2022.