TECH
mahindra

Understanding The Right Fit for Your Organization: Data Fabric or Data Mesh?

Abstract

The key objective of setting up data mesh or data fabric architecture is to enable the availability of quality data in a timely fashion to the right people in the right format.

A data fabric is an architecture framework and a set of data services that provide frictionless data capabilities across a choice of endpoint applications or services spanning hybrid or multi-cloud and on-premises, by using rich metadata foundation and artificial intelligence/machine learning (AI/ML) automation.

Data mesh essentially refers to the concept of breaking down data lakes and siloes into smaller data domain-specific sets with a self-serve design, to enable data-driven decisions using data products and providing scalable delivery of data with a flexible data governance model.

Key Takeaways

Introduction

The difficulty of managing data in a distributed environment architecture makes adoption of modern data management solutions inevitable for leading organizations. Data mesh and data fabric are two such modern architecture concepts that help in achieving scalable and reliable modern data architecture.

While data fabric enables a single, virtual management layer at the top of heterogeneous distributed data applications, data mesh enables data products for individual groups of consumers to manage data with global and federated governance processes.

If we look closer into these two modern concepts, the major distinguishing factor is how the overall data and governance is managed. In the case of data fabric, it is centralized (consolidation centric) compared to a decentralized approach for data mesh (end consumer centric).

Challenges Prompting the Need for Data Mesh or Data Fabric

Multi-platform, multiple formats and applications, varied and adaptive governance, and reporting needs necessitates organizations to adopt data fabric or data mesh based solutions but these come with many challenges as well.

Some of the key business challenges include the following:

Effective insights from huge volumes of telemetry and internet of things (IoT) data to improve customer experience and retention
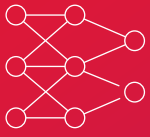
Managing regional compliance and security standards for multinational organizations specifically in BFSI, retail, and telecommunication sector

Enabling advanced analytics for each stakeholder group as per individual use cases

Availability of timely, consistent and accurate data for robust and reliable forecasting analytics

A few noteworthy implementation challenges are as shown below:

DATA MESH	DATA FABRIC
Optimized Data Domain and Product Defining independent data products and reducing data duplicity across data products.	**Consolidation and Data Availability** Identification of all sources across enterprise with varied data type and integration framework
Federated Data Governance Strategy Identifying independent business and data processes across different communities.	**Centralized Data Governance Strategy** Bringing enterprise-wide data and business process for identifying data governance polices and organization structure along with change management.
Infrastructure Planning Planning optimised infrastructure usages across enterprise to optimise cost and licensing.	**Performance Optimization** Optimizing performances for data ingestion and sharing for organization with high data growth

Use Case Trends

In this age of technology and business transformation fueled by competition, growth and business model changes post the pandemic, it has become imperative that organizations upscale their enterprise architecture to enable consistent, secured, reliable, and timely data to all stakeholders (internal and external) for analytical needs.

In addition, enabling data platforms such as CDP and Product 360 or for initiatives around cloud modernization, it needs to be ensured that data is made available across the enterprise into a central governed platform. In this context, data fabric and data mesh have become the most desirable enterprise architecture models and listed below are specific popular use cases that are driving these architecture pattern needs.

DATA MESH	DATA FABRIC
Availability of Telemetry Data and Analytics Automatic recording and transmission of data from remote or inaccessible sources to an IT system in a different location for monitoring and analysis. For e.g.: Network monitoring, equipment, and sensor data	**Regional Compliance and Regulations** Establish individual data products for regional compliance and operational data(multilingual) For e.g.: Multinational retail organizations compliance strategy (CPG, finance)
Availability of Telemetry Data and Analytics Automatic recording and transmission of data from remote or inaccessible sources to an IT system in a different location for monitoring and analysis. For e.g.: Network monitoring, equipment, and sensor data	**Manage Excessive Data Growth** Managing operational analytical data for large organization with a very high data For e.g.: data from sensors and equipment, Traffic data analysis for large cities.
Analytics for Multi Cloud and Platform Establish data fabric and data mesh for hybrid and multi cloud environments For e.g.: Enterprise Data Warehouse consolidation	**Sales and Marketing** Demographic based sales and demand forecasting For e.g.: usages reports for video streaming organizations Customer behavior analysis for retail
Data Marketplace Provide data as a service from a curated data platform on premise (big data) or on cloud (data lake) For e.g.: Customer data platform (CDP) , data as a service – for enterprise services (Operations and analytics)	**Federated Enterprise Governance** Federated data governance for key data elements such as customers and products to enable independent data and business process. For e.g.: CDP for banking customers - B2B and B2C, inventory management for individual regions for global manufacturers

Data Mesh Solution Approach and Benefits

Data Mesh essentially refers to the concept of breaking down enterprise data into smaller data domain specific data sets with a self-serve design, to enable data driven decisions using data products and providing scalable delivery of data with flexible data governance model. The key success factors and solution enablers for realizing the best

benefits of a data mesh architecture includes enabling data product as a service, creating efficient data products with least data duplicity, metadata based flexible integration framework with a global gateway, focused automation, reliable and consistent data with self-healing capability, and a self-serve platform for stewardship and advanced analytics. The implementation journey starts with readiness assessment and could be summarized in a 9-step approach as shown below:

Data Mesh Solution Approach

Data Security
Metadata Management

#	
1	**Readiness Assessment** Data maturity level, data sources/domains, and governance framework
2	**Data Product Assessment** Feasibility, cost vs ROI, and KPI
3	**Implementation Methodology** Agile/Waterfall methodology, plan, CRP/demo-based checkpoints, Data-Ops
4	**Architecture Road Map** End state definition, transition state definition, adaptive global source/subscriber/third party onboarding
5	**Design of Data Product** Code, infrastructure, data and metadata, modelling (including DP relationships)
6	**Federated Data Governance:** Governance and standards, automation and computation, domain specific governance
7	**Data and Integration Infra** Data code, data storage, data processing optimization plan, and data growth computation
8	**Visualization** Global/data validation and quality rules, databases and file data extraction, domain specific, and global UI platform
9	**AI/ML Platform** Adaptive and augmented governance and metadata management,

Visualization and Monitoring

The key benefits of a data mesh are as follows:

- ▶ Domain-oriented data owners and pipelines
- ▶ Data-as-a-product
- ▶ Self-serve data infrastructure as a platform
- ▶ Federated computational governance
- ▶ Interoperability and standardization of communications
- ▶ AI/ML based data pipelines and data quality transformation rules

Data Fabric Solution Approach and Benefits

Data fabric is an architecture that provides frictionless data capabilities across a choice of endpoint applications/services spanning across hybrid/multi cloud and on-premise. It uses rich metadata foundation and AI/ML automation, and has key success factors for realizing the best benefits of a data fabric architecture. These benefits include enabling desktop-as-a-service (DaaS), establishing metadata based global CDIF, centralized governance across all CDE, AI/ML based automation, timely, reliable, and consistent data with self-healing capability, self-serve platform for stewardship, and advanced analytics. The implementation journey is shown as below:

Data Fabric Solution Approach

Data Security
Metadata Management

1	**Readiness Assessment** Data maturity level, data sources/domains, and governance framework
2	**Unified Ingestion Process** Global and reusable ingestion framework
3	**Data Quality Standardization and Transformation** Business rules applied on incoming data to perform first level validation and map into a common data model
4	**Consolidation** Data Consolidation from multiple sources and create data lineage
5	**Data Cataloging** Segregate data per business/functional need
6	**Global Data Governance:** Apply global data governance rules, manage and govern master and reference data
7	**Data Services** Provide multi platform multi format data services based on business use cases
8	**Reporting and Analytics** Provide automation for predictable processes; e.g., metadata management, DQ processing
9	**AI/ML Augmentation** Provide automation for predictable processes; e.g., metadata management, DQ processing

Visualization and Monitoring

Key Benefits of Data Fabric

- ▶ Enables self-service data consumption and collaboration

- ▶ Automate data engineering tasks and augment data integration across hybrid, cloud, and on-premises resources

- ▶ Automate governance, protection, and security; enabled by active metadata graph

- ▶ Accelerate insight delivery with a single view of all relevant information across your enterprise

Best Practices to Choose Data Mesh or Data Fabric for an Organization

When considering the pros and cons, it's important to keep in mind that data mesh and data fabric are concepts and not technologies and aren't mutually exclusive. An organization can adopt both a data mesh and data fabric approach across partial, or all, departments as appropriate.

Data fabric and data mesh both provide an architecture to access data across multiple technologies and platforms, but a data fabric is organization-centric, architectural driven while a data mesh is technology centric and data domain driven.

The data fabric supports simple and centralized data sharing by principle whereas for data mesh the data sharing is complex and use case based.

The following table provides insights into some of the key factors that can help decide between data fabric and data mesh architecture:

Comparison	Data Fabric	Data Mesh
Data Volume and Domains	Medium to High	High to Very High
Governance	Centralized, Medium-High Complexity	Decentralized, High to Very High Complexity
Data Maturity	Medium-High	Low-Medium
Data Complexity	Low-Medium	High-Very High
Approach	Organization Centric, Architecture Driven	Technology Centric, Data Domain Driven
Cost/Budget	Low-Medium	Medium-High
Self-Serve Capability	Low-Medium Centralized	Medium-High Federated
Skill Demand	Low-Medium	Medium-High
Data Sharing	Simple and Centralized	Complex and Use Case Based
Infrastructure and Enterprise Architecture	Few Technologies, Centrally Managed	Distributed Architecture, Federated Management

Best Practices

DATA MESH	DATA FABRIC
Use Case Based Data Products Create consumption-oriented use case-based data products to optimize data usage and reduce data duplicity with an eye on performance and scalability.	**Meta Data based Common Ingestion Framework** Establish common ingestion framework based on centralized data model (consolidation based) using meta data graph.
Hybrid and Agile Data Governance Federated governance across data products with a touch of centralized data governance for key areas such as data security, data quality, and data ingestion framework.	**Modular Governance Approach** Identify data policies based on critical data elements and business processes.
MDM and DQ as a Service Provide master data management and data quality processes as a centralized services to maintain data consistency across enterprise.	**Scalable Infra Planning and Monitoring** Scalable infra planning based on data and processing or compute growth rate enabled by dashboard reporting through comprehensive monitoring.

Making the Right Choice for Your Business

Technically data fabric advocates a centrally managed and centrally governed system for organization whereas data mesh takes the use-case based data product approach thus enabling the data management with individual data product owners.

In summary, data fabric and data mesh will remain equally relevant for the foreseeable future. While each involves different elements, they are towards the same goal of bringing modern and advanced analytics to an organization while managing a wide range of growing data and infrastructure.

Authors

Saurabh Jha

SVP and Global Head – Data and Analytics, Tech Mahindra

Saurabh Jha is currently serving as the SVP & Global Head for data and analytics competency at Tech Mahindra. He has 20 years of global experience spanning across technology industry and management consulting. Saurabh is an MBA from the prestigious IIM Bangalore and holds a degree in Bachelor of Engineering from NIT, Bhopal.

Rajeev Dubey

Rajeev Dubey is the CoE Head for data management in data and analytics competency at Tech Mahindra. Rajeev has over 15 years of experience in the domain of data analytics across verticals.

TECH mahindra