

TECHNICAL WHITE PAPER: March 2024



Data Mesh as Your Data Layer

Proposing a
Data Mesh to Data Fabric Architecture

Table of Contents

Executive Summary	3
Problem Statement	3
Initial Solution: Implement A Data Mesh Architecture	4
Maturing Solution: Organization-Wide Data Fabric	5
Next Steps	6
Summary	7

Executive Summary

When organizations have a decentralized nature, meaning several components within them act with some degree of autonomy, implementing traditional organization-wide data governance becomes challenging. Looking toward more modern architectures, such as a data mesh and/or a data fabric, will accelerate organizational goals relating to data and its governance, as well as provide a flexible, modern, data layer supporting dynamic, data-driven mission operations and enabling decision advantage. To achieve near real-time, machine-augmented decision making across multiple operational domains, data must be delivered to the right user or machine with seamless consistency across multiple networks, classification levels, and infrastructures. This consistent data exchange must extend to on-premises data centers and operations centers, as well as thousands of tactical edge devices distributed across the globe, in addition to the multiple Cloud Service Providers (CSPs). VMware stands ready to assist your organization in achieving these data-driven mission imperatives.

Problem Statement

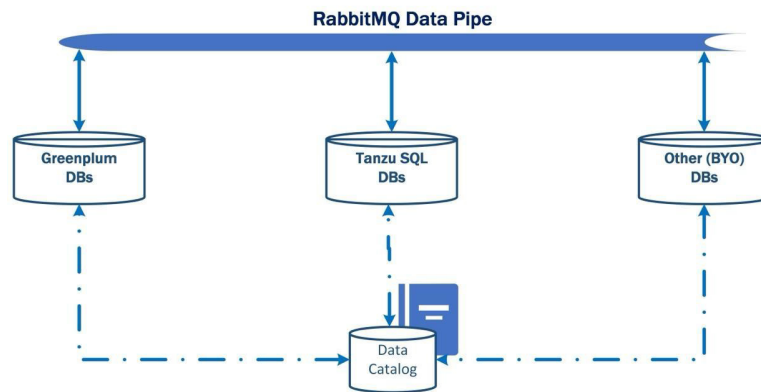
Your organization faces several challenges as it relates to data, both in terms of existing organizational data needs, as well as new data it acquires and generates. Some of those challenges may include the following:

- Data models are certain to be disparate across organizational components, with non-uniform entity and attribute names, data types, and structures.
- Within the public sector, defense organizations are exploring novel methods to connect and share data, specifically in an impromptu manner. The adage of combat is “aggregate to mass effects and disperse to survive.” Historically, this behavior was designed, if not discretely planned. But tomorrow’s battlefield requires more agile options which will require a degree of preplanning for how data must come together.
- Some components may lack any effective data governance, allowing data owners to “roll their own” in terms of defining both data security requirements and sharing procedures.
- Organizational components have varying motivational levels to perform the non-trivial task of remodeling and adjusting to data governance at a higher organizational level. Further, the timelines for doing so would not support mission needs, as the static nature of such a top-down data structure would not support the dynamic nature of evolving mission needs.
- Often when organizations are formed from several other preexisting organizations, such as those supporting joint or coalition mission operations, there are existing data holdings and processes by which data is shared and governed. In other words, these organizational components maintain data sovereignty and have their own, often incompatible, policies on data security. It is almost certain these processes, definitions, and access patterns will not be uniform.

Due to these challenges, as well as others that might be uncovered as new data architectures are defined, a decentralized architecture is a logical choice. But how does such an architecture still allow interoperability between diverse organizations with disparate data requirements and architectures?

Initial Solution: Implement A Data Mesh Architecture

A data mesh is a modern data architecture that is more decentralized in its implementation and its governance. For an introductory article on what a data mesh is, refer to [Why Data Mesh?](#)¹ by Gurnam Madan.



At its core, a data mesh is where individual data producers, such as business units or mission units, inside (and potentially outside) an organization keep ownership and governance over their own data holdings. These units come together, defining entities (a Personnel entity, or a System Element entity) which two or more of them would have in common. For data sharing, a high-bandwidth conduit between domains (units) is established. [Data Contracts: The Mesh Glue](#)² is an excellent article that provides additional information regarding this collaborative process.

Note that there are several non-trivial [organizational measures](#)³ that are also necessary with a data mesh architecture.

Implementing a data mesh provides some advantages for your organization:

- Existing departments within your organization can begin sharing their data in many instances without the need to significantly re-model elements, and modeling efforts will naturally concentrate on those higher-priority elements which must be shared by many parts of the organization.
- Data governance processes will be collaboratively developed by many distributed stakeholders, including those operating autonomous devices, creating a natural laboratory environment by which organization-wide governance processes can be defined and refined.

The following capabilities become available or improve significantly with a data mesh architecture :

- **Formalized business intelligence (BI) and reporting capabilities** – Each domain within the organization has access to more data, and that data is well-documented within the data catalog. These definitions and specified data formats allow integration with data from within that organizational component, allowing for a much wider array of reporting and dashboards. Since the broader organization does not specify tooling or capabilities, each domain is free to develop just the level of reporting it needs and no more. Since domains are working together on shared data governance, the sharing of expertise and experiences regarding such necessary services as BI and reporting occurs organically within organizations using a data mesh.

1. Madan, Gurnam. "Why Data Mesh?" Medium. October 26, 2022. medium.com/wesfarmers-onedigital/why-data-mesh-e6b85bb09dfa

2. Velasco, Luis. "Data Contracts: The Mesh Glue." Medium. July 23, 2022. towardsdatascience.com/data-contracts-the-mesh-glue-c1b533e2a664

3. Ciraci, Ugo. "Shaping Organizational Culture to Promote a Data Mesh Initiative." Medium. November 27, 2022. medium.com/@ugociraci/shaping-organizational-culture-to-promote-a-data-mesh-initiative-cd68ba95a33e

- **Predictive and advanced analytics** – With the broader availability of data from elsewhere in the organization, domains begin to have greater context for the data they routinely use. This enables deeper insights into how the domain is meeting its mission goals, and how it can improve its processes. Creation of advanced analytics products by data science professionals becomes increasingly possible, as well as desired. These advanced, and even predictive, analytics are the stepping stone into the creation of more advanced AI/ML models.
- **Artificial intelligence (AI) models and applications** – Domains find significantly more training data available for the creation of AI/ML models, allowing for the development of capabilities not previously available. As organizations rush to implement AI assistance for mission tasks, having more data available from related organizational components allows for better-trained and tuned models, whether they be traditional ML capabilities like data classification, or much more advanced ones like large language models (LLMs) to provide ChatGPT-like functionality.

Maturing Solution: Organization-Wide Data Fabric

As a collaborative culture of distributed governance and common data practices in the organization matures, subordinate components become more integrated into the broader data mesh and a more integrated architecture becomes possible: the data fabric. [Data Mesh vs Data Fabric](#)⁴ by Christian Lauer provides an explanation of the differences between a data mesh and a data fabric. At its core, a data fabric is a maturing of the mesh architecture. As the organization learns more about what data is essential to which business processes or mission process and how data moves from one place to another and between organizational elements, structures become more formalized and permanent. This facilitates the following higher-level data capabilities:

- **Bringing data from the edge** – Domains using significant and data-intensive edge devices, such as sensors and drones, will create mechanisms for such data to be efficiently ingested, and then for the data to be converted promptly into publishable data formats. This allows data products to be available across the mesh as quickly as possible. As the mesh matures into a data fabric, these mechanisms mature, become more permanent, and become directly accessible by other domains within the organization.
- **Proactive security analytics** – Cybersecurity within government organizations has primarily been reactive to incidents, with analytics designed or intended to discover what happened, improve the security posture to close the gap, and to prosecute offenders if possible. Contrastingly, proactive security analytics discover suspicious activity before a security incident happens, with the intention of avoiding the incident entirely. A key approach to proactive cybersecurity is anomaly detection, which requires an understanding of steady state behaviors. These behaviors are best characterized by data exchange and process execution. Having broader access to data from across the organization, centralizing relevant data entities within one owning domain, and implementing best practices from one domain into others within the organization all aid in making security more proactive and forward-looking in nature. Further, combining proactive analytics with machine learning enhances the speed and efficacy of the cybersecurity posture needed to combat dynamic threat actors.
- **Military use of decision-point tactics** – A model use case for consideration is the military's regimented approach to warfighting. This will facilitate a more defined pathway by which to share data, centered on military planning (course of action development) and optimized for the operational processes that are the heartbeat of these organizations. Mission operations enhanced by collaborative data sharing ecosystems drives *decision advantage*, the holy grail of future multi-domain operations. This becomes increasingly important as the data mesh matures into a data fabric.

4. Lauer, Christian. "Data Mesh vs. Data Fabric." Medium. November 3, 2022. medium.com/geekculture/data-mesh-vs-data-fabric-f396916865bc

Next Steps

VMware has products and expertise to help your organization adopt the culture and practices that align with implementing modern data architectures. It should be noted there really is no such thing as a *data fabric in a box*, or an *off-the-shelf data mesh*. These architectures are built through organizational enablement and the thoughtful integration of many products aligned to mission outcomes. Hence, the evolution toward enterprise data meshes that retain architectural fluidity and fully focus on iterative mission needs requires enabling disaggregated data owners to evolve their data services and collaborate within a framework and culture that is collaboratively established by all relevant stakeholders. The focus on organizational enablement in data practices that scale cannot be overstated.

The VMware® data portfolio has the elements necessary to construct either a data mesh or a data fabric:

- [VMware Greenplum®⁵](#) provides a scalable, robust platform for organizational elements with significant data holdings, such as those used in intelligence processing, who have immediate AI/ML, BI, or other analytics requirements.
- [VMware SQL™ and Postgres⁶](#) provide open-source database platforms with the flexibility to support other applications among government organizations.
- [VMware GemFire®⁷](#) offers a scalable, performant key-value store for use cases where a NoSQL solution is preferred. GemFire is also specifically suited to requirements for data replication across geographically dispersed areas.
- [VMware RabbitMQ®⁸](#) offers a resilient, scalable messaging platform to ease data movement and other messaging at enterprise scale.

In addition to a technical portfolio with all the necessary components, the VMware Data Team can implement these architectures and systems in the mission's context. The modernized approach to data is an organizational shift that must be addressed in the people, process, and technology framework. With the implementation of a modern data mesh or data fabric, embedding the right mix of data scientists, data engineers, operators, analysts, and business advocates is important for the success of the modern platform. With this paradigm shift of a modern address to data, continued operational success is just as important as the technology. The prioritization and operationalization of information—with prescribed outcomes—grows and changes with the mission; thus the team surrounding the technology must function in a flexible manner in order to deliver. At VMware, our team has the background and experience in organizational development necessary to enable you to set up and refine the governance processes your team will own and use for the long haul. Our teams include veterans, former civil servants, and others that have served in the public sector and understand the processes needed for success.

5. tanzu.vmware.com/greenplum

6. tanzu.vmware.com/postgres

7. tanzu.vmware.com/gemfire

8. tanzu.vmware.com/rabbitmq

In implementing a data mesh, the following steps are necessary to set up a minimum viable product (MVP):

1. Define the initial domains (organizational components) involved in the data mesh and the criteria and process for additional domains to join after the data mesh is established. Each domain should name the persons who will speak for them at the necessary governance forums which are emerging and begin figuring out which data elements will be its products to share within the mesh.
2. Set up a central Data Platform Team to serve across the data mesh. This team is responsible for implementing and maintaining the data catalog. This team will also set up pipelines for data exchange between domains, likely leveraging RabbitMQ as a primary tool. A best practice for organizations with a data mesh architecture is for the Data Platform Team to explore solutions as reference architectures. Domains may choose to adapt these reference architectures to their specific needs as they see fit, but are not mandated to do so.
3. [Necessary processes and organizational structures](#)⁹ are set up to supply the organization's data governance. These organizational structures are created from the representatives selected by each domain, and such personnel should have the requisite skill set to 'own' and support the equities of their domain's data needs. For example, members of data security governance forums should be made up of representatives for each domain who are knowledgeable and have skills relating to data security. These collaborative forums establish [shared requirements](#)¹⁰ for each domain to publish and share its data products across the data mesh.

Summary

This overview of data mesh concepts and the data practice enablement organizations use to employ them is intended to help you think through your data modernization journey and illustrate the potential value to your mission. VMware stands ready to help your organization design and develop more modern data practices, culture, architecture, and environments. Please reach out for an introductory discussion or whiteboard session facilitating enablement in this area.

For further information on how to get started, contact us: pubsecdata.pdl@broadcom.com.

9. Oh, Patrick. "The Synergy of Policies: Towards Effective Data Governance." Medium. October 20, 2023. medium.com/@patrick-oh-sglion65/the-synergy-of-policies-towards-effective-data-governance-e8f0a771d981

10. Tyagi, Deepanshu. "Data Governance: Best Practices for Managing Data as a Strategic Asset." Medium. March 8, 2023. blog.brilliantprogrammer.com/data-governance-best-practices-for-managing-data-as-a-strategic-asset-297baf88134

