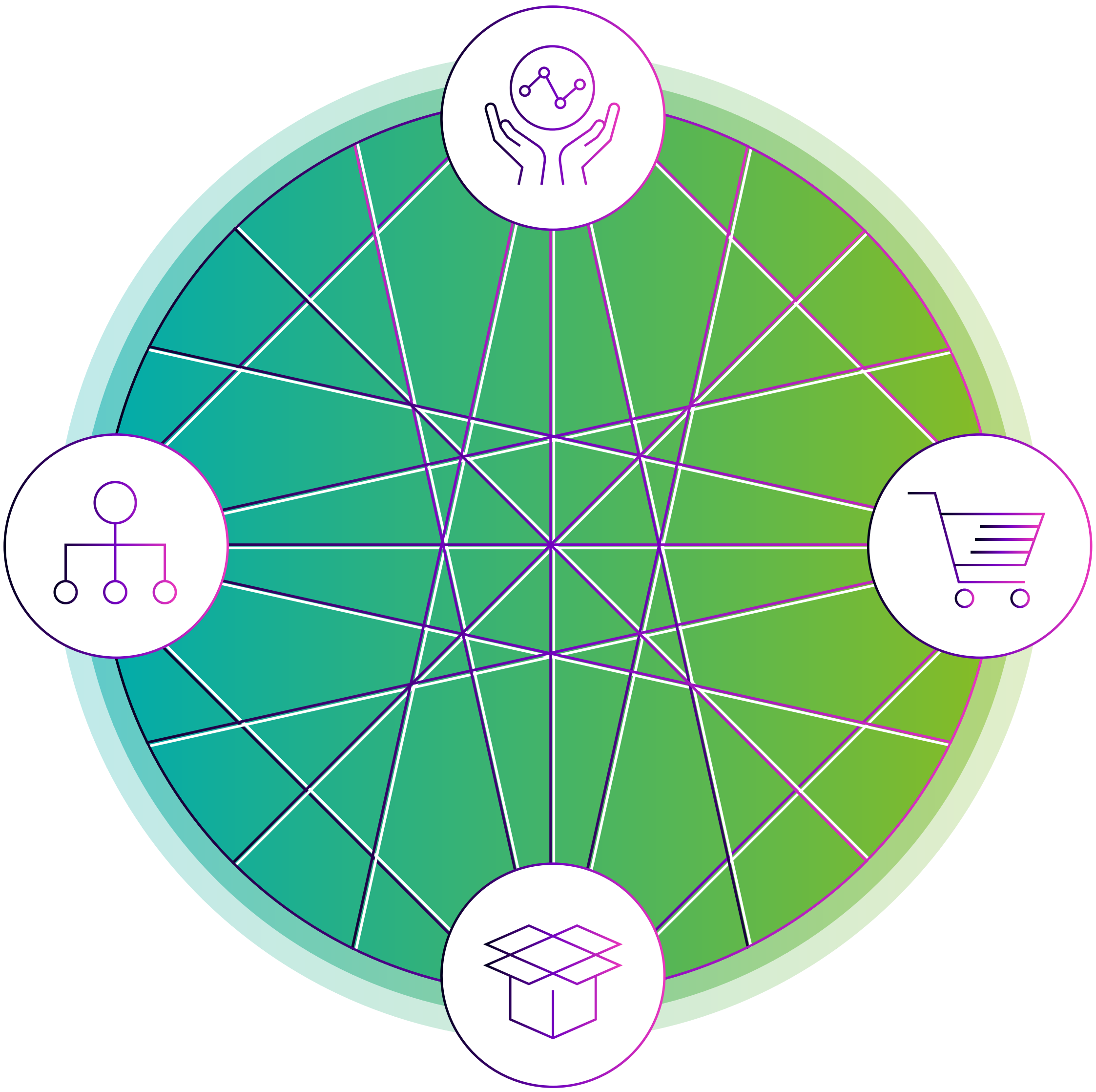


Treating data as a product in the era of GenAI

Break up the data monolith and instill a product mindset to use data more effectively across your organization

Contents

Data mesh and data as a product

As we discussed in our last article, [“Are you ready to ‘mesh’ your data?”](), many organizations are adopting data mesh, a federated data architecture that allows them to meet the growing demand of data across the enterprise.

In the previous article, we presented the evolution of enterprise data platform’s architecture along with the driving forces behind data mesh. We introduced the four key principles of data mesh and discussed how organizations consider data mesh as their choice to organize enterprise data, as well as the key considerations to take into account across four dimensions: business values, data and technology, talent and culture, and data management and governance.

Data mesh architecture is based on four key principles:

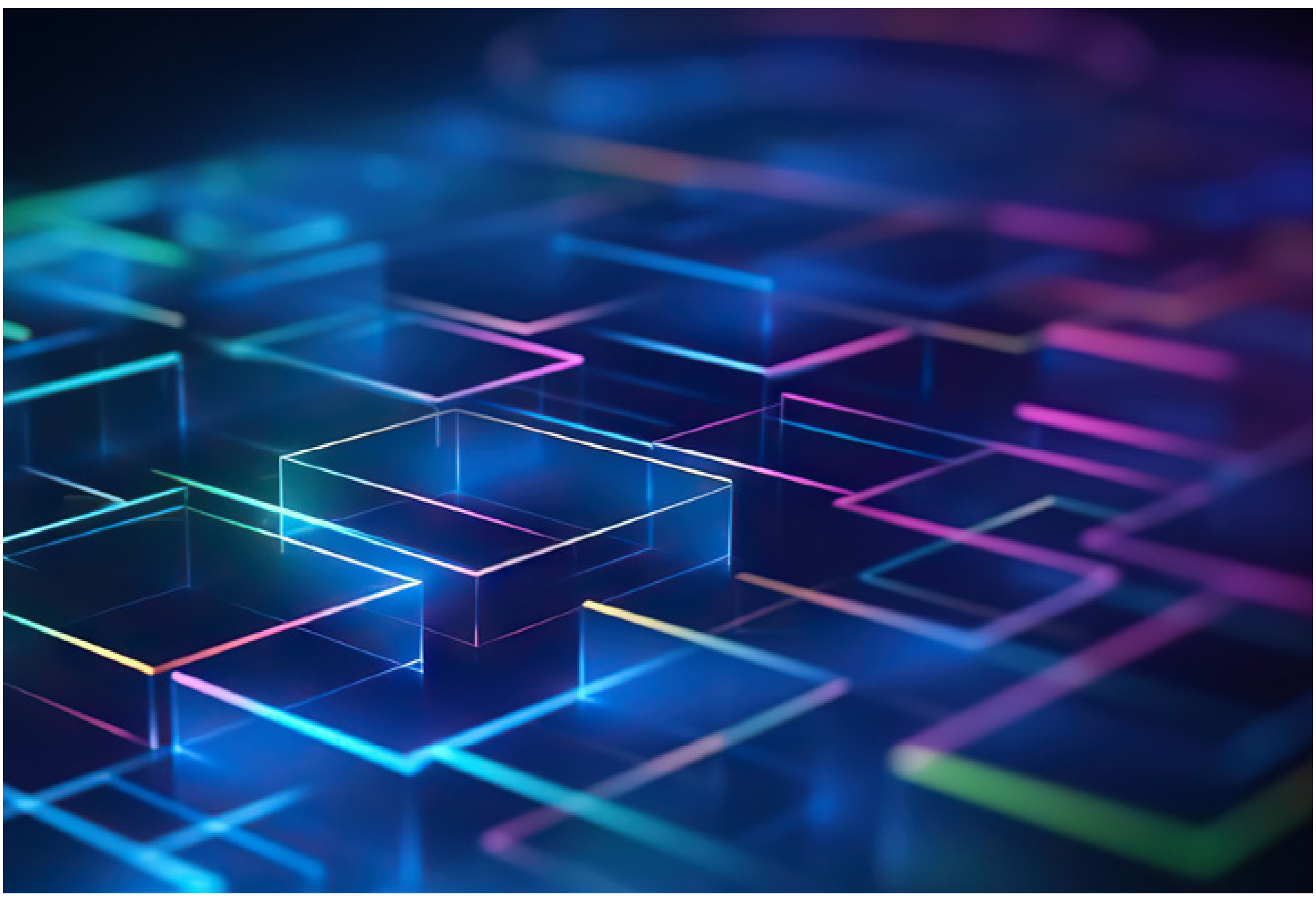
Domain-oriented data

Data as a product

Self-serve infrastructure

Federated computational governance

Below, we explore the importance and power of the second principle, treatment of data as a product. In the data mesh, data isn’t just an asset; it’s a product with well-defined owners, consumers, and quality standards. Treating data as a product motivates domain teams to manage their data as a product and treat the rest of the organization as their customer.



What does it mean to treat data as a product?

As organizations adopt data mesh, a transformative paradigm for their data architecture, the treatment of data as a product is at its heart. Treating data as a product calls for breaking down the data monolith into smaller data products, inspired by the "API-first" modular, reusable design principles adopted for applications over the last decade. This evolution will lead to data being treated as an enterprise asset and managed by source- or consumption-oriented domains.

SILOED ANALYTICS

Applications, the data they produce, and analytics are owned by each individual line of business in separate environments

There is little information shared across business units, and customers are managed independently

"Fragmented hindsight"

DATA LAKES

Organizations seeking to reduce data storage costs invested significantly building enterprise data lakes to hold raw application data

With relatively little curated data for analytics available on these platforms, analyst must "wrangle" the data before it is suitable for analytics

This process may consume up to 70% of their time which is a huge cost and waste to the organization

"Information overload"

ENTERPRISE MODELS

In an effort to become "customer-centric" and enable analytics, organizations developed enterprise data models to create a unified data structure for the organization

What went wrong?
For large organizations with a complex data landscape, just conforming the data required multiyear roadmaps with little value realized

One-size-fits-all model ended up requiring mass customization for unanticipated special case scenarios

Reliance on technology teams to design, build, and maintain data pipelines created a bottleneck

"Data purgatory"

DATA PRODUCTS

What can we do to fix this?
A paradigm in the perspective of treating data as a stand-alone product with its own life cycle, instead of a byproduct

Managing data as a product helps organizations take control of their data assets and realize business values

"Integrated insight"

What is a data product?

Data products are assets that help organizations take control of their data and generate business value. They are shareable and can help unlock the potential of data in a way that benefits both internal and external customers.

Organizations can lay the foundation for value delivery by instilling product thinking within the data ecosystem. Data products allow them to unlock the power of their vast datasets by empowering analysts to focus on business insights. Here are five ways data products can help an organization:

1. **Providing a single source of truth:** Cater to multiple use cases with standardized key datasets to ensure consistency of the insights across the organization.

2. **Enabling trust and accountability:** Increase confidence in data with consistent definitions, usage, and results across teams.

3. **Empowering efficiency and collaboration:** Make the standardized and reusable datasets easily searchable in a data marketplace where analysts can share their knowledge and findings.

4. **Increasing speed to market:** Let analysts focus on producing valuable insights for your organization instead of constantly battling with data.

5. **Reducing infrastructure and resource spend:** Reduce IT overhead with a data-simplified architecture. Federate IT development costs back to LOB analysts with a chargeback model.

What are the different types of data products?

Adoption of product thinking for enterprise data requires a clear understanding of different types of data products. Data products can be broadly categorized based on their purpose and the degree of transformation performed to use them. Understanding these types can help organizations manage and leverage their data products for optimal use cases and derive value from enterprise data.

a) Raw data products

This represents the most fundamental type of data generated and gathered from multiple sources. This data is provided in its original format, although some minor processing and cleansing steps may be performed. Organizations can decide how to utilize this data by further processing it to derive additional value. *For example—a list of all transactions related to products over a specified time.*

b) Conformed data products

These products are data products that bring data from raw or cleansed data products into a conformed data model per enterprise taxonomy for a specific data domain. *For example—conformance of deposit data generated by various deposit platforms to create an enterpriser data domain for deposits.*

c) BI-ready data products

These data products are further structured and optimized for analysis. The data undergoes further aggregation to be readily available for data analytics tools and applications. *For example—the total number of products sold by region.*

d) AI-ready data

The AI models require context along with the data to generate valuable insights from the enterprise data. The structured, semi-structured, and unstructured datasets undergo cleaning and organizing the data, tagged and categorized appropriately to tune and train robust LLM models to create models such as GenAI models with enterprise context. *For example—automated customer support.*

What is a typical data product life cycle?

The data product life cycle is a sequence of development and enhancements a data product passes through. It starts from intake and proceeds to prototyping, development, deployment, enhancements, and again to intake based on feedback.

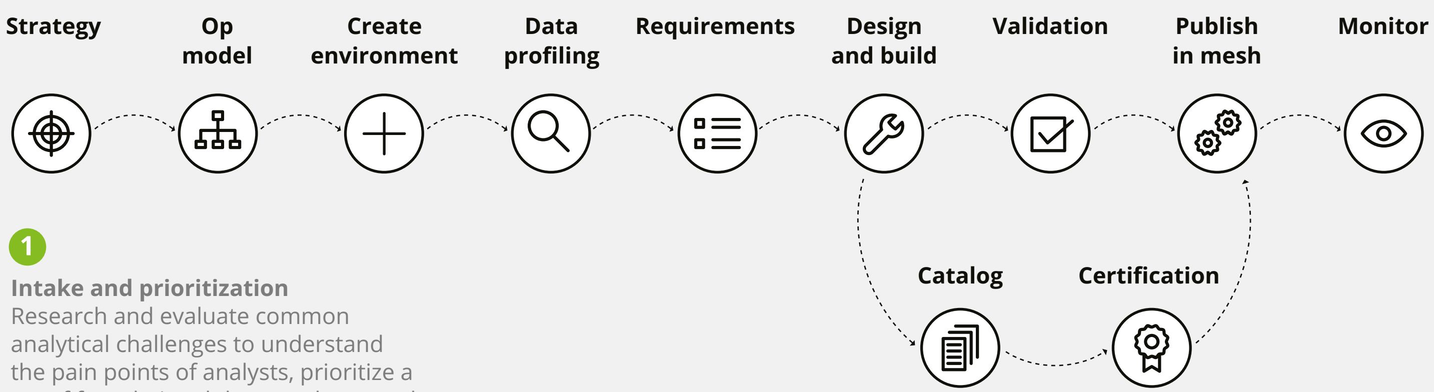
Data product life cycle

1 Intake and prioritization
Research and evaluate common analytical challenges to understand the pain points of analysts, prioritize a set of foundational data products, and identify champions and influencers.

2 Requirement gathering and prototyping
Create a sandbox environment where the data product team and dataset owner can collaborate and co-develop.

3 Product development
The data product team "productized" the dataset with core components (e.g., metadata, product marketplace(s), data source connectors).

4 Deployment, marketing, and monitoring
Deploy code and additional components through standard SDLC processes and procedures. Monitor performance and data product usage to identify opportunities for enhancements based on how analysts are interacting with the data.



5 Feedback and enhancement
Solicit feedback actively based on usage patterns and insights obtained from telemetry results, and work with dataset owners to vet new requirements. When there is mutual agreement on a change, start the agile delivery process to deliver as quickly as possible.

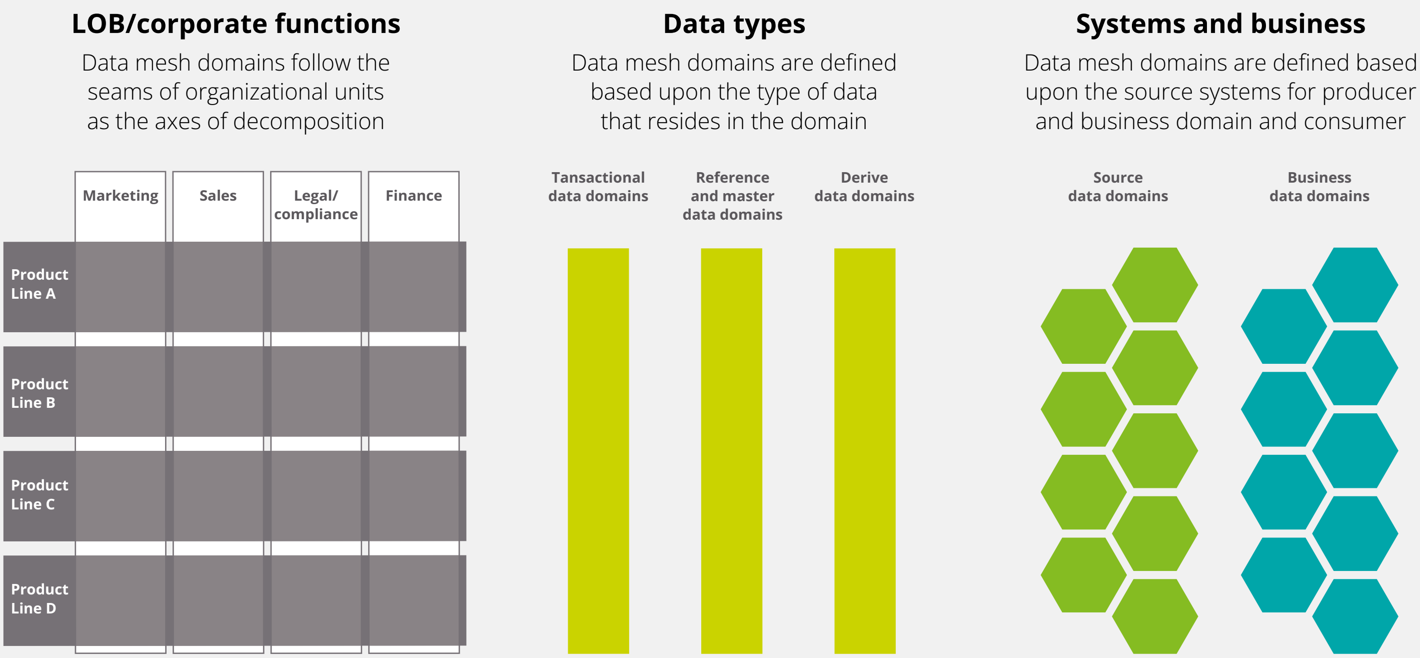
How to embark on a journey to build and treat data as a product

Instilling product thinking doesn't happen overnight. Throughout the years, organizations have experimented with various solutions to manage the rapid proliferation of data and drive business insights. To adopt this mindset, you will need to overcome some challenges:

1 *How do we break the monolith data ecosystem into products?*

The first step is eliminating the constraints associated with centralized monolithic architectures by breaking data into domains and distributing steps within the data processing pipeline. "Decompose and distribute" is a critical step that focuses on segmenting data across domains within an organization into more manageable, domain-specific pieces and distributing these across a network to improve accessibility and scalability.

Below are some ways to distribute data domains that involve identifying and organizing data around distinct business contexts and capabilities:



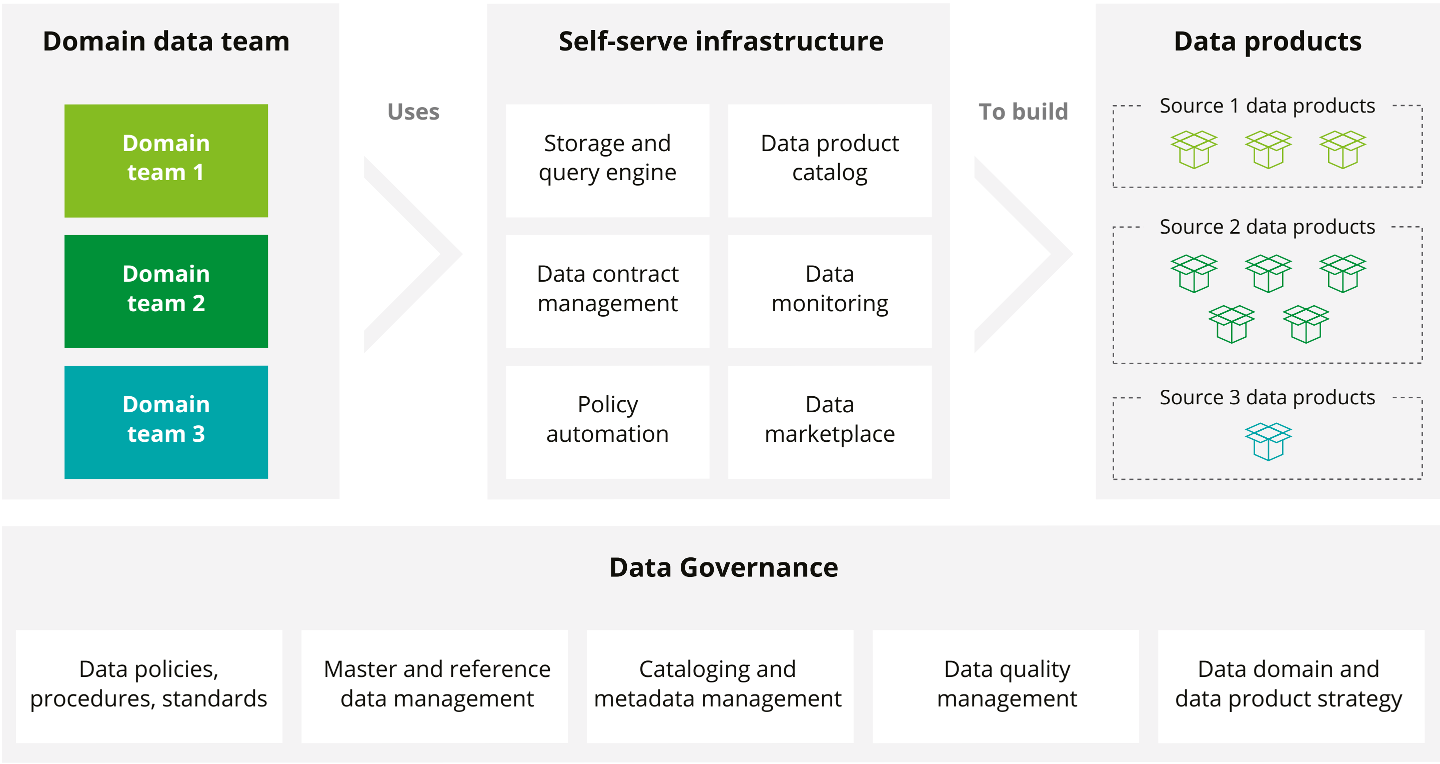
Each approach offers benefits but has associated challenges. Aligning data products per LOBs and corporate functions provides ease of alignment given IT systems and teams are mostly aligned with LOBs/corporate functions; however, the organization's structures change over time (e.g., due to mergers and acquisitions), which may affect the domain definition for the data mesh.

Breaking down the monolith by data type—transactional, master, reference, or derived data—offers a suitable mechanism for building domains based on data types. However, this approach requires buy-in and clear ownership assignment. Whereas with third approach—aligning data products with sources and analytical use-case systems—leads to many domains. A smaller organization can take this approach; however, large organizations may end up creating duplicate data products.

This selection of ways to distribute domains is not exhaustive. Thus, before breaking down the monolith, organizations should consider their organizational context—data, technology, structure, size, needs, and other requirements to ensure number of data products are mutually exclusive and collectively exhaustive.

❷ *How do we provide a technology platform for teams to build data products?*

Data product teams require different tools and technology. A cohesive data infrastructure is a platform that provides capabilities for building data pipelines, storage, streaming, and management. The data infrastructure team can own and deliver the necessary tools for domain teams to capture, process, store, and serve their data products. Set up domain-agnostic data infrastructure and platforms and provide the data infrastructure components in a self-service manner.



It is essential for the data infrastructure team to lay down some standards for the domain teams. Here are some ways they can help:

- **Right mix of tools—**There is no one-size-fits-all approach due to domain teams having different use cases. Therefore, the data infrastructure team needs to set guardrails, including the organization's security and compliance regulations and the rules that must be adhered to regarding new tools. On top of this, it's the infrastructure team's responsibility to provide guidance on which tools should be used based on the domain team's business needs. This is imperative to ensure data products can be shared easily without creating redundant copies of data.

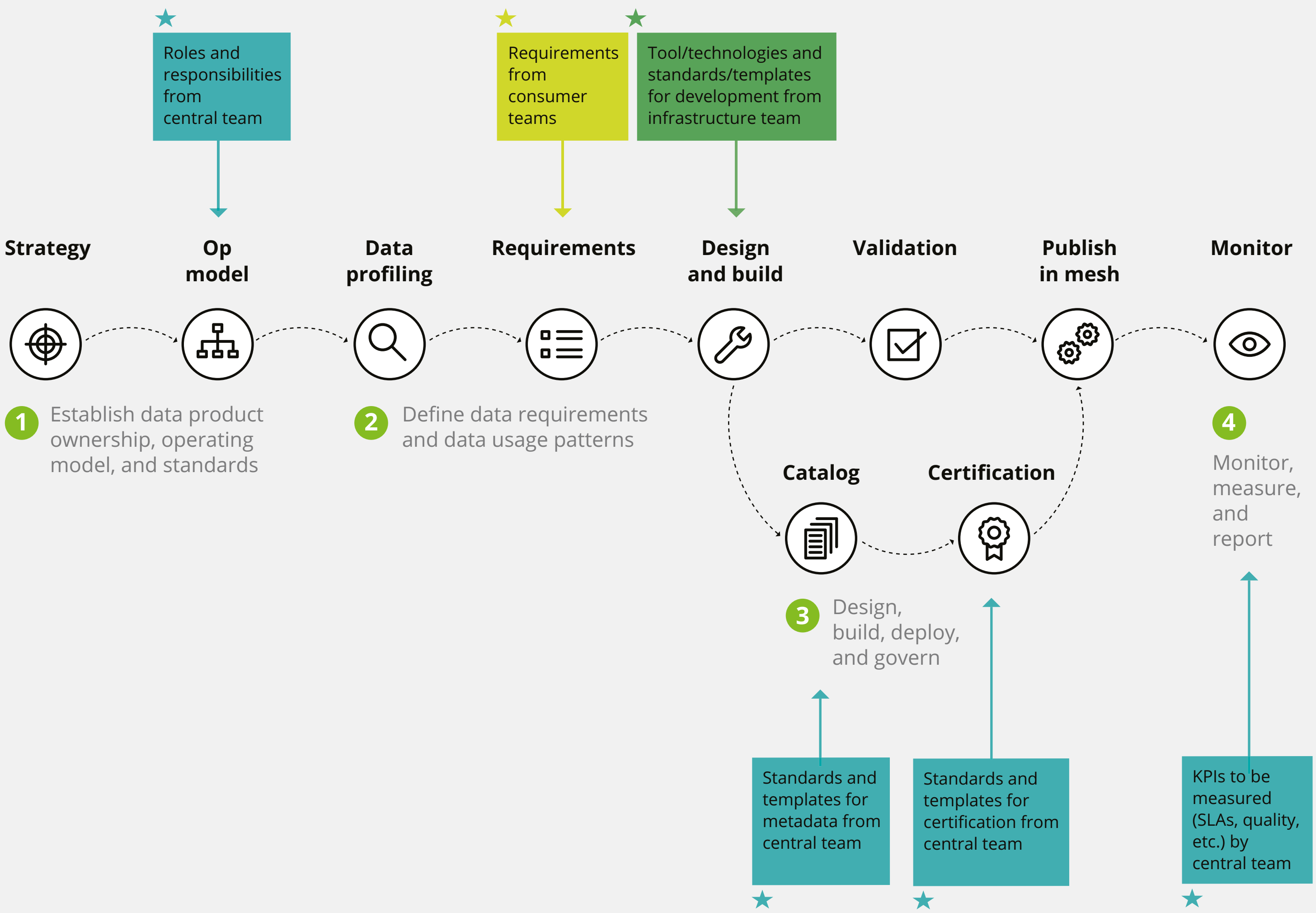
- **Defining standards and templates for development—**Create standards for encryption, anonymization of data, etc. Create pipeline templates and document best practices such as error handling and logging. Publish them to the domain teams.

- **Strong FinOps—**Domain teams already have a product mindset—they prioritize the user experience and the value the product delivers to the customer. Similarly, data infrastructure treating domain teams as customers would prioritize them to provide the best services, tools, and guidance. Treating them as customers also includes being explicit about the cost of services and billing them appropriately.

3 *How do we govern data products during their life cycle?*

Data governance is crucial to maintaining quality, consistency, and security across the data product life cycle especially within a regulated industry. A federated custodianship model distributes ownership of governance policies and enforcement across central and domain teams, promoting a culture of collaboration and tailoring products to specific domain needs. This collaboration between autonomous domain and central teams aims to meet the organization's overall data needs effectively. It also fosters a scalable and agile governance approach, encourages ownership and buy-in from domain teams, and ensures consistent data quality and compliance across the entire life cycle of data products.

Iterative data product delivery model



Breaking it down a little further, federated governance allows the domain teams to:

- Define, create, and publish data products to the mesh.
- Create and maintain the quality of the data products.
- Create and maintain the domain catalog.

Meanwhile, central teams can:

- Defines standards, policies, and procedures that need to be followed by the domain teams.
- Define and create metadata management processes and templates for domain teams.
- Define roles and responsibilities.

Central teams are also responsible for defining standardized data contracts to enable seamless data exchange and consumption across domains, preventing silos and ensuring interoperability, workflows for provisioning data access, defined checks and balances, and key performance indicators (KPIs) to ensure data products comply with data governance standards.

4 *What talent model is required to build and manage data products across their life cycle?*

The process of breaking down the data monolith into smaller data products should account for the changes required in the organization's data engineering and analytics talent model. Data product and domain teams should comprise a specialized group of individuals focused and aligned to produce and deliver high-quality data products. However, the success of the data mesh and data product teams within the organization also requires support from other teams.

Domain teams

The domain teams are the cornerstone of the data mesh and the powerhouses of creating and maintaining high-quality data products. This team needs to be augmented with new skill sets around the data product owners and the domain data product developers (e.g., data engineers, data analysts) to collaborate effectively with other teams. They are responsible for:

- Understanding the high-level architecture and domain definitions.
- Developing and operating the data products.
- Understanding of business domain and process to ensure data is fit for purpose across the organization.
- Offering them on the marketplace with specified service-level agreements and KPIs.

Data infrastructure platform team

The data platform team is crucial to successfully implementing and maintaining the underlying tools and technologies needed for data product development and the overall data mesh architecture. They are responsible for:

- Designing the data mesh architecture to enable data sharing with clear domain definitions and boundaries.
- Operating the marketplace for efficient data product discovery.
- Reviewing and approving domain designs for consistency.
- Installing and managing governance tools to ensure compliance and security.
- Providing self-service modules and training for domain teams to effectively utilize the platform.

Central team

The central team holds immense importance. It is momentous for ensuring enterprise data integrity, quality, and compliance. It is responsible for:

- Determining data governance needs and necessary adaptations to governance processes.
- Developing and maintaining enterprise governance standards, ensuring alignment with organizational goals and industry best practices throughout the data product life cycle.
- Maintaining charter and facilitating governance forums where key stakeholders collaborate, track metrics, and address emerging challenges.
- Leveraging advanced governance tools to oversee enterprise data, ensuring adherence to established standards and regulations.

Enablement team

The enablement team is cardinal to the successful implementation of mesh.
It is responsible for:

- Enabling data mesh adoption and transition to the data mesh.
- Providing technical training.
- Guidance on best practices for data management.
- Fostering cross-functional collaboration.
- Promoting data literacy.
- Establishing a continuous improvement and feedback loop.

Domain teams	Data infrastructure platform team	Central team	Enablement team
Data engineer	**Data architects**	**Data governance lead**	**Change management specialist**
Data product owner	**Platform product owner**	**Data stewards**	**Change manager**
Domain experts	**Site reliability engineer**	**Data security and privacy engineer**	

What are the must-do's to get this right?

When embarking on this data mesh journey with product thinking, there are some must-do's to make it successful:

- **Number of data products:** Start small, i.e., one product at a time, and evolve incrementally into a full data mesh. Make sure to start with the right set of use cases that cater to a business goal. In short, don't build data products just because you can.
- **Culture change and enablement:** Implementing a data mesh requires both a technical change and a culture and enablement transformation. Embedding a culture and enablement team is vital for the success of the transformation. It plays an imperative role in the shift from a traditional monolithic data infrastructure to a distributed data architecture.
- **Measurable KPIs and customer experience:** Another important must is to define success criteria and business-aligned measurable KPIs, especially around data products. For example, how many data products are created, how many are consumed successfully, the rate of use of data products, and data-centric metrics like data quality and accessibility.
- **Leadership buy-in:** For any initiative that requires organizational-level changes, success hinges on leadership buy-in. It is vital for making decisions, creating momentum, and ensuring commitment in terms of securing resources.

Beginning your journey

For any organization, big or small, the journey to modernize the enterprise data platform starts with understanding the business drivers. Breaking down the enterprise data monolith into smaller data products will enable organizations to address growing data demand and maintain data quality across products.

Organizations must, however, holistically evaluate the maturity of their data and technology footprints, governance processes, culture, and talent before embarking on the data mesh journey. Understanding context is imperative to break the monolith into data products. This will allow organizations to develop the right number of products and to derive value from their data assets.

Authors

Parth H Patwari
Principal
ppatwari@deloitte.com

Tanmay Kumar
Managing Director
tkumar@deloitte.com

Abhishek Kumar Verma
Senior Manager
abhishverma@deloitte.com

Jitin Arora
Manager
jiaroradeloitte.com

Deloitte.