



How-To Guide

# Accelerating Your Cloud Migration Journey with a **Data Mesh Architecture**



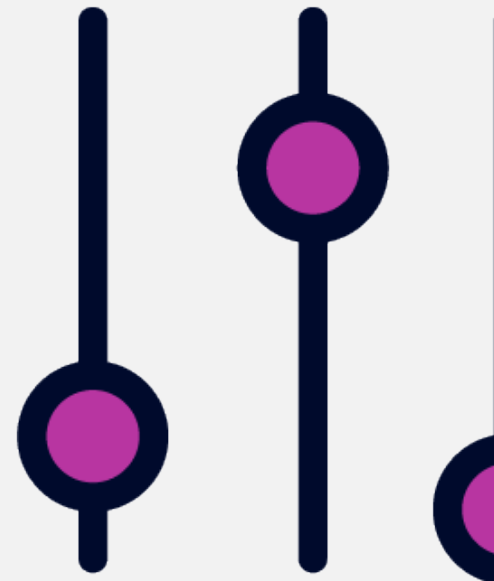
# Competitive Data Analytics Before, During, and After Migration

One of the primary goals of digital transformation is to become a data-driven organization – one that extracts the greatest value from data through analytics and machine-language (ML) modeling.

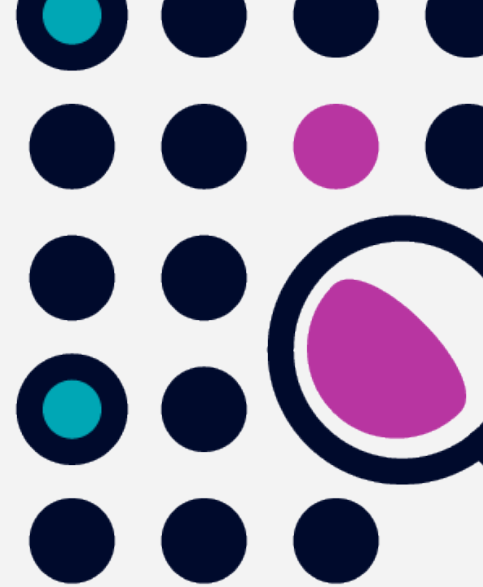
The path toward digital transformation is largely via migration of workloads to a reliable, scalable, secure, and performant cloud environment.

Data migration to the cloud cannot happen overnight, and sometimes, organizations may choose to keep certain types of data in-house and not migrate it at all. **So, how can organizations quickly leverage the data analytics capabilities they need to compete before, during, and after data migration?**

The solution is **data mesh** – a concept of distributed data ownership that eliminates the need for a centralized data architecture while also allowing organizations to think about data as a product. In this guide, we'll explore the meaning and value of data mesh as a broad concept as well as ways it can facilitate and accelerate cloud migration.



# Roadblocks on the path to digital transformation



The concept of “digital transformation” has been around for more than a decade and has since become widely accepted among enterprises across multiple industries.

**Typically, organizations saw this transformation as a long-term evolution that would take years to complete.** But circumstances over the last couple of years have accelerated this outlook dramatically. Workforces became hybrid or remote. Rising customer expectations forced a complete reconsideration of the customer experience – not to mention business models. And partner ecosystems grew in size and complexity in response to supply chain challenges. **Suddenly, cloud migration and data analytics became more urgent concerns.**

But migrating mission-critical workloads from on-premises servers to the cloud is easier said than done. **It can be a slow, complex, costly, and disruptive process** – especially if you hope to modernize your applications along the way,

as most organizations do. Some data, whether for regulatory reasons or corporate policy, cannot be removed off premises at all. These factors can slow down time to value beyond what is acceptable.

Even beyond the issue of cloud migration, the path toward successful data analytics has its own speedbumps. The traditional approach – with a central IT team serving as gatekeeper and data locked in siloed systems – makes efficient use of data difficult to achieve. **When data is disconnected from the business units closest to it, it can become stale and untrustworthy. When sharing is difficult, the risk of redundancy increases.** All this stifles innovation and leaves data value on the table.

# What is a data mesh?

As mentioned above, data mesh helps organizations overcome these challenges, but in order to understand how, we must begin by understanding the concept itself.

The founder of data mesh, Zhamak Dehghani, describes the concept as **“a decentralized sociotechnical approach to share, access, and manage analytical data in complex and large-scale environments – within or across organizations.”** Dehghani chose the word “sociotechnical” deliberately, as data mesh is both a way to organize data on a technological level as well as on an organizational level around business domains.

**Data mesh is concerned with analytical rather than operational data.** Operational data is transactional and streams in real or near-real time, describing the business in its current state. By contrast, analytical data is a compendium of historical data that remains largely unchanged. It is the primary source for developing insights and predictions to help optimize the business in its future state.

<sup>1</sup>Zhamak Dehghani, *Data Mesh: Delivering Data-Driven Value at Scale* (Sebastopol, CA: O'Reilly, 2022).

The following **four principles** mean that data mesh distributes data both on a technological and an organizational level. Individual data sources might live on-prem, in one or more data centers, or on one or more clouds, but they are accessible through a unified interface. The platform provides a **single, consolidated view into both data sources and products – maintaining governance while allowing users from any domain to access data and develop products.**

# Dehghani organizes **data mesh** around four principles:

## 1 Domain ownership

In data mesh, ownership of analytical data is decentralized. Data belongs to business domains closest to the data. These domains can either produce the data or be the primary consumers of it.

## 2 Data as a product

The productization of data is one of the most exciting aspects of data mesh. Organizations share domain-oriented data as a product directly with data users, such as data analysts and data scientists. Each data product is autonomous, and its life cycle and model are managed independently of others. Furthermore, every data product is highly usable because it is:

- Discoverable
- Addressable
- Understandable
- Trustworthy
- Accessible
- Interoperable and compostable
- Valuable on its own
- Secure

A data product is the smallest unit of architecture that can be independently deployed and managed. It has all the structural components that it requires to do its function: the transformation code, the data, the metadata, the policies that govern the data, and its infrastructure dependencies. Data mesh scales out by adding and connecting more data products, which teams share freely rather than collecting and siloing. This is the foundation of a data-driven culture of innovation.

## 3 Self-serve data platform

The operational technology that supports data mesh also encourages self-service to democratize data, empowering teams across domains to share data with minimal friction and without the involvement of a centralized IT team. This platform manages the full lifecycle of individual data products as well.

## 4 Federated computational governance

Even though data within a mesh is distributed among domains, it is governed by an operating model that creates a centralized incentive and accountability structure. This structure balances the autonomy and agility of domains with the global interoperability of the mesh as a whole.

# Accelerating the hybrid cloud migration journey



As you can see, data mesh is a valuable end state for data management, sharing, and productization that balances governance with innovation. (We'll delve into this end state later in the guide.) But it also serves as **a practical solution for accelerating time to value during cloud migrations** – no matter where your organization may be currently along its cloud journey.

With traditional, centralized data architecture, the ideal state would be to have all data reside in data lakes, warehouses, or lake houses in a single cloud where it can be transformed and analyzed. In practical terms, however, migrating data from its origins – multiple data centers and clouds – can take years of time, money, and disruption.

Imagine, for example, a pharmaceutical company that makes multiple acquisitions each year. Every new acquisition means multiple new data sources, in multiple locations, and in multiple formats. Spending time and resources consolidating this into a single location – even

a public cloud – would be impractical. This impracticality means that this data would remain siloed, greatly reducing its value and potentially negating the purpose of the acquisition itself. The growing company would need to transfer data on an *ad hoc* basis, a slow and error-prone process at best.

**Data mesh decouples the decision to migrate data from one location to another from the ability to leverage its full value for analysis.** Any part of the business can fully utilize data wherever it's located and in whatever state it exists: individual databases, data lakes, warehouses, or lake houses.

This means that data migration could be organized according to a number of priorities, such as cost, usage, or performance – allowing organizations to make smarter decisions around data migration without impacting its accessibility and productization. **Connect to a data source today while on-prem, and you'll stay connected to it tomorrow once it has migrated to the cloud.**

# Evolving data mesh and data migration side-by-side



The beauty of leveraging data mesh during data migration is that the two processes can evolve simultaneously. Just as data migration is a developing journey – one your organization is likely already on – with different maturity levels, so too is data mesh.

In one sense, the technology is the easy part: **[a solution like Starburst](#) provides the self-service platform that connects data sources from wherever they are to business users.** The greater challenge is the cultural and organizational shift to domains and domain-based data ownership.

Traditional organizational structure centralizes the responsibility of sharing analytical data to a team that performs activities related to the technology, such as the IT team. Data mesh shifts responsibility for data – except governance, security, etc. – to teams with a “sphere of knowledge, influence, or activity” such as sales, finance, marketing, etc. **This transformational shift to domain-oriented data ownership requires careful change management in order to succeed.**

<sup>2</sup>Eric Evans, *Domain-Driven Design: Tackling Complexity in the Heart of Software* (Upper Saddle River, NJ: Addison-Wesley, 2003).



# Starburst on Google Cloud for data mesh

The partnership between Google Cloud and Starburst can facilitate both cloud migration and the simultaneous development of a robust cloud mesh solution.

**Starburst can help organizations ready to implement data mesh, using this three-step process:**

## 1 Connect to data sources where they reside

A key data mesh implementation principle is to connect your data sources by leveraging your existing investments: cloud or on-premises, structured warehouse or non-structured lake. Starburst offers [50+ connectors](#) that enable you to connect to data sources.

## 2 Create logical domains

After connecting across all the various data sets, the next goal is to create an interface for business and analytics teams to find their data. You can create these interfaces in Starburst so that all the data they need resides in your domain alongside domain teams that are empowered to work autonomously. Starburst also provides self-service where data consumers can independently do more on their own.

## 3 Enable teams to create data products

When you provide a domain team access to the data they need, the next step is to teach them how to convert data sets into data products. Then, with a data product, create a library or a catalog of data products. Starburst has a built-in catalog that enables you to very quickly search, discover, and identify data products that might be of interest.

Google Cloud complements Starburst's data mesh capabilities in a number of ways – starting with the data migration process – for which Google cloud offers multiple automated tools such as the [Database Migration Service](#).

In addition, Google protects more people online than anyone else and has translated this knowledge and experience into a comprehensive set of enterprise security features that help you protect your data whether on-premises or on Google Cloud.



**Most significantly, Google Cloud provides a full set of data analytics tools so you can extract the greatest value from your data. These include:**

### **Big Query**

Streamline your migration path and modernize from any on-premises or cloud data warehouse with BigQuery. Solve for today's most challenging demands and seamlessly scale your business with built-in AI and powerful multicloud capabilities.

### **Dataproc**

Empower your teams to securely and cost-effectively ingest, store, and analyze large volumes of diverse, full-fidelity data. Use Dataproc for data lake modernization, ETL, and secure data science.

### **Streaming Analytics**

Ingest, process, and analyze event streams in real time to make data more useful and accessible from the instant it's generated. Dataflow unifies streaming and batch data analysis and builds cohesive data pipelines. Pub/Sub ingests hundreds of millions of events per second.

### **Business Intelligence**

Drive better outcomes through data-driven experiences with Looker and BigQuery.

### **AI and ML**

Build, deploy, and scale more effective AI models with our unified machine learning platform, Vertex AI. Derive insights from unstructured text and images with Natural Language AI and Vision AI

### **Google datasets**

Increase the value of your data assets when you augment your analytics or AI initiatives with external data. Discover and access unique and valuable datasets and pre-built solutions from Google, public, or commercial providers. With fully managed data pipelines, you can stay focused on what matters most: delivering insights and business value.

And finally, Google Cloud continues to develop its rich partner ecosystem of system integrators and independent software vendors to augment your data mesh with third-party integrations with Google Cloud capabilities in AI, ML, data, analytics, and security.

# Data mesh: The next generation of data management

Data mesh brings data management into the future by facilitating the adoption of hybrid cloud, democratizing data access, and removing data silos so that teams across the enterprise can reap the benefits. **With data mesh, your organization gains this accessibility to data – no matter where it actually resides.**

When you don't have to worry about data location in order to realize its full value, you can make smart migration decisions. Starburst and Google Cloud have the technologies and know-how you need to lay the foundation for a future-proof data mesh architecture.

To learn more, check out [Starburst on Google Cloud](#).

