

The Ins & Outs of Data Mesh

Data Mesh is a new socio-technical data management paradigm proposed by Zhamak Dehghani that challenges the traditional monolithic data architectures with a decentralizing approach to ultimately make data usable in companies. The challenges of the current enterprise data architecture are that they still aren't able to fully manage the large number of sources and enable a diverse set of consumers.

Therefore, this framework tries to solve the bottlenecks of a centralized data management architecture such as data lakes and to bridge the gap between the business needs and the technology. The idea is to shift the mindset to make data to be a product that is used rather than an asset that is treasured. This will be achieved by transferring data ownership in a decentralized infrastructure to the domains (the organizational units) which generate data and understands it the best. The domains are in charge of preparing and sharing the data so other domains can consume it on their own in a self-service manner. Automated processes and centralized governance and security measurements are further fundamental components of the data mesh.



A data mesh considers domains as a first-class concern, applies platform thinking to create self-serve data infrastructure, treats data as a product, and introduces a federated and computational model of data governance.

Zhamak Dehghani



● What is Data Mesh?

To enable the paradigm shift of moving beyond the traditional monolithic approaches, data mesh is based on four key pillars:

● Domain-oriented decentralized data ownership and architecture

Decentralization and distribution of responsibility to the domains are the core aspects of data mesh. The idea is that the data should reside and be managed by the people who are closest to it as they are most familiar with it. They understand best what the data represents and how it is used. The users can then make the adjustments in the most efficient way. Thereby, the bottleneck of centralized infrastructures which require a separate entity to take care of all tasks related to data management can be removed, e.g. data scientists looking for data in a data lake environment.

● Data as a product

As the decentralized approach can lead to an increasing number of data silos and suffering data quality, the data needs to be treated differently. The perspective needs to change from data being an asset to data being a product that can be discovered, understood, and trusted by the consumers who should be treated as its customers. In order to satisfy these customers, the domain owners need to provide data as a product that is of high quality, trustworthy, and which decreases the lead time of data consumption.

● Self-Service data infrastructure as a platform

Managing data infrastructures is complex and requires special skills which won't exist in all domains. In order to still enable the different domains, a self-service platform with capabilities for data integration and transformation, implementation of security policies, data lineage, and identity management is needed.

● Federated computational governance

To ensure a healthy and interoperable ecosystem in this decentralized set up, the data mesh implementation requires a governance model with global standardization and harmonization. The standardization efforts address concerns in regards to data semantic/syntax modeling, metadata formatting, identification management, etc.

● Why is Data Mesh gaining so much traction and who should consider building a Data Mesh?

Despite the technological developments, evolutions, and improvements of the past years, the pain points in becoming data-driven, using data to compete or using data at scale to drive value have hardly changed. Companies are still struggling to solve the ongoing and increasing data challenges such as data silos, leveraging data for analytical as well as transactional use cases, long lead times in adding new data sources, and bad data quality. All of this, caused by the underlying architecture and unchanged organizational culture towards data.

Data Mesh is a compelling concept for companies that deal with a large number of disparate data sources. It is also relevant for companies that deal with different sets of consumers coming from operational and/or analytical systems of the business. The domain-driven approach together with the other three pillars of data mesh is technology-agnostic with the focus to bridge the gap between the data and all individuals within an organization who work with it. It balances the human as well as technical part of an organization and aims to create a better common understanding of the data via this mechanism and as a result also improves the management. On the technical side, it recognizes and respects the nature and topology of the data and the different use cases that it can enable. On the human side, it looks at the individual personas of data consumers, their diverse access patterns, and their domain specific knowledge.

● What are typical Data Mesh use cases?

- **Customer 360:**
More automated processes providing better personalized and contextualized customer experience. Results are reduced average handling time, increased first contact resolution, and improved customer satisfaction.
- **Marketing:**
Marketing teams are enabled to run the targeted campaigns to the right customer, at the right time, and via the right channels.
- **Data Privacy:**
Customer data can be protected by complying with the ever-emerging regional data privacy laws, like GDPR. Security rules can be easily applied through the integration of external data governance, policy, and security tools (such as Collibra) on the global level prior to making it available to data consumers in the business domains.
- **IoT Device Monitoring:**
Insights into the device usage patterns help to continually improve product adoption and profitability.
- **AI and Machine Learning Training:**
Machine learning (ML) and artificial intelligence (AI) models can be easily fed with data from different source to help them learn, without running the data through a central place.
- **IT and DevOps:**
Data latency can be reduced by providing instant access to query data from proximate geographies without access limitations.
- **Loss Prevention:**
The domain-oriented decentralization allows to analyze the data and run fraudulent behavior models on local level and thereby, detect and prevent fraud in real-time.
- **M&A:**
The combination of decentralized data ownership and federated governance allow data sovereignty and data residency on regional level while complying with data governance rules on the global level.



However it's the management and access to the analytical data that remains a point of friction at scale. This is where data mesh focuses.

Zhamak Dehghani

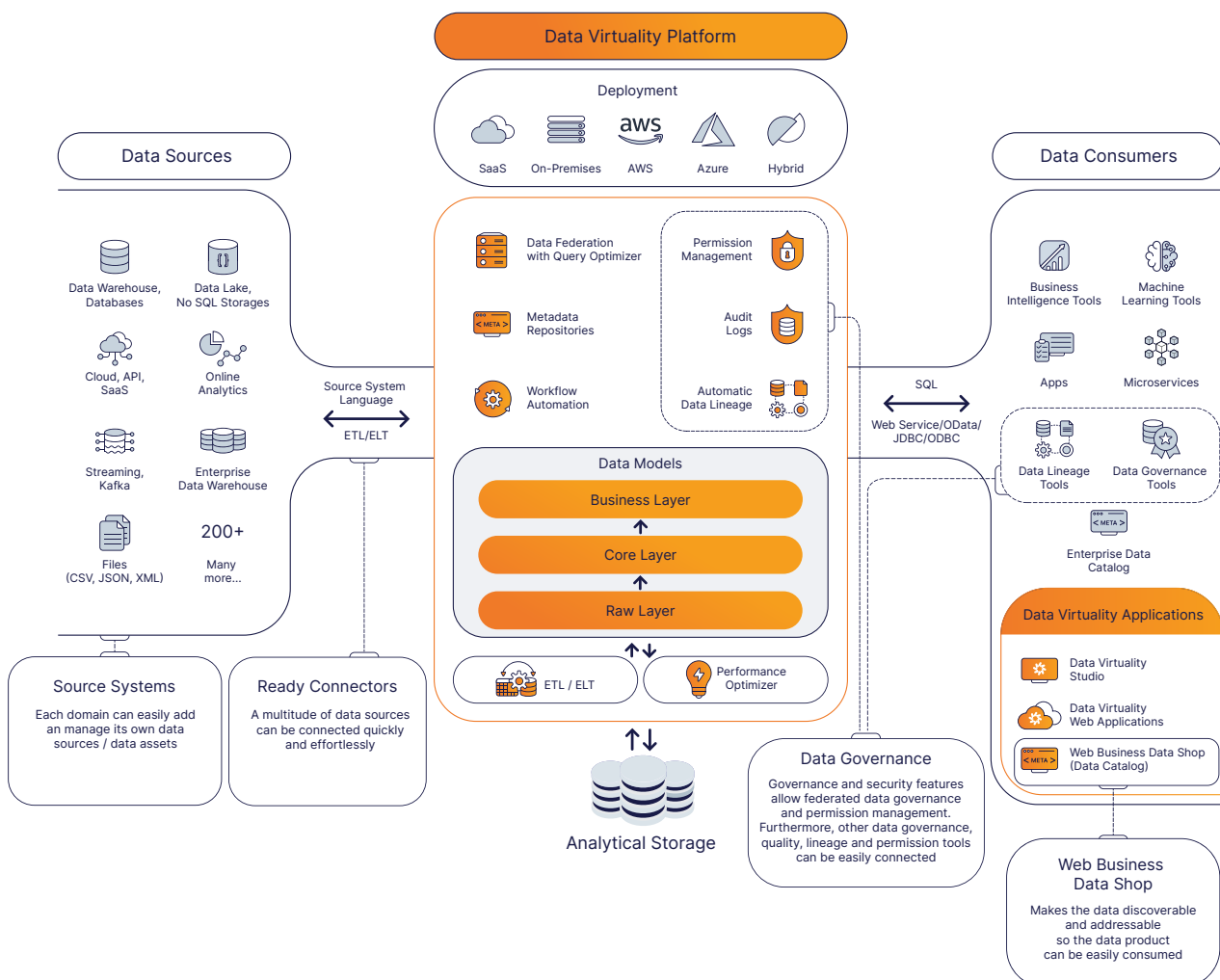


● How does the Data Virtuality Platform enable a Data Mesh?

The Data Virtuality Platform enables logical and distributed architectures that are essential to data mesh by combining two technologies; data virtualization and automated ETL.

- Data models can be created and shared in the virtual layer by the individual domains.
- The data products can be accessed via SQL (JDBC, ODBC) or APIs.
- Global governance, quality, and security measurements can be enforced and managed by a central team with the governance and security features provided in the Data Virtuality Platform.
- Metadata repositories make the data discoverable and addressable so the data products can be consumed.
- Other specialized governance, policy, and data catalog tools can be easily connected.

How the Data Virtuality Platform enables Data Mesh



● Why should I enable the Data Mesh with the Data Virtuality Platform?

The Data Virtuality Platform has been designed to provide a unified, governed, and secure data layer on top of multiple distributed data systems and is a natural fit for enabling a data mesh. In detail, the four pillars of the data mesh are addressed as follows by the Data Virtuality Platform:

Data Mesh Pillars	Data Virtuality Platform
Domain-oriented decentralized data ownership and architecture	<p>Scalability through decentralized data ownership</p> <ul style="list-style-type: none"> • Domains can autonomously set up data assets and data models • Data virtualization with or without data replication supports decentralized architecture • SQL makes the use of the tool easier in all the different domains
Data as a product	<p>Data becomes a product as it can be easily discovered, understood, and securely used</p> <ul style="list-style-type: none"> • Web Business Data Shop makes the data discoverable and addressable • Over 200 connectors enable easy access to data assets • Data delivery layer provides a powerful and easy-to-use foundation for using data as products • It's easy to prepare the data for the other domains to use (with our transformation option within our platform)
Self-service data infrastructure as a platform	<p>Reduced complexity supporting autonomous data consumption through self-serve platform</p> <ul style="list-style-type: none"> • Data marketplace provides self-service access to data • Data modeling layer which can be used by business users and can be easily shared • Analytical as well as operational data can be easily combined and used
Federated computational governance	<p>Healthy and interoperable ecosystem through global rules for federated governance</p> <ul style="list-style-type: none"> • Unified governance of decentralized data • Security and data protection features as well as row/column level security, data masking are provided within the Data Virtuality Platform • External centralized identity and policy management platforms (such as Collibra and Infogix) can be easily integrated

● About Data Virtuality

- **Founded:**
2012 by Nick Golovin (PhD) in Leipzig, Germany after 8 years of research
- **Offices:**
Munich, San Francisco, Leipzig
- **Solutions:**
Data Virtuality Platform SaaS
Data Virtuality Platform On-Premises
Data Virtuality Pipes Professional
Data Virtuality Pipes
- **Acknowledgements:**
Honorable Mention in 2022 Gartner Magic Quadrant for Data Integration Tools
- **Awards:**
Most Innovative Data Management Provider 2022, 2021 and 2019 (A-Team Insights)
2020 and 2019 Deloitte Technology Fast 50

Message: info@datavirtuality.com

Visit: datavirtuality.com

Request Demo: demo@datavirtuality.com

Data Virtuality Platform SaaS Free Trial: <https://eu.platform.datavirtuality.com/#/start-trial>

