Masthead

(https://mastheadata.com)

Start now
(https://app.mastheadata.com/sign-up)

August 20, 2022

What is Data as a Product, and What to Consider Implementing It?

In this blog post, we unfold the 'data as a product' principle of Data Mesh and elaborate on why product thinking is so essential for data, and how to convert data into something that organizations can convert into irreplaceable assets.

Data as a product is an underlying concept of the Data Mesh architecture.

Data Mesh is a modern data architecture philosophy that emphasizes decentralized data management and governance. Its main objective is to scale with organizations' ongoing growth and data change. Data Mesh spreads and promotes four main principles:

1. Domain-driven data architecture.
2. Data as a product.
3. Data infrastructure is a self-service platform.
4. Data governance is defined locally, not in a centralized manner.

The article below will unfold the data as a product principle and explain in detail why it is important, its caveats, and how to implement it.

Talking about the "data as a product" principle, one of the biggest risks here is abusing the concept itself. Data owners could treat decentralization as such where each team cares about their data to the point that they ignore the data users and turn data into a "My Precious" asset that nobody can actually use.
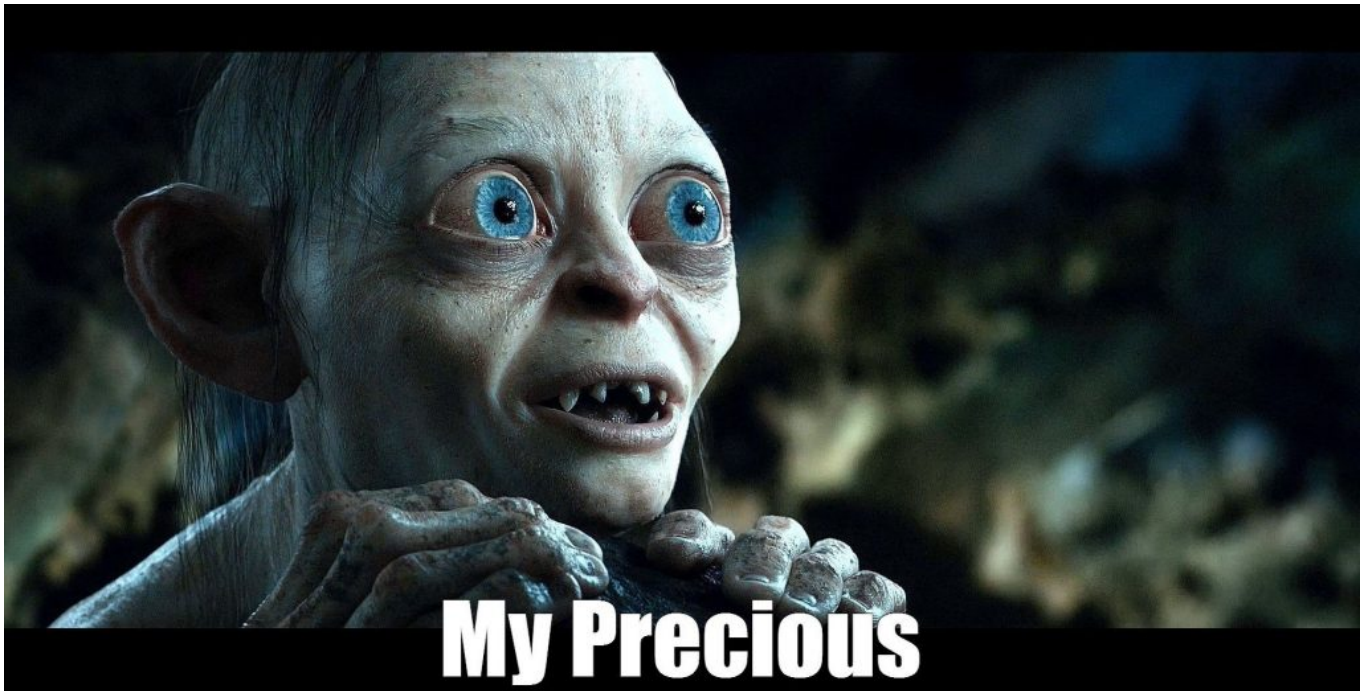
This, in turn, might create new problems with data quality or add new to the existing ones. Instead of using decentralization as a way to own and govern the data, individual teams risk creating data silos where they don't need to care what's going on with data beyond their domain and how data from their department serves the others in the organization.

Adopting the 'Data as a Product' mindset

Implementing Data Mesh demands a thorough and thoughtful approach from all parts of an organization.

A decentralized Data Mesh approach is a route to transforming the organization's culture. Treating data as a product in a company urge having teams responsible for it, and it also requires product thinking where data is managed as a product and stakeholders are being treated as customers. This means that the data owners' goal is to provide clean, valid, and reliable data products.

In other words, the 'data as a product' concept requires a shift in the mindset of data owners. Data stops being ephemeral stuff that your team just inserts, updates and deletes in the data warehouse to get their job done. Data becomes the core output of every system which has its consumers.

Bringing Product Thinking to Data

The 'data as a product' approach requires every team in an organization to adopt product thinking. Why is product thinking so important with Data Mesh? Generally speaking, it requires data owners to think beyond their department and understand how their particular data product influences and integrates into the data of the entire organization. One of the main objectives here is to make data discoverable, clean and complete so other teams in the organization can make good use of it. Product thinking requires each team that produces data in an organization to start treating internal data users as first-tier customers.

Adopting product thinking can help data teams see the impact of their work and improve the quality of data products they deliver. It also allows them to focus on delivering value to customers rather than just managing data.

You may think of data products as an output of a single node inside a cluster of interconnected nodes (Data Mesh).
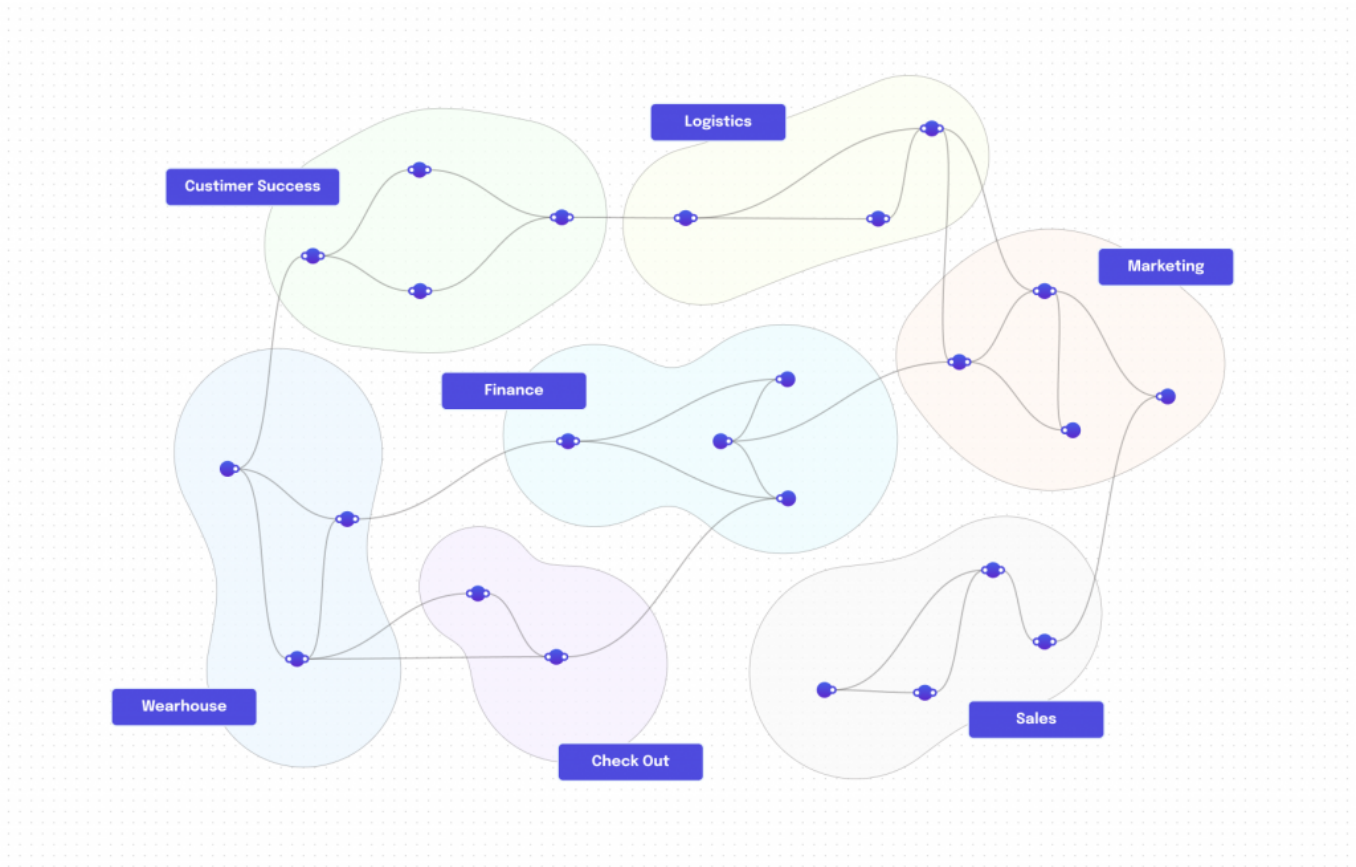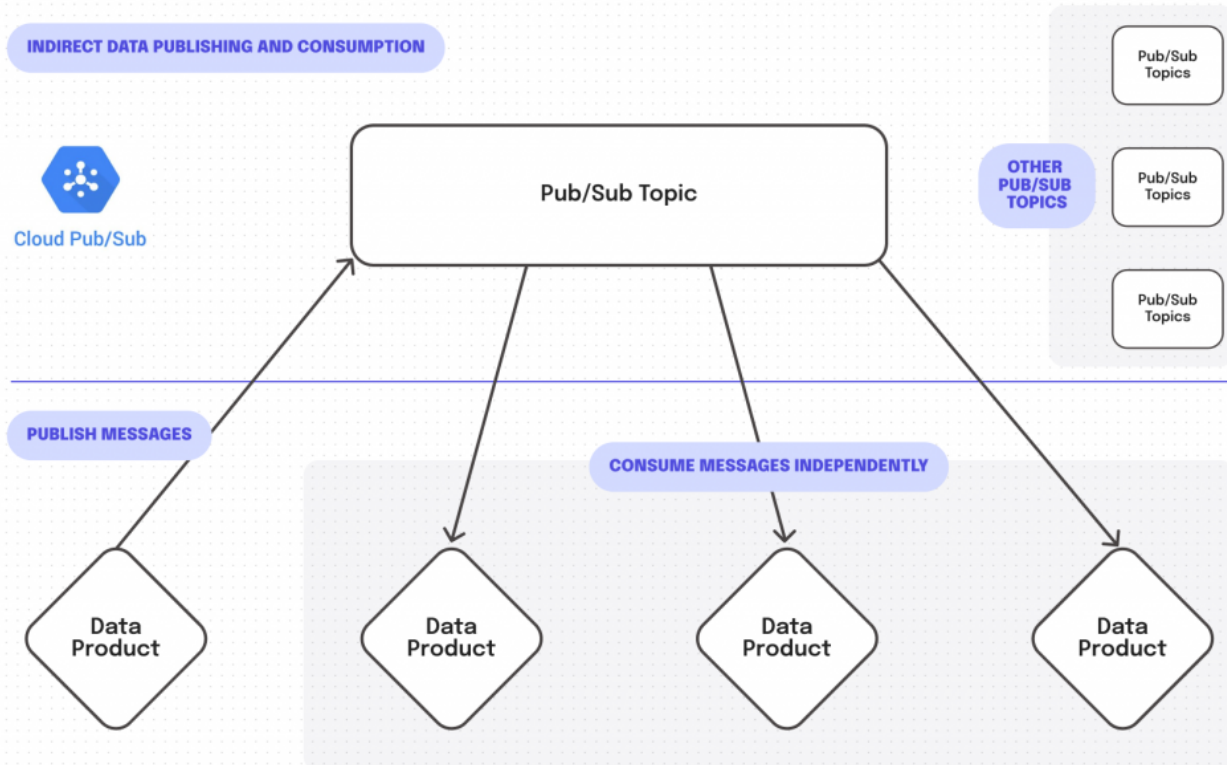

Courtesy of Masthead Data.

The role of each data product is to produce and sometimes to consume data within the Data Mesh. Data Mesh is a lot like the nervous system where each neuron (Data product) continuously receives and sends impulses to other neurons in the system. And the data is continuously flowing, being processed, analyzed, and used for decisions. Everything works as a coherent system, and things go smoothly together.

How to Create a Product Data Feed/Output

Please note that we use the 'nervous system' analogy as a logical view of how data products communicate within the Data Mesh. When it comes to physical implementation, message brokers kick in. Such as Kafka or Pub/Sub. If also worth emphasising that message brokers process data indirectly, not point to point. Meaning that, the way data products communicate inside the Data Mesh is in fact way simpler than the nervous system. Messages are published and consumed indirectly.

Courtesy of Masthead Data.

Here's how this works in reality. Data product communication is powered by message brokers, and data products publish messages to topics where other data products consume these messages. The principle here is a lot like a message board where one writes a message – and many readers can consume it.

Message brokers enable each team to create topics where they publish their data and data consumers can subscribe to topics and consume the data they need. All of that can happen asynchronously, and may not involve a lot of interactions between teams. Message brokers facilitate implementing the data product principle. This, in turn, helps to ensure that the data is reliable and meets quality standards, so it can be consumed by those who need it.

Once you decide to implement Data Mesh in the Google Cloud infrastructure, Pub/Sub is the go-to solution to propagate data:

- it propagates data immediately,
- it can handle large data volumes,
- messages can be stored and recalled,
- Pub/Sub allows to limit recipients and make subscriptions addressable.

Yet message brokers are not the only way to deliver data in and out of data products. You can also go for Res/Req APIs or send Snapshots via ETLs. These options work both ways to deliver data to data products and to send output, though ETLs deliver data in batch and there is less control over the data schema compared to Res/Req APIs.

Now, what happens after you've fed the data to the data product? The responsibility now lies with the team in charge of the data product. One of the core ideas behind Data Mesh decentralization is the data owners are accountable for delivering high-quality data to the destination where data users can access

it. The decentralization principle gives autonomy to each data product team to decide what language to use to propagate data, and which infrastructure to use to store the data. In short, they have carte blanche, when it comes to technology choices within their department. Anything works as long as the output of the data product meets data quality standards of being addressable and reliable.

Best Practices with Data Mesh and Data as a Product principal

- Start with deciding what data should be streamed and what can be delivered in batch – both options are valid. The decision has to be made based on your business requirements rather than a mere pursuit to use new shiny technology.
- Get data from its original source rather than from intermediaries. That has a lot to do with data quality. Do you remember the 'broken telephone' game? That is, the more intermediaries, the more risks of having data quality issues. This can be effectively solved through the event streaming through Pub/Sub by subscribing directly to the source the data originates from. By the way, you can use **Pub/Sub to stream the data directly in Google BigQuery tables (https://cloud.google.com/blog/products/data-analytics/pub-sub-launches-direct-path-to-bigquery-for-streaming-analytics).**
- Avoid the evil practice of fixing data locally at any cost. Instead, make sure that the data comes in a format that is ready to consume for everyone in the organization. Use data constructs to help you govern the data schema and any unauthorized changes related to schema changes and data types.

The takeaway is Data Mesh requires users to embrace a new culture, where data become a central product in their day-to-day work. Data Mesh involves a lot of tech investments and configurations from every department in the organization. But there is something more that should not be underestimated. To succeed with Data Mash architecture - buy-in from all departments and the entire organization's commitment is indispensable. Because every department turns into a data producer and data owner who are in charge of data quality and take full accountability for it.

Post Tags : **Data Mesh**

CEO, Masthead

(htt ps:/ /ww w.lin kedi n.co

(htt ps:/ /ww w.yo utu be.c om/ cha

n.co
m/i
n/y
uliia
_
tka
cho
va/)

nnel
/UC
Pq0
eC1
R6D
9A6
XAa
6tH
JXx
w)

Recent Posts

> Maximizing Data Trust through Data Observability

> No Trust Without Data Reliability, No Privacy Without Data Integrity

> What is Data as a Product, and What to Consider Implementing It?

Follow Us

(htt
ps:/
/ww
w.yo
utu
be.c
om/
cha
nnel
/UC
Pq0
eC1
R6D
9A6
XAa
6tH
JXx
w)

(htt
ps:/
/ww
w.lin
kedi
n.co
m/c
omp
any/
mas
the
ad-
dat
a/)

Stay in touch

Learn more about how Masthead secures your data quality!

Contact Us ()

Start now
(https://app.mastheadata.com/sign-up)

Privacy Policy

(https://mastheadata.com/privacy-policy/)