Open in app ↗   Sign up   Sign In

◐|| 🔍   👤 ⌄

🟦 Published in Google Cloud - Community

👤 Thinh Ha (Follow)
Feb 22, 2022 · 6 min read · ▶ Listen

🔖 Save   🐦   ⓕ   in   🔗

Recipe for building your first Data Product in a Data Mesh

A journey of a thousand miles begins with a single step. For a Data Mesh, this journey begins with a single Data Product.

In this article, I will outline a recipe for building your first Data Product. Like all recipes, you are welcomed to adjust it to your context and preferences.

To bring this article to life, I will also provide an example of what a Data Product architecture looks like on Google Cloud.

This article assumes that the reader is either the business sponsor for the Data Mesh, or someone who is part of the Data Platform team enabling the Data Mesh.

1. Find a business case that can be solved with Data Mesh

First, you need to find a business sponsor who has a specific business opportunity or challenge that cannot be implemented effectively on existing data infrastructure. A few examples are:

- Provide new service offerings through APIs, potentially leveraging predictive analytics

- Build an integrated customer database for improving customer service, marketing, and operations

👏 132 | 💬 1

- Streamlining risk and regulatory reporting

The ROI for solving this challenge or opportunity needs to justify onboarding a small team of full-time engineers to build and run the solution. Ideally, this business sponsor should 1) have a track record in delivering transformative change programmes and 2) tends to be an early adopter in a underline technology adoption lifecycle.

2. Ensure business readiness and acquire funding

Second, you need to get buy-in from this business sponsor to own the development of this new Data Product. It is crucial that they understand the goals of a Data Mesh and their role in building this Data Product. Specifically:

- The goal of the Data Mesh is to empower distributed teams to create business value from data. It aims to create a network effect by treating Data as a Product and providing distributed teams with a self-service Data Platform that has embedded governance.

- The business sponsor owns the business requirements for this new Data Product as well as accountability for the end-to-end product development lifecycle.

- The business sponsor will allocate funding for a cross-functional team to deliver the Data Product. This team is responsible for putting together and delivering on a product roadmap to ensure the Data Product captures value for the given business opportunity. They are also responsible for the day-2 operations as well as any ongoing enhancement and maintenance of the product.

- Progress is measured in terms of business outcomes achieved, and funding for the Data Product team can be adjusted upon business review without a change control process.

3. Prove Data Platform readiness and empower people for success

Third, ensure your Data Platform has the technical capability to onboard this new Data Product. You need to ensure that any technology or infrastructure services required by the Data Product are approved for use within your organisation.

The goal is to reduce friction and lower the barrier to entry for adopting these services. As such, there should be clearly documented security guardrails,

developer guides, runbooks, and architecture blueprints on how to use these services.

Crucially, the service should be provided as a self-service capability. This can take the form of an infrastructure-as-code module to deploy a cloud database, or a templated data pipeline that can be easily adopted and extended by distributed teams. The guidance needs to also cover:

- production deployment process following CI/CD best-practices

- how to meet platform-level security and governance guardrails

- how to integrate with the platform's day-2 operations tooling for logging, monitoring, observability, audit endpoints, etc…

As a Data Platform team, you are successful when a distributed team can deploy and manage a Data Product without needing to involve you in the process. This proof-point can only be achieved once you've onboarded a number of Data Product teams as customers.

At beginning stages, you may benefit from finding a customer who is willing to co-invest to build the Data Platform components they need. Your priority as a Data Platform team is to develop the capabilities for this Data Product in a way that will be reusable by other customers of the platform, while your customer's priority is to ensure the platform is fit-for-purpose for a live Data Product application.

You may also benefit from embedding Data Engineers from the Data Platform team into the business domain team. In this engagement model, the business domain team owns the product backlog, while the Data Engineers provide the technical expertise to deliver on the backlog **as part of the same team**. At the end of the engagement, the Data Engineers can choose to stay in the product team or return to the Data Platform team to start repackaging what they developed for reuse by other Data Platform customers.

It is worth nothing that the Data Platform itself is an internal product, and in the words of Matthew Skelton from Teams Topologies fame, should be a "curated experience for engineers". The Data Platform's goals of empowerment, lowering technical barriers, and minimising collective technical debt are all in service of the

communities using it, such as the Data Product teams, and therefore require constant feedback from its customers and refinement to achieve its purpose.

4. Tell your success story

Developing a Data Product is an incremental process, as such you will benefit from setting intermediate milestones for regularly reviewing progress and the business impact achieved with your sponsors. This is also your opportunity to celebrate any successes and share your success stories with the rest of the organisation in order to attract additional customers to the Data Mesh.

5. Rinse and repeat

With every Data Product, you should aim to shorten the time-to-value for the next sets of Data Products on-boarded onto the Data Platform. The value of the overall Data Mesh should be at least proportional to the combined value of your Data Products. With a larger number of Data Products, the value of the Data Mesh should grow exponentially due to the network effect allowing Data Products to build on top of each other.

As you can guess, adopting Data Mesh may require a huge shift in how you think about Data Culture and Data Operating Model. I have previously discussed cultural and organisational challenges you might encounter when adopting Data Mesh in 10 reasons you are not ready to adopt Data Mesh. The goal of this article is to balance out the message from the previous article with a key message: **you can start small and evolve into the full Data Mesh, one Data Product at a time!**
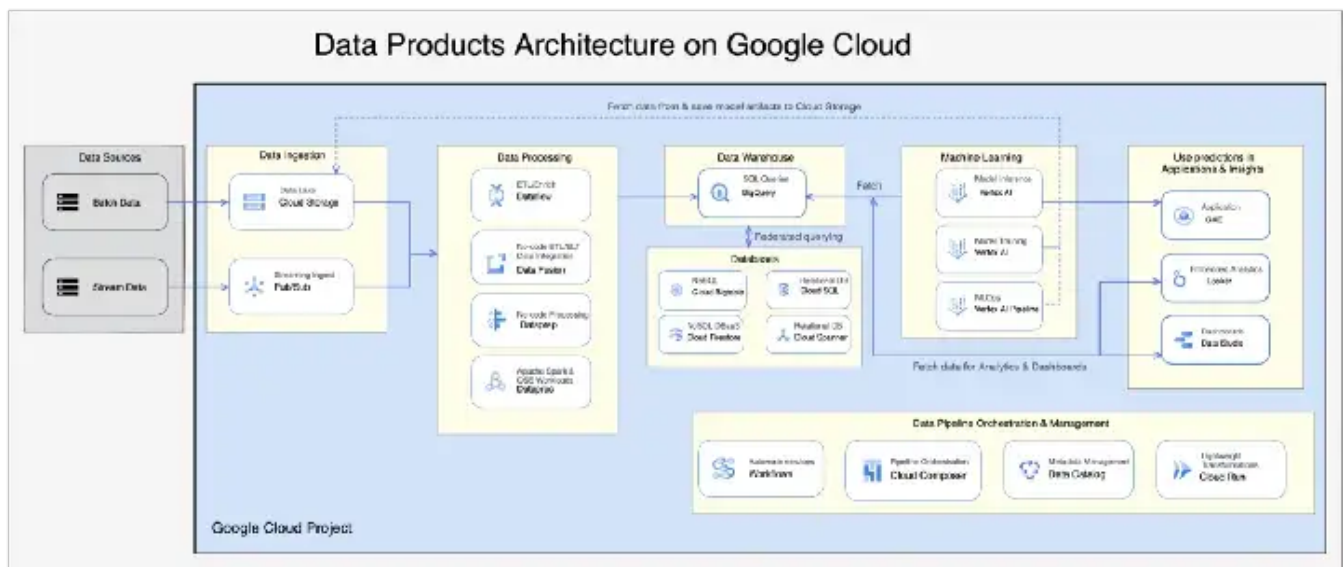
Example Data Product Architecture on Google Cloud

Many organisations have built successful Data Products on Google Cloud, leveraging our comprehensive suite of serverless and integrated data services such as:

- BigQuery for serverless data warehousing

- Dataflow for unified batch and streaming data processing

- **Dataproc** for **serverless Spark** and for running open-source Data Analytics applications

- **Looker** for embedded analytics and visualisation

- **Vertex AI** for end-to-end MLOps

- **Cloud Composer** for a managed **Apache Airflow** workflow orchestration service

- **Cloud Spanner** for a serverless and battle-hardened relational database service

The diagram below illustrates a typical Data Products architecture on Google Cloud. The diagram was taken from **Google Cloud Architecture Diagramming Tool**, courtesy of **@pvergadia**.

In this architecture, each Google Cloud tenant has their own **project(s)** where they can manage the end-to-end lifecycle of their Data Products, leveraging deployment templates and guardrails provided by the Data Platform team. Most of the services are serverless so there is minimal infrastructure to manage, allowing Data Product teams to focus on generating value from their data. To ensure global interoperability and governance, Data Platform tenants can additionally benefit from the following horizontally integrated services across the Google Cloud data stack:

- **Analytics Hub** for zero-copy data sharing with BigQuery

- **Dataplex** for unified governance across Data Lakes and Data Warehouse environments

- <u>Cloud IAM</u> for unified access management across all Google Cloud services

- <u>Cloud Operations Suite</u> for integrated logging and monitoring

- <u>Cloud Asset Inventory</u> for auditing assets across all projects and services

For more information on how to build a Data Mesh on Google Cloud, please see the new <u>Google Cloud Data Mesh Whitepaper</u> published as part of the Dataplex GA launch.

Hope you found the article helpful. Is it missing something? Let me know what you think by reaching out to me on Twitter at <u>@thinh_ha</u> or on <u>LinkedIn</u>.

Special thanks for my colleagues Akhilesh Singh, Anant Vikram, andRubén Fernández who helped review this article.

Data Mesh   Data Engineering   Data Analytics   Data   Google Cloud Platform