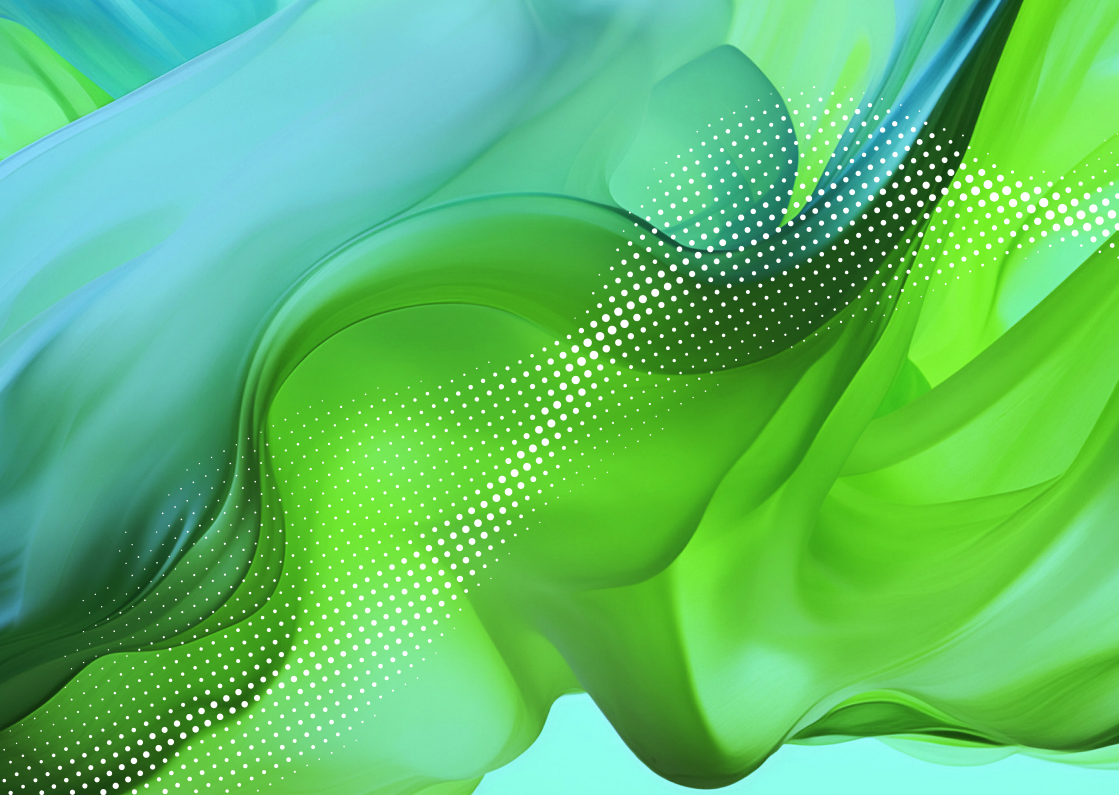


Unlocking the full value of Data as a Product with Microsoft Fabric and Purview

thoughtworks | Microsoft

Executive summary

In 2020, Thoughtworkers introduced the world to a revolutionary new concept called Data Mesh. The thinking and philosophy behind the architectural paradigm have proven extremely valuable for all kinds of organizations — especially the principle of Data as a Product (DaaP).

Managing DaaP supports higher data quality, makes data more accessible and discoverable, streamlines innovation, and simplifies data governance across diverse domains.

This whitepaper outlines a number of lessons learned and best practices derived from years of experience building data products and modernizing data foundations. The whitepaper also shares secrets that can help your organization build and scale high impact data products using Microsoft Fabric and Purview platforms.

What you'll learn:

1. What data products are, and how they can unlock and accelerate innovation.
2. How to work backwards from your desired outcomes to create impactful data products — and prove their value.
3. The pre-requisite data platform capabilities proven to improve and accelerate data product development.
4. Approaches that enable data product teams to become good stewards of their products.
5. What good looks like: Thoughtworks' data modernization reference architecture that demonstrates how Microsoft Fabric, Purview and Azure can be leveraged to create clear distinctions between data platform capabilities and data product acceleration.

Questions answered in this whitepaper:

- As a data scientist, how can I more quickly experiment and find impactful solutions in our existing data?
- As a data platform engineer, how can I drive experimentation and help scientists generate more impact with the data we have?
- At the same time, how do I preserve the quality, integrity, lineage, and proper governance over our most valuable asset, our data?
- What common pitfalls should I prepare for or avoid when considering Data as a Product?

The guidance in this whitepaper will help establish a model for success which separates platform and product responsibilities. This paper will set your teams on a path that enables data product teams to build their own data products. At the same time, it will prepare the data platform teams for the efforts needed to consistently manage, govern, and operate your data estate.

Thoughtworks' experience plus the features of Fabric and Purview make these objectives attainable and realistic. We hope this paper demonstrates these statements and sets you on a path to be successful for data products in your enterprise.

How Data as a Product unlocks and accelerates innovation

DaaP is a style of data management which — as its name suggests — helps organizations manage and structure data in the same ways they build and maintain products. It applies the principles of product thinking to data, treating data users (both internal and external) as customers, and ensures the data available to them meets their needs and helps them achieve their goals.

In organizations that have adopted DaaP, teams start with a customer need or desired use case, then work backwards to build a data product that meets that need. So, rather than simply providing access to wide data sources that customers can hopefully use to solve their challenges, they're empowered with purpose-built products — helping unlock the business value of their data.

What is a data product?

Data products are processed, analyzed, and curated data sets designed to provide solutions or insights, and packaged as self-contained units with their own independent life cycle. Typically, they contain both the data required to meet a specific use case, and any supporting resources (such as models or capabilities) required to execute the use case they're built for.

A data product in Fabric could be a workspace, where people can request access to the dashboards or data it contains, through the Purview interface, which would also hold the documentation and business context.

Practically, DaaP is all about creating domain-aligned, cross-functional data product teams with a clear focus on business value. With domains in control of their own bespoke data products, and responsible for their daily management, core data platform teams can narrow their focus to providing the tools those domains need to seamlessly produce and consume data products.

That's where self-service data platforms come into the picture. Platforms bring essential capabilities together in ways that support domain workflows and help everyone achieve their unique data-driven goals. They enable seamless developer experience, and easier data discoverability allows data product teams to accelerate business value creation and innovation.

Self-service data platforms ensure that data products are:

- **Discoverable**, so that users within their core domain and across other domains can easily find them and apply them to new use cases to maximize their value.
- **Standardized** to ensure interoperability and make it as simple as possible for data products to be combined to unlock new value.
- **Easy for teams to utilize, build and manage** without specialist platform engineering skills, empowering as many people as possible to use them.
- **Of a consistently high quality**, by applying guardrails at the platform level that govern what data products can look like and ensure all domains adhere to the same standards.

That's why it's so important to get your platform and its underlying strategy right. Your choice of foundation and enabling capabilities can have a huge impact on your developers' and domains' ability to innovate and unlock the full value of data.

They define everyone's experience of working in a DaaP-enabled organization — and ultimately shape the value and impact you realize from adopting DaaP.

Best practice at a glance: Working backward from your use case to design data products

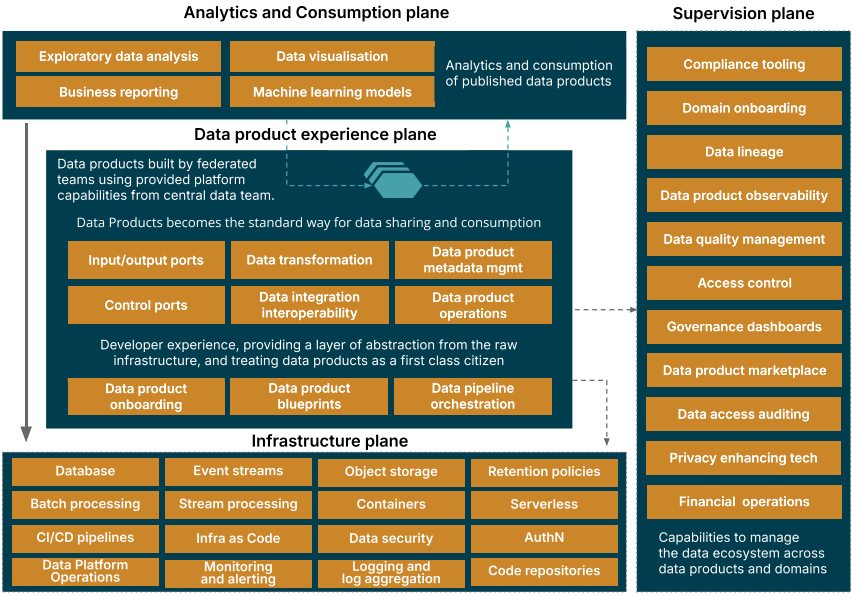
The success of DaaP hinges on your ability to work backward from real business use cases and challenges toward the creation of new data products. For a lot of organizations, this represents a significant shift in data culture. Instead of thinking first about your data and how it can be connected and integrated, you need to begin with a clear view of the use case you want to enable or the problem you want to solve.

For example, in an industrial manufacturing organization, a team may decide that it wants to improve the performance and output of a specific piece of machinery, and the processes that surround it. By starting with that vision, data leads can devise the best way to make that happen, and bring together only the most relevant data sources. This refines project scope, keeps everyone aligned to a single, tangible goal, and ensures that what's delivered at the end is fit for purpose and generates high ROI.

Inside the multi-plane data platform

Self-service data platforms have a lot of different customers. To ensure they can meet the diverse needs of the unique stakeholder groups that depend on them, we break platforms down into multiple planes — each with its own purpose, and the capabilities required to deliver on that purpose.

Here's what that looks like in practice:



The **Analytics and Consumption plane** ensures that data across the platform is visible, and easy to turn into actionable insights. It comprises essential capabilities for data analysis, visualization, and reporting, including leading capabilities powered by AI and machine learning.

The **Data Product Experience plane** is where data products are built, maintained, and orchestrated. The capabilities here ultimately define the experience that developers across

diverse domains will have when creating data products for their unique use cases, and accessing existing data products for consumption.

The **Supervision plane** keeps everything standardized, interoperable, secure, and compliant. This is where the rules that govern how the platform can be used and what data products should look like are set and upheld.

The **Infrastructure plane** — as its name suggests — contains the core infrastructure required to build and operate the platform.

The multiplanar approach to data platform building reminds us that any platform we build to support DaaP must meet the needs of a lot of different groups across an organization. Ideally, it will support those groups to work completely autonomously from one another.

When the four planes are combined, the result is a self-service data platform which meets the needs of all its user groups, provides them all with intuitive experiences, and reduces their dependence on one another to complete common data tasks. Through that combination of autonomous empowerment and a high standard of user experience, the platform plays a critical role in accelerating data-driven innovation across the business.

Best practice at a glance: Separation of concern between data product and data platform teams

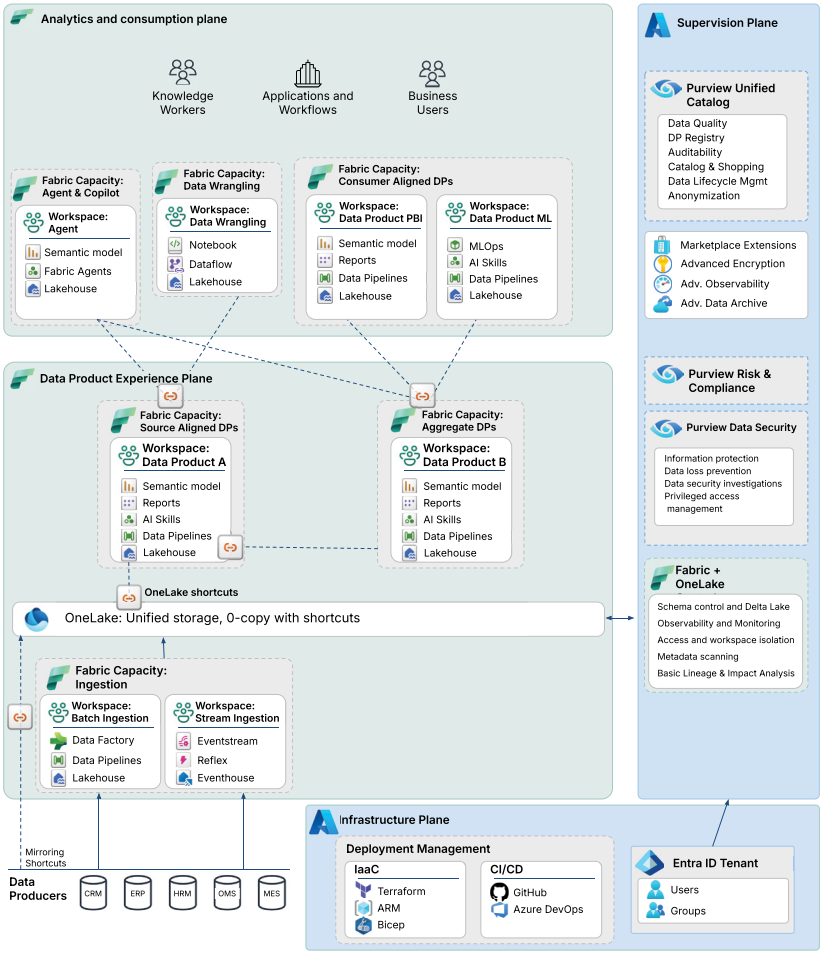
In environments that embrace DaaP, data products and data platforms are built by very different teams. While organizations still have core data teams responsible for their data platforms and the overall experience of engaging with them, the platform itself empowers individual domain teams to build data products, rather than simply consume them.

For many organizations, this demands a significant shift in mindset when building a data platform. Stakeholders outside of core data teams aren't just consumers of these platforms — the platform empowers them to become developers and owners of data products. So, it needs to deliver a great and intuitive developer experience, as well as a strong consumer experience.

Implementing DaaP with Microsoft Fabric and Purview

When they're constructed correctly, data platforms enable the seamless creation and management of data products. At Thoughtworks, we've built many such platforms over the last five years and have gained a deep understanding of which capabilities deliver the best results for organizations and the diverse domains that engage with data products.

Along our journey, Microsoft Fabric and Purview have emerged as especially robust and well-suited solutions for supporting Data as a Product. In this section we'll take an in-depth look at why, and explore how some of the specific capabilities they deliver can help organizations unlock the full value of DaaP.



Microsoft Fabric: The foundation of multi-planar data product environments

The Fabric workspace naturally lends itself to the creation and management of data products.

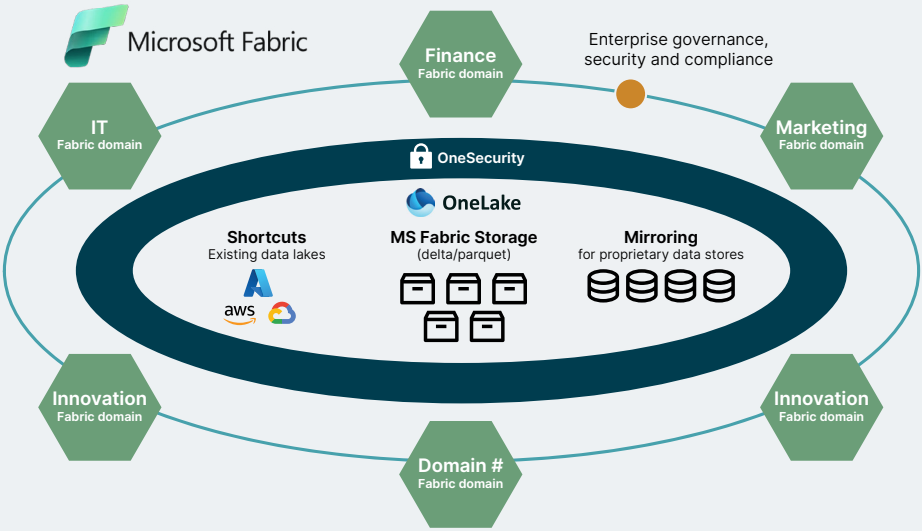
Fabric supports the creation of autonomous and empowered workspaces, with OneLake as the main data storage and sharing layer between them. Rather than continuously duplicating data from the core lake to individual domain environments, Fabric provides sharing capabilities that create shortcuts for data, allowing each workspace to see relevant data in their context, without duplication. This shortcutting ability works across tenants in Azure, with external data sharing, and even across providers, with support for AWS S3 and GCP cloud storage.

OneLake shortcuts are a key Fabric feature in enabling the interoperability and data sharing across data products.

Fabric has advanced capabilities for governing access control and security. The governance is complemented seamlessly by Purview. Purview has both a native understanding of the concept of data products and provides workspace scoped scanning capabilities to match.

Domains in Fabric

One of the biggest reasons why Fabric is so well-suited to supporting Data as a Product is that it enables organizations to easily build domain-oriented data environments, as shown in the diagram below.



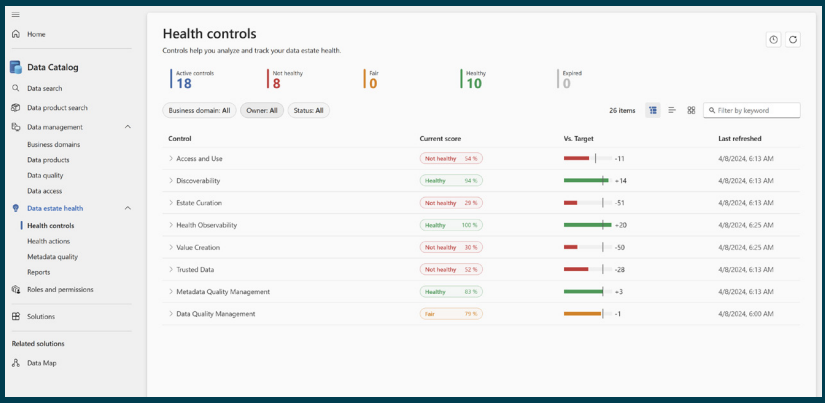
This ensures every domain has their own workspaces, complete with the specific capabilities they need to both construct their own highly relevant data products, and consume those data products in ways that deliver high, specific value for them.

Data domains are also a native concept to the Microsoft Purview Unified Catalog that enables a seamless experience between the different planes that share concepts.

Best practice at a glance: Use a consistent meta model for data products to drive alignment across domains

While data products serve different needs and are built by individual domains, to unlock their full value they must be interoperable and consistently constructed, so other domains can apply them to new use cases.

A consistent meta model and catalog — such as the ones provided by Microsoft Purview — ensures that data products remain consistent, visible, and interoperable across domains. While each domain has control over their products, the catalogue and meta model ensure that the entire organization can see that data, and stay aligned to a consistent data standard by offering estate and domain scoped compliance reports, ensuring the standards are followed.



Microsoft Purview: Supporting compliance and governance

Purview Unified Catalog allows data product users to discover them, and their relationships. It also enables user access management to the data.

Purview Unified Catalog provides the vital automated governance layer needed to enable efficient domain-oriented data management, continuously upholding standards and governance requirements to ensure all data products are created and consumed in safe, secure ways. But it isn't just concerned with security. With native support for data products, Purview also provides a view of data product health, enabling the continuous improvement of data product quality, and the quality of outputs and outcomes data products are used to deliver.

Purview Unified Catalog helps you monitor domains and the whole data estate, providing the tools to scale the usage of data products to large enterprises. Combined with advanced data classification capabilities, this guarantees the interoperability of the different domains on the semantic business level, where Fabric guarantees technical interoperability.

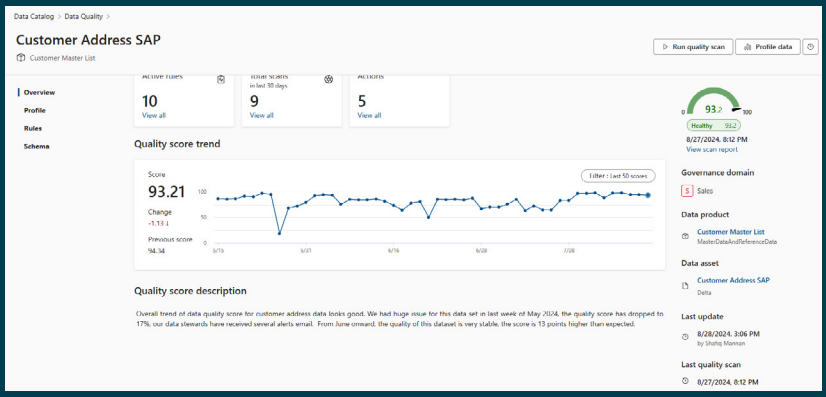
Purview Data Security ensures information classifications are enforced, and provides a seamless integration in Fabric like the integration for Microsoft 365, making sure data is safe across the entire Azure cloud.

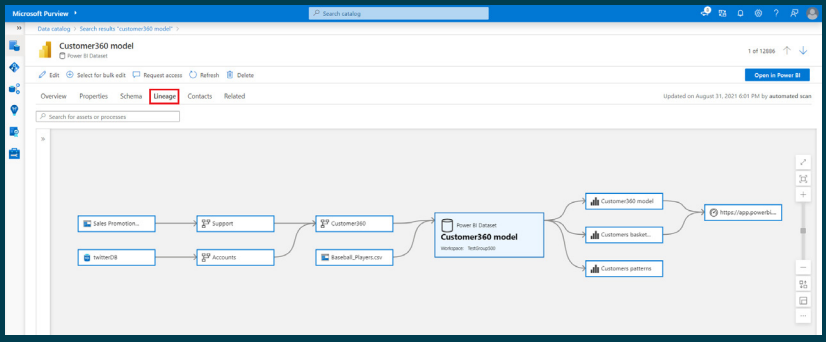
The end result is a robust governance layer that natively supports data products, and is tightly integrated with both OneLake and Microsoft Fabric — creating the ideal environment for teams that want to make the most of what DaaP can deliver, and prepare their data estate for the AI era.

Best practice at a glance: Make data product metrics, quality, and lineage visible to all

Managing data as domain-owned products marks a major cultural change for many organizations. As you make the shift, you need to focus on building trust in your new platform, environment, and data products.

To help, use capabilities like Microsoft Purview to surface reliability metrics and data product lineage data. With them in place, anyone can learn anything they need to about data products before using them, with total confidence in what that data can help them achieve.





Supercharging data with Gen AI

One of the most compelling reasons for embracing Data as a Product today is that domains can create and curate high quality data products designed to support AI and ML workloads. But with Microsoft Fabric and Purview, AI-powered capabilities can be embedded at the platform level, increasing productivity for the diverse teams responsible for data products and their consumption.

In addition to supporting the creation of custom Gen AI experiences based on your data, Microsoft Fabric provides built-in Copilots that provide productivity boosting Gen AI experiences for a wide range of data-centric roles. It currently offers Copilots to support tasks based in:

- Data Factory
- Data Warehouse
- Data Science
- Real-time Intelligence
- PowerBI

With the help of this Gen AI copilot, teams are empowered to spend less time on routine data product management tasks, and more time focused on extracting the maximum possible value from their data.

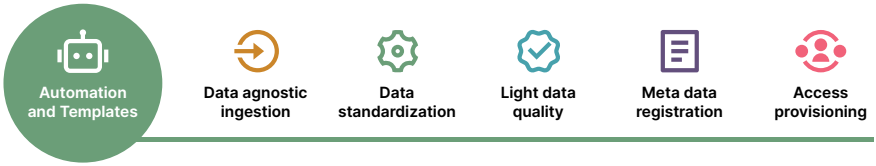
Microsoft Purview takes that capability a level deeper, adding AI-powered features such as:

- **AI-enabled product search**, which enables the rapid searching of existing data products using natural language. Teams from any domain can simply ask for the data they need, and instantly locate and start using it.

- **Data product Copilot**, which helps teams perform a wide range of product-centric tasks and accelerate routine data product management activities.
- **Automatic data quality** optimization suggestions that help teams clearly see how to improve the quality of their data products using new rules that can be turned on with a flip of a switch.
- **Data product adding assets** which streamline the process of adding new data products in a compliant, standardized way.

Kickstart your DaaP journey with proven expertise and accelerators

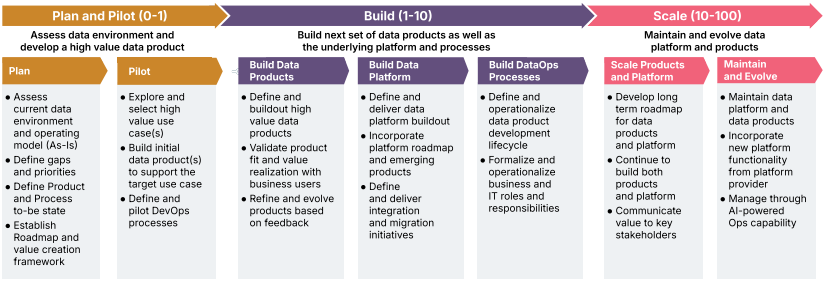
The journey from embracing Data as a Product to building and orchestrating your new data platforms and environment can be complex. But it's one that Thoughtworks has successfully navigated countless times for clients across diverse verticals.



With deep Microsoft Fabric, Purview, and Azure expertise, and a proven set of IPs and accelerators, we can help you realize the full potential of Data as a Product. As the team behind many of today's leading DaaP best practices, we can help you:

- **Design unique workflows and architectures** to support how your domains work, and enable cross-functional innovation in even the most siloed organizations.
- **Build and implement the right combination of Microsoft products and services** for your organization, and align the tools available with your business and data goals.
- **Evolve your data culture** alongside your technology and empower your people to make the most of the data products available to them, and become passionate data product custodians.
- **Automate complex and time-consuming tasks** like data ingestion and standardization to accelerate time to value and help you build valuable data products at speed.
- **Harness the power of AI** by embedding it into your data workflows at the platform and capability level, and enabling you to bring your most ambitious AI use cases to life.

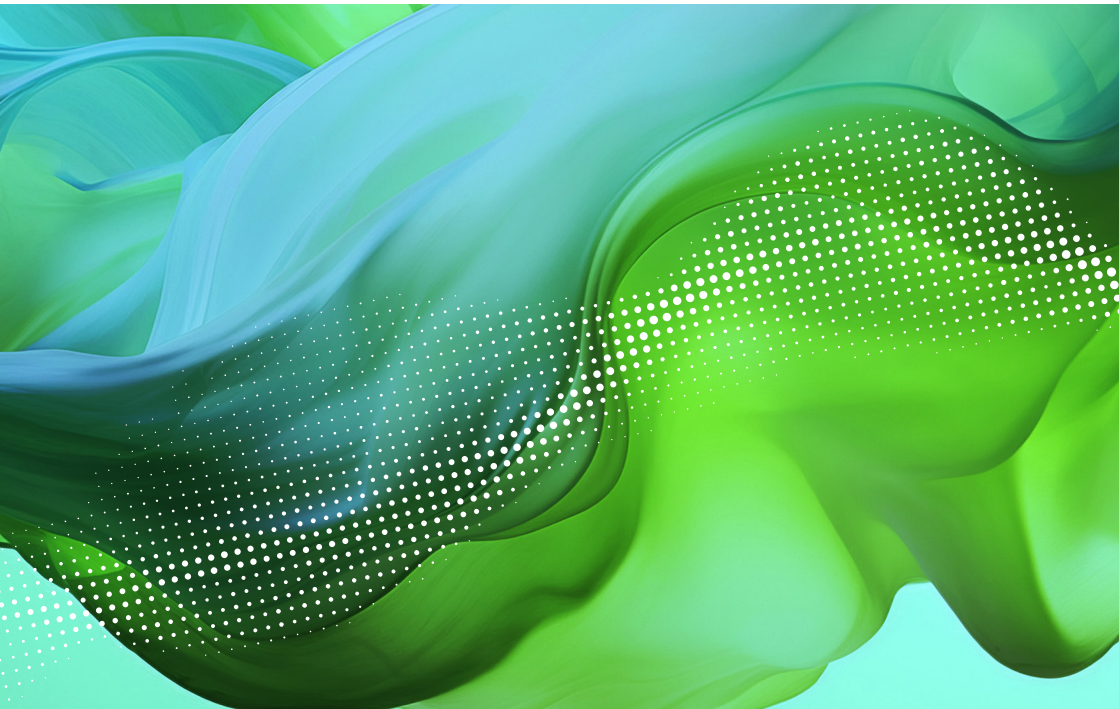
To make it as simple and cost effective as possible for our customers to plan, create and scale data products within their organization, we follow a three-horizon plan as outlined below.



It starts with an eight-week plan and pilot phase where we work with clients to help them define a high level data strategy, select an initial use case for DaaP and map out a corresponding set of data products that can help support that use case. In a maximum of just eight weeks, we can prove the value of using Microsoft Fabric, and make a robust case for scaling use cases out and building further data products.

Throughout this process we also help clients unlock partner funding from Microsoft, making it a highly efficient and accessible way of adopting DaaP. Pre-built accelerators help you navigate your journey as quickly and predictably as possible, while unlocking unique value for your organization.

To find out more about our data mesh and DaaP services and offerings, and start your own three-horizon DaaP journey with Thoughtworks and Microsoft, visit this page, or contact us today.



Contributors

This whitepaper has been created by Thoughtworks, in partnership with Microsoft. Key contributors to the project were:

Prashant Atri
Senior Cloud Solution Architect Data & AI
Microsoft

Alice Crohas
Microsoft Alliance Manager Europe
Thoughtworks

Lauris Jullien
Lead Consultant
Data & AI
Thoughtworks

Kiran Prakash
Principal Consultant
Data & AI
Thoughtworks

Brian Blanchard
Global VP of Cloud
Thoughtworks

References

The following resources have been used to inform and support the content within this whitepaper:

martinfowler.com: Designing data products

Microsoft whitepapers:
Open Lakes, Not Walled Gardens
Modern Analytics and Governance at Scale

We are a global technology consultancy that delivers extraordinary impact by blending design, engineering and AI expertise.

For over 30 years, our culture of innovation and technological excellence has helped clients strengthen their enterprise systems, scale with agility and create seamless digital experiences.

We're dedicated to solving our clients' most critical challenges, combining AI and human ingenuity to turn their ambitious ideas into reality.

thoughtworks.com

/thoughtworks

Design. Engineering. AI.