K2 VIEW.

Book a Demo

What is a Data Product?
A Practical Guide

This guide discusses data products and product-driven data management, addressing the questions what is a data product, where does it help, and how to get started.

INTRO

Data Products – A Key Principle of Data Mesh

The concept of "data products" was first introduced as a core component of the data mesh architecture and operating model. Data mesh introduces the following 4 principles:

1. **Decentralized, domain-oriented data ownership**
2. **Data as a product**

K2 VIEW.

Book a Demo

The second principle suggests that for a distributed data platform to be successful, domain data teams must apply product thinking to the datasets they deliver – considering their data assets as their products, and the rest of the organization's data consumers, as their customers.

Analyst firm Gartner describes data mesh as an architecture designed with "the specific goal of building business-focused **data products**".

Data products are also relevant for centralized data management architectures, such as data fabric, where data products are created, managed, and adapted by central data teams for consumption by authorized data consumers across the company.

This paper first defines data products and their attributes, and then goes on to discuss the market need, examples, the evolution from project- to product-driven data management, the advantages of, and prospects for data products, and how to get started.

Table of Contents

Please select Chapter ▼

CHAPTER 01

The Business Need for Data Products

Today, data is produced at an unprecedented rate, due to the staggering amount of digital services and offerings, combined with ubiquitous Internet connectivity. At the same time, data is a company's most important asset, and critical to business success.

According to McKinsey, data-driven companies are:

- **23x** more likely to acquire customers, and

- **19x** more likely to be profitable.

K2
VIEW.

Book a Demo

- **Fragmented** in 100s of systems

- **Captive** in vendor-owned applications that lack a rich API set for data access

- **Locked** within in-house legacy systems, with little or no knowhow of the underlying data models

- **Variably structured**, or unstructured, in dozens of technologies and formats

- **Non-compliant**, containing sensitive personal information, which must be anonymized to adhere with regulations (GDPR, CPRA, LGPD,…)

The end result: over 80% of enterprise data remains "in the dark", in the sense that it is inaccessible and unleveraged. This dark data is not driving business decisions nor is it being used to improve customer experiences or operational efficiencies. It is weighing companies down.

CHAPTER 02

What is a Data Product?

A data product is a reusable data asset, engineered to deliver a trusted dataset, for a specific purpose. It integrates data from relevant source systems, processes the data, ensures that it's compliant, and makes it instantly accessible to anyone with the right credentials.

A data product shields data consumers from the underlying complexities of the data sources, decoupling the dataset from its systems, making it discoverable and accessible as an asset.

A data product generally corresponds to one or more business entities (customers, suppliers, devices, orders, etc.) and is made up of metadata and dataset instances:

Data product metadata

- **Static metadata**, including the tables and fields used to capture the data product's dataset

K2
VIEW.

Book a Demo

- **Business logic** to process, mask, and enrich the raw dataset, prior to delivering it

- **Data governance policies**, to ensure the dataset's quality and privacy are enforced according to internal and external regulations

- **Active metadata logs**, for capturing data product performance and usage statistics

- **Access controls**, including authentication and credential validation

Data product dataset

- **Managed as a unit**, simplifying data processing and access

- **Always-fresh**, clean, and compliant – integrated, cleansed, masked, and enriched

- **Stored**, cached, or virtualized

- **Automatically audited**, to log every access and change to the dataset

- **Accessible** by any authorized data consumer

The data product is built, versioned, tested, deployed, and monitored, to ensure that it continues to serve its customers, the data consumers.

CHAPTER 03

7 Examples of Data Product Use Cases

Data products are engineered to drive specific business outcomes, through operational and analytical workloads. Here are 7 examples of data product use cases:

1.

4. **Tokenizing** sensitive customer data prior to AI/ML analysis

5. **Delivering** a consolidated, real-time, and holistic customer dataset to a CRM application, including customer transactions, interactions, and master data.

6. **Publishing** the latest updates on the spread of COVID-19 to an HMO's patients in high-risk areas

7. **Moving** a legacy application's data into a new cloud computing environment safely and quickly, while ensuring business continuity

While data products are often associated with analytical workloads, they are vitally important to a company's operational workloads.

CHAPTER 04

Operational Data Products

such as a customer, vendor, or order — or anything that's important to the business. An operational data product moves data between sources and targets, in both directions, and in fractions of a second. And it can selectively store data, to act as an operational datastore, when necessary.

What makes an operational data product so special, is that its dataset is always:

- **Unified**, and complete, for any business entity

- **Up-to-date**, and enriched with operational intelligence

- **Protected**, compliant with privacy regulations, and properly governed

- **Accessible**, in real time, via data services, and a wide range of data delivery methods

In a data tokenization use case of operational data products, Comcast deployed K2View Data Product Platform, enabling business domains to build, publish, and maintain data products. Authorized data consumers across the company, can auto-discover data assets using the platform's data product catalog.

In this implementation, each data product manages and persists the dataset for each individual customer, in its own high-performance Micro-Database™ – or mini data lake. In the case of Comcast, the platform manages over 30M Micro-Databases, one for each customer.

Comcast created a data product to tokenize sensitive data, where the tokens for each customer are persisted in the customer's specific Micro-Database, each secured with its own 256-bit encryption key. In a sense, the Micro-Database becomes a "mini-vault", with zero risk of a mass data breach.

Micro-Databases are foundational for operational workloads because they're always:

- **In perfect sync**, with all underlying data sources

- **Secured**, each with its own 256-bit encryption key

- **Compliant**, in the context of data privacy regulations, with any sensitive data tokenized or dynamically masked on the fly

- **ACID-compatible**, so that they can transformed into databases of record for new apps

- **Accessible**, in milliseconds, by authorized data consumers across the enterprise

- **Federated:** Making operational datasets easily accessible across the company

- **Tightly governed:** Providing a single, trusted, real-time view of any business entity, to anyone with the proper authorization

CHAPTER 05

The Data Product Lifecycle

Data-driven enterprises have one thing in common: they build data products, as opposed to one-off data projects. Data products are reusable assets focused on business outcomes.

Every data product follows a lifecycle, similar to that of a software product, to iterate and assure that it delivers the desired business outcomes. It looks something like this:
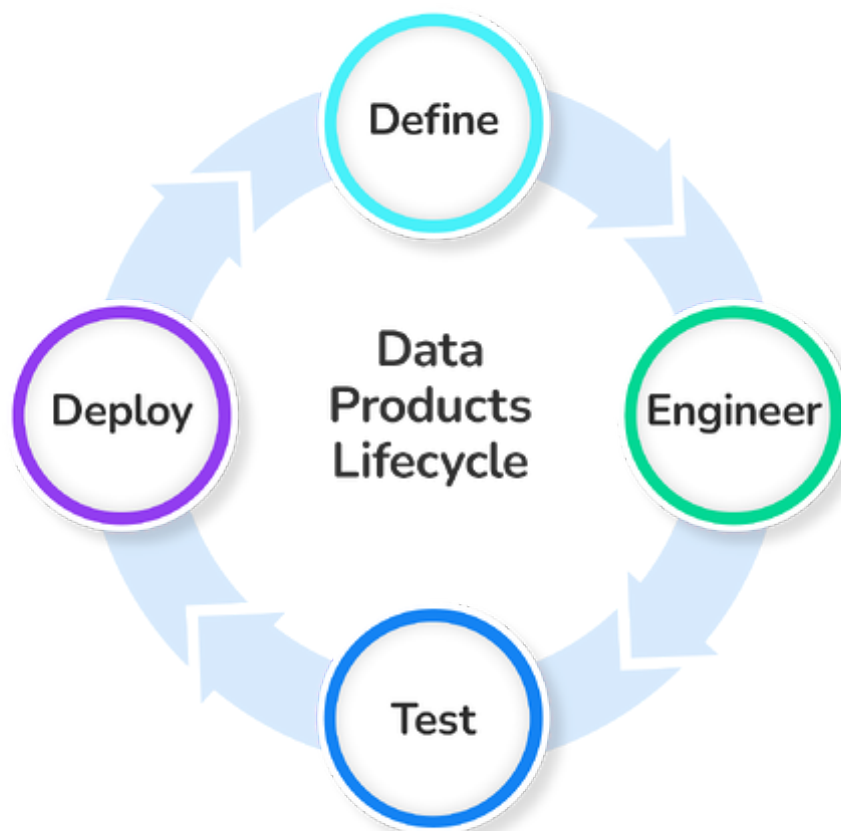
2. **Engineer**

A data product is engineered by locating, collecting, and integrating the source data, and then processing it as needed. Data services are created to provide consuming applications with access, while data pipelines are designated to prepare and deliver the data to authorized analytical data consumers. The data product is versioned and designed to comply with performance SLAs.

3. **Test**

Data products only add value once they're run in production. But, before that can happen, they must be tested to ensure that the datasets they deliver perform as expected, and are fresh, cleansed, complete, compliant, and ready for high-scale consumption.

4. **Deploy**

The data product is deployed, monitored (for usage, performance, and reliability), maintained, and supported – to quickly address any issues that may arise.

Enter the data product manager

Similar to the software product manager, a data product manager is responsible for delivering business value and ROI from the data products – defining their goals and priorities (together with data engineers and consumers), and continuously working to ensure that the promised value is attained.

Why the cycle?

Data teams are constantly experimenting – implementing new services, deploying them, and monitoring the results. The quicker they go through the cycle, the quicker they learn, and the quicker they deliver incremental value to their customers.

CHAPTER 06

Project vs Product Data Mindset

K2
VIEW.

Book a Demo

typically raises a request with the central data engineering team. That request represents a project to identify, collect, prepare, and deliver the relevant dataset to the business domain. This same pattern is followed every time a new use case emerges, from any domain in the organization.

This "data as a project" approach has some major drawbacks, including slow time-to-delivery, lack of reuse, rigidity, and risk of delivering wrong, and/or incomplete data.
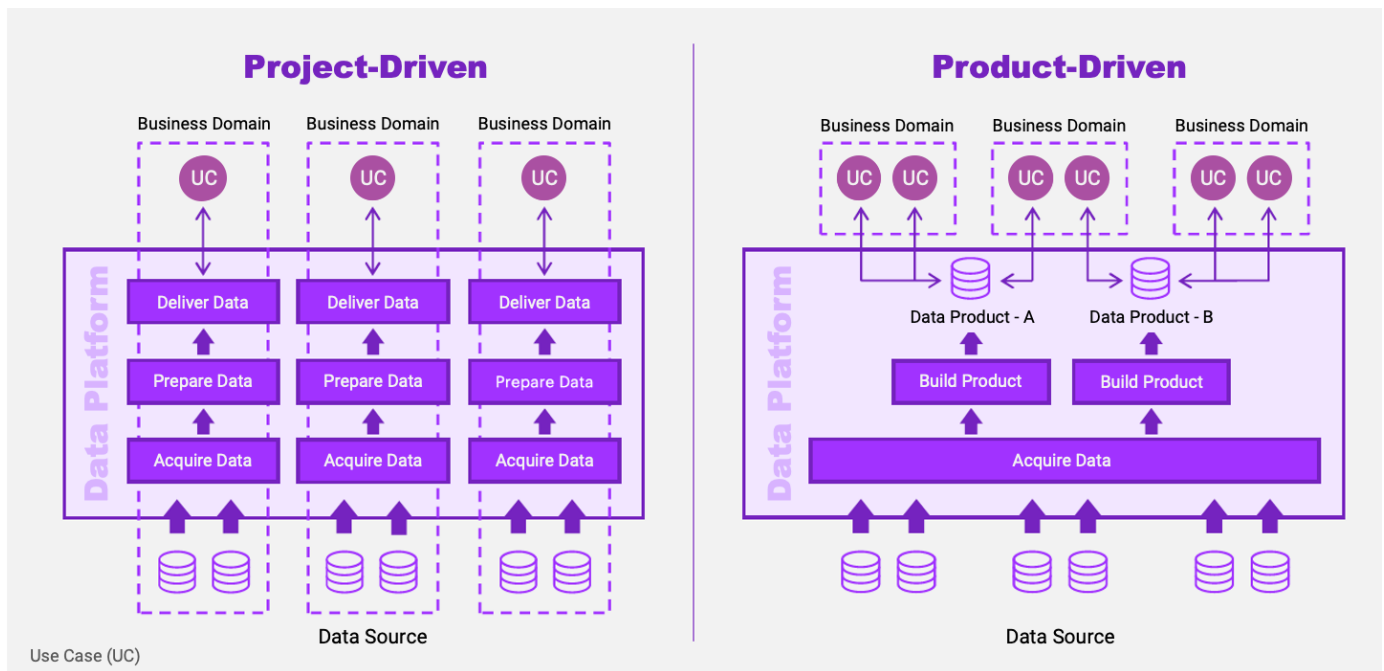
*A **project**-driven approach to data drives greater complexity and minimal reuse, compared to the simpler, more agile **product**-driven approach to data.*

On the other hand, a *product*-driven approach keeps the entire enterprise's data needs in mind. Data products can be reused to support any number of use cases, serving any number of domains.

Advantages of Data Products

Over time, data products deliver better ROI, and cost-per-use, than data projects. Despite some upfront costs, they quickly evolve to support multiple outcomes, addressing emerging use cases – where the focus is always on use case accommodation.

For data consumers, data products offer:

- **Quicker time-to-insight:** Using pre-built data products (instead of initiating a new project)

- **Real-time response times for operational use cases:** Timely and informed decision-making

- **Data governance:** Data is of high quality and compliant

- **Always accessible:** Data is easily discoverable and instantly accessible

For the enterprise, data products are:

- **Business-driven** and outcome-focused

- **Agile**, with value delivered incrementally

- **Built once**, but used again and again

- **Future-proof**, in terms of data architecture

- **Enhancing data trust** and integrity

- **Forging a common language** between business and IT

CHAPTER 07

How to Get Started with Data Products

Deploy the right platform

K2View provides a Data Product Platform to engineer, test, deploy, and monitor data products, in serving a broad variety of workloads.

The platform's Data Product Studio enables data teams to quickly define and maintain the metadata for data products, including the data schema, connectors, sync policies, data transformations, governance, and more.

Once deployed, the data product uniquely manages each dataset instance in its own hyper-

A "Customer" data product collects data from all sources, prepares it, and delivers it to authorized data consumers – end-to-end – in real time.

Appoint data product managers

Data product managers need a broad range of skills in the areas of data, analytics, enterprise applications, business analysis, and DataOps. Ultimately responsible for the entire data product lifecycle, they:

- **Develop data strategies**, determine performance metrics, and promote data literacy across the company

- **Ensure business value** from data, and maximize the Return On Data Investment

- **Close the gap** between business and IT, by communicating the needs of data consumers across business domains and working with data engineers to improve data accessibility.

In the same way a software product manager defines user needs, prioritizes them, and works with R&D to assure delivery, data product managers collect the needs of data consumers, and

Book a Demo

data products within the organization.

Go for flexibility

To maximize flexibility, enterprises should choose a platform that deploys on premises, in the cloud (iPaaS), or across hybrid environments – with support for all modern data architectures.

A data fabric architecture is a modular data management framework, which integrates with your existing data and analytics tools. It assumes that data products are defined by a central data and analytics organization, and adapt over time based on automated analysis of active metadata.

A data mesh architecture shifts data strategy to a federated data network. It gives business domains the autonomy and tools to create data products for their needs, and creates a common framework for building, and scaling, product-driven data solutions, in real time.

There are advantages and disadvantages to data mesh vs data fabric, but both architectures leverage data products as a fundamental construct.

CHAPTER 08

Conclusion

Data products are an emerging data construct, adopted by leading, data-driven organizations. Their value stems from quick discoverability access to trusted data, cutting the time to insights, and driving informed, timely decision making.

Data products fuel operational and analytical workloads, and may be deployed in a data mesh or data fabric architecture - on premises, in the cloud, or in a hybrid environment.

Data teams should seek a data product platform to manage the entire lifecycle of data products,

Book a Demo

Achieve better business outcomes with the K2View Data Product Platform

Book a Demo

K2view

Follow us on LinkedIn   Follow us on Twitter   Follow us on Facebook
Follow us on YouTube

Platform

Data Integration

iPaaS

Data Virtualization

Data Catalog

Micro-Database

Data Governance

Data Orchestration

Data Masking

Book a Demo

Architecture

Data Fabric Architecture

Data Mesh Architecture

Data Hub Architecture

Industries

Telco and Media

Financial Services

Healthcare

Initiatives

Agile Customer 360

Test Data Management

Data Migration

Data Privacy Management

Data Pipelining

Customer Data Hub

Multi-Domain MDM

Operational Intelligence

DataOps

Book a Demo

Data Masking

Test Data Management

Customer 360

Data Fabric

Data Mesh

iPaaS

Data Migration

Legacy Modernization

Data Virtualization

Book a Demo

Customers

News

Resources

Partners

Academy

Careers

Support

Contact Us