



# The Ultimate Guide to Data Products

The comprehensive guide to designing  
and shipping data products your  
company will actually use.



# TABLE OF CONTENTS

---

**01** Introduction

---

**02** What is a data product?

---

**03** Moving from data to data-as-a-product

---

**04** How to design a data product?

---

**05** Data product best practices

---

**06** Data product examples

---

**07** Data product marketplaces

---

**08** Data + AI observability for data products

---

**09** Data product FAQs

---

# Introduction

Data serves as the foundation for businesses today.

A data ecosystem is no longer an afterthought. It's now a fully formed and essential element of a business' technical architecture.

With data continuing to cement its importance in the business tech stack, the majority of teams are adopting a new framework: treating data as a product.

But, what exactly does it mean to shift from “data” to a “data product”? How does a business build a data product? And what is a data product marketplace?

We'll dive into examples, best practices, and strategies to give you everything you need to know about data products.

# What is a data product?

First things first: let's define data products.

**A data product is a critical data asset, such as a key dashboard or machine learning model, that has been developed with the same reliable processes and architecture that would be used to develop a product for an external customer.**

## Benefits of data products

There are several benefits to treating your critical data assets as data products. Let's dive into a few.

### Data products increase trust

By increasing your focus on data quality, you retain trust in the data. Business stakeholders will typically trust the data until something goes wrong, and from then on, trust may be broken. Ensuring you have the data observability and data quality management processes in place to fix data issues fast will enable your team to create a data-driven organization built on trust.

### Data products provide a framework to measure (and report on) the value of the data team

If you're a data engineer, data scientist, analytics engineer, or a similar role, you're likely very close to the data. That means you have an intimate understanding of your technical architecture, how your pipelines are running, and the data they're delivering to the business. But, that intimate data knowledge typically isn't shared by everyone in the organization – especially business stakeholders. Because of that, it can be difficult to explain the value of the reliable, high-quality data you've worked so hard to maintain.

Treating data like a product can enable you to provide hard metrics to demonstrate how your team is performing and contributing to the business. These metrics might include the number of data incidents, time to detection (TTD), time to resolution (TTR), data downtime, the cost of that downtime, and more.

If you're not sure exactly how to calculate the cost of data downtime in your organization, you can [use this calculator](#) to input your numbers and get an estimate.

Below is a [cheat sheet](#) of additional data quality metrics.

Quality Metric	Definition And Calculation	Industry Standard
<b>Number of incidents (N)</b>	<p>An instance of incorrect, missing, or incomplete data.</p> <p>Total incidents (N).</p> <p>Incidents per 1,000 tables per month.</p>	<p>6 per 1,000 tables per month.</p> <p>(37% schema; 23% volume; 28% freshness; 12% quality)</p>
<b>Time to detection (TTD)</b>	<p>The average time from when incidents occur to when they are detected. The time to response—the time between an alert and investigation— can also be a helpful metric.</p>	4 hours
<b>Time to resolution (TTR)</b>	<p>The average time from when the issue has been detected to when it has been given a resolved status.</p>	9 hours
<b>Data Downtime</b>	<p><math>N \times (TTD + TTR)</math></p> <p>It can be helpful to calculate data downtime in the aggregate as well as by specific data products.</p>	793 hours/month*
<b>Data Downtime Cost</b>	<p><a href="#">Our calculator and eBook are the best resources for this.</a></p>	<p>~26% of revenue</p> <p><b>OR</b></p> <p>~\$280K labor cost</p> <p>~\$1 million in efficiency cost*</p>

## Data products improve the team's focus and capacity

Surveys have shown that data engineers typically spend anywhere between a third and a half of their time on data quality. That's huge – and it can feel even more time-consuming when you don't approach your data with a data product framework.



Source: The 2024 State of Reliable AI, Monte Carlo

By treating your data as a product, data teams can automate and expedite a significant amount of the tedious work that goes into fixing data errors and instead, spend more time on innovative work that actually moves the business forward.

Plus, giving self-service access to data products to users across the business can dramatically reduce the amount of ad-hoc questions directed toward the data team.

# Moving from data to data-as-a-product

Simply put, data is just data until it's given context. Data becomes a data product when it takes on the framework that ties data directly to the business. Data products like these might take the form of tables, reports, machine learning models, and other key assets used by the business.

Sounds simple enough, right? Well, this shift from data to data-as-a-product also needs to be accompanied by a mindshift change: the data must be treated with the same diligence as production-grade software products in order for your data and AI products to realize their full potential.



You can't just declare that your data is now a data product, just like you can't declare bankruptcy. Just ask Michael Scott.

Unfortunately, you can't just declare that an executive dashboard or internal machine learning model is now a data product. To turn your data into a data product, you need to make the necessary changes to your data's governance and reliability frameworks.

In the next section, we'll dive into the changes you need to make to go from data to data as a product.

# How to design a data product

To effectively turn your data into a data product, consider the following steps:

- 01. Create SLAs** to ensure different engineering teams and their stakeholders are confident they're speaking the same language, caring about the same metrics, and sharing a commitment to clearly documented expectations
- 02. Assign ownership and accountability** within the data team at the organizational, project, and pipeline/table levels
- 03. Establish documentation** to enable discovery and self-service across the organization (pro tip: document your key assets first, then iterate, automate, and measure your documentation levels accordingly)
- 04. Factor in data compliance** (for external stakeholders) and data governance (for internal stakeholders)
- 05. Certify your datasets** by taking tiered SLAs (gold, silver, bronze, and occasionally, platinum) and assigning them to data assets
- 06. Assess and demonstrate the value** of your data product with KPIs that measure the breadth of customers and applications, how you're building for the future, how you're delivering business impact today, and data quality.
- 07. Consider DataOps and Agile methodologies** to merge data engineering and data science teams to make it easier for data engineering teams to provide analysts and other downstream stakeholders with reliable data to drive decision making.

# Data Products Best Practices

When you're building a data product, it's important to keep in mind a few key essential truths.

Every organization – and every data product – is different, so keep the following best practices in mind to stay on track and keep your team from getting overwhelmed.

**01**

## Everything doesn't have to be a data product.

Don't minimize the value of focus. Sometimes you just need to spin up a quick dashboard and that's that!

**02**

## Don't do everything at once.

Building a data product takes time and effort. Start small, launch an initial version, then gather feedback and iterate. Don't waste time building something no one will use.

**03**

## A data product doesn't have to be a single, universal source of truth.

Remember, data products are meant to enable specific teams to accomplish their specific goals.

**04**

## You don't have to be decentralized to have a data product.

While data products are one of the core principles of the data mesh, you can build a great data product with a centralized data team.

# Data Product Examples

As we mentioned earlier, data products can take many different forms depending on the needs of your business, industry, customers, and more.

What do these different data products look like in action? Here are a few examples:

- An airline's flight tracking system that combines real-time GPS data, flight manifest tables, and historical arrival and departure information
- A customer relationship management platform syncing data across marketing tools
- An AI algorithm that trains disparate financial and investing data from thousands of sources to forecast future stock returns

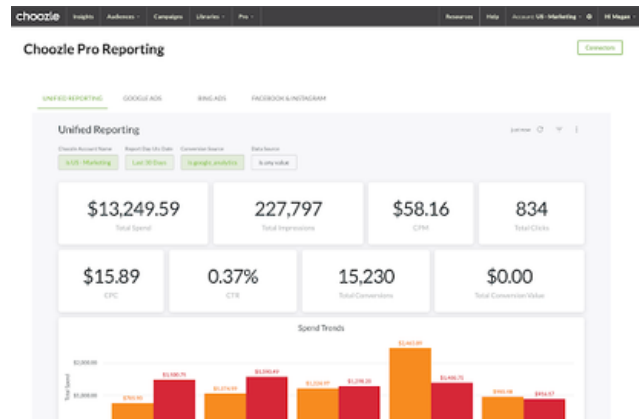
It's essential for data products like these to manage their data quality effectively, and data observability is an ideal tool for the job. Let's look at how Choozle, a leading digital advertising software company, leverages data observability to maintain the reliability of its data products.

Choozle's main data sources are demand-side platforms (DSPs), with which they facilitate the buying of media on behalf of its users. These data sources are static with defined ingest patterns, and are powered by ELT pipelines that bring data into a unified schema where Choozle abstracts the difference in how data is presented between the platforms. That equation changed and a data quality issue arose when Choozle released its massively powerful unified reporting capability, which allows users to connect external media sources.

"When our advertisers connect to Google, Bing, Facebook, or another outside platform, Fivetran goes into the data warehouse and drops it into the reporting stack fully automated. I don't know when an advertiser has created a connector," said Choozle Chief Customer Officer Adam Woods. "This created table sprawl, proliferation, and fragmentation. We needed data monitoring and alerting to make sure all of these tables were synced and up-to-date, otherwise we would start hearing from customers."

Monte Carlo gave the Choozle team deeper visibility into data product issues that otherwise may not have been proactively caught.

“Without a tool like this, we might have monitoring coverage on final resulting tables, but that can hide a lot of issues,” said Adam. “You might not see something pertaining to a small fraction of the tens of thousands campaigns in that table, but the advertiser running that campaign is going to see it.”



**“With Monte Carlo we’re at a level where we don’t have to compromise. We can have alerting on all of our 3,500 tables.”**

**– Adam Woods, Chief Customer Officer, Choozle**

Every organization experiences some level of data downtime, or periods of time when data is partial, erroneous, missing or otherwise inaccurate.

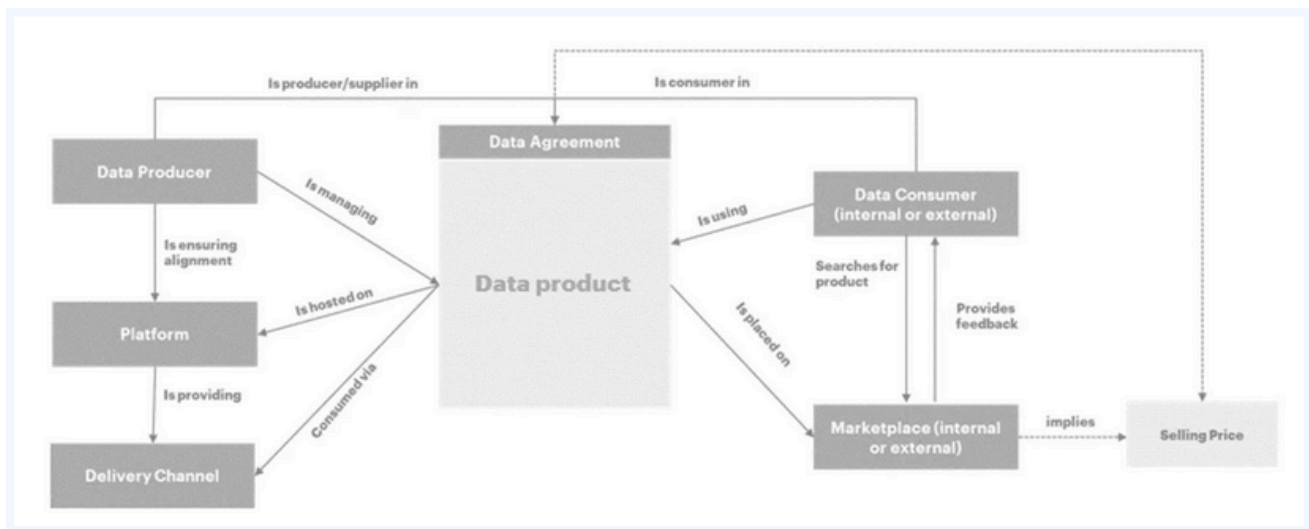
By leveraging Monte Carlo, Choozle has reduced their data downtime approximately **88%**.

# Self-Service Data Products: Data Product Marketplaces

Treating your data as a product is one thing. Now, you also need to make it easy for the data consumers within your organization to access those data products.

That's where a data product marketplace can come in.

A data product marketplace is kind of like an online shopping marketplace, like Amazon, that serves as a one-stop self-service shop for data consumers. It's a centralized, standardized experience that enables data product owners to share data products within the organization.



Data product marketplace architecture. [Source](#).

# Self-Service Data Products: Data Product Marketplaces

Typically, a data product marketplace should make it easier for data consumers to do three key actions.

## Data Product Marketplace Checklist



**Find or search for data products.** Users should be able to explore and discover the details of various data products, read data contracts, preview samples, submit requests for future data products, and leave feedback on current data products.



**Publish data products.** Data product owners should be able to make their data products clearly visible to data consumers, with descriptions, definitions, performance KPIs, proper usage instructions, and more.



**Manage data products.** Data product owners should also be able to establish the correct standards, governance, and policies. They should also be able to monitor the performance of their data product, get notified of data quality issues or incidents, and maintain a changelog of product updates.

Data product marketplaces can be sophisticated, with personalized recommendations for relevant data products for the user. Data consumers should be able to sort data products by popularity, relevance, or rating and receive notifications about changes or issues with the datasets they use.

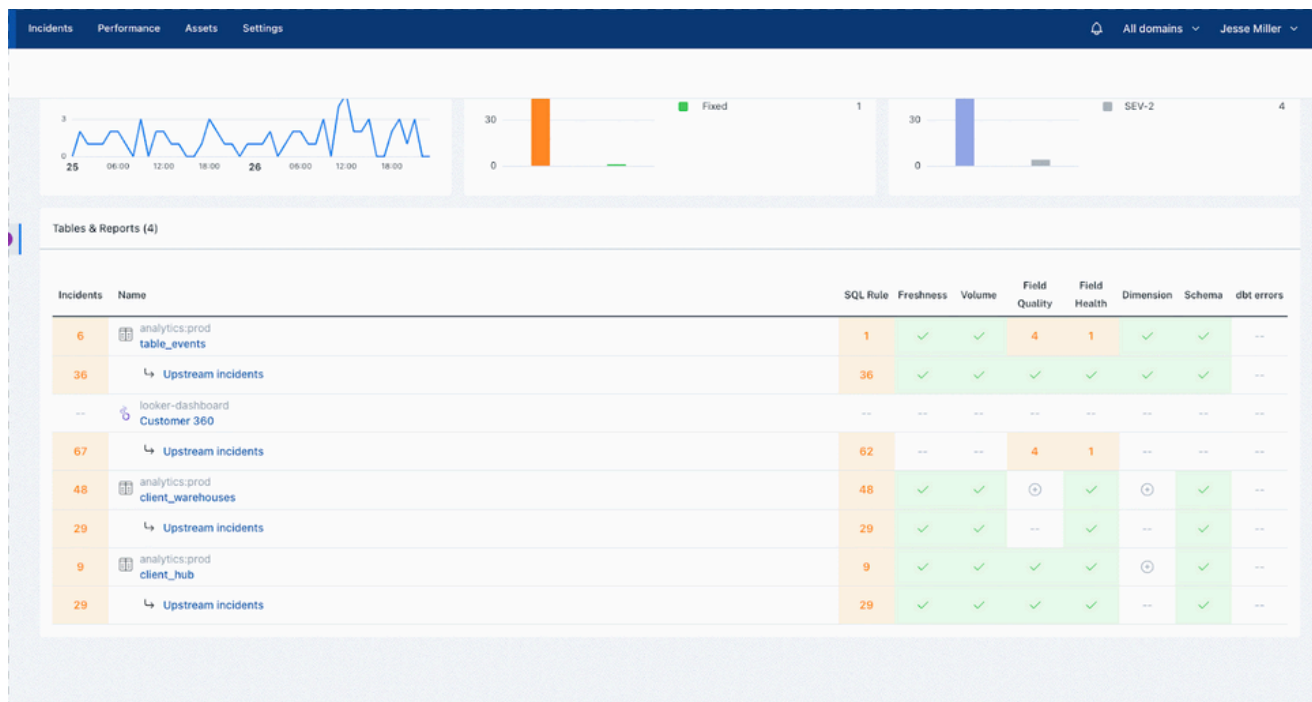
Data product marketplaces are a fast-growing option for data-driven organizations looking to drive efficiencies and ensure business users have access to the data products they need.

# Data + AI Observability for Data Products

Data products provide a fantastic framework to enable data teams to reduce tedium, automate the small stuff, and focus on innovative projects that move the business forward.

But your data products are only as valuable as they are reliable. A data asset ceases to be a product when business users lose the ability to trust its output. That's why working with a data + AI observability tool is so important.

By taking an end-to-end approach to your data quality management, you can understand at a glance how a product is performing, where it needs coverage, and how to fix it if something goes wrong.



# Data + AI Observability for Data Products

In fact, tools like Monte Carlo's include features like the [Data Product Dashboard](#) that help organizations manage and improve the quality of data and tables powering their most critical product—and foster cross-functional collaboration in the process.

With our tailored dashboards and automated data profiling, you'll be able to identify coverage gaps instantly, scale coverage automatically, and understand the health of all your critical products at a glance. Plus our AI enabled monitor creation gives your stakeholders the power to create their own monitors too—if you're into that sort of thing.

And when your data environment (and products) grow, your data quality coverage will grow right along with it.

That goes a long way towards building trust with your stakeholders—and protecting your sanity.

To learn more about how data observability can facilitate the reliability of your data products, talk to our team.

# FAQ: Data Products At-a-Glance

## What is an example of a data product?

Examples of data products include an airline's flight tracking system that integrates GPS and historical flight data, a customer relationship management platform syncing marketing tools, or an AI algorithm forecasting stock returns based on diverse financial datasets.

## What are types of data products?

Types of data products include dashboards, reports, machine learning models, and unified schemas for specific business use cases. These vary based on industry needs and user demands.

## What does a data product do?

A data product turns raw data into actionable insights by combining datasets with [product management practices](#), business logic, and access controls. It delivers value by aligning data to specific business goals and improving trust, reliability, and user adoption.

## What is a data product in data mesh?

In a data mesh, a data product is a domain-specific, decentralized data asset designed for self-service use. It adheres to governance and quality standards, facilitating its use across teams without central control.

## Why do you need data products?

You need data products to increase trust in data, provide measurable value to stakeholders, reduce time spent on ad-hoc queries, and enhance focus on innovative tasks. They also improve team efficiency and ensure data quality by structuring assets with governance and reliability practices.