🧪 **[New eBook] Data Testing 101: Take the First Step in Your Data Quality Journey** →
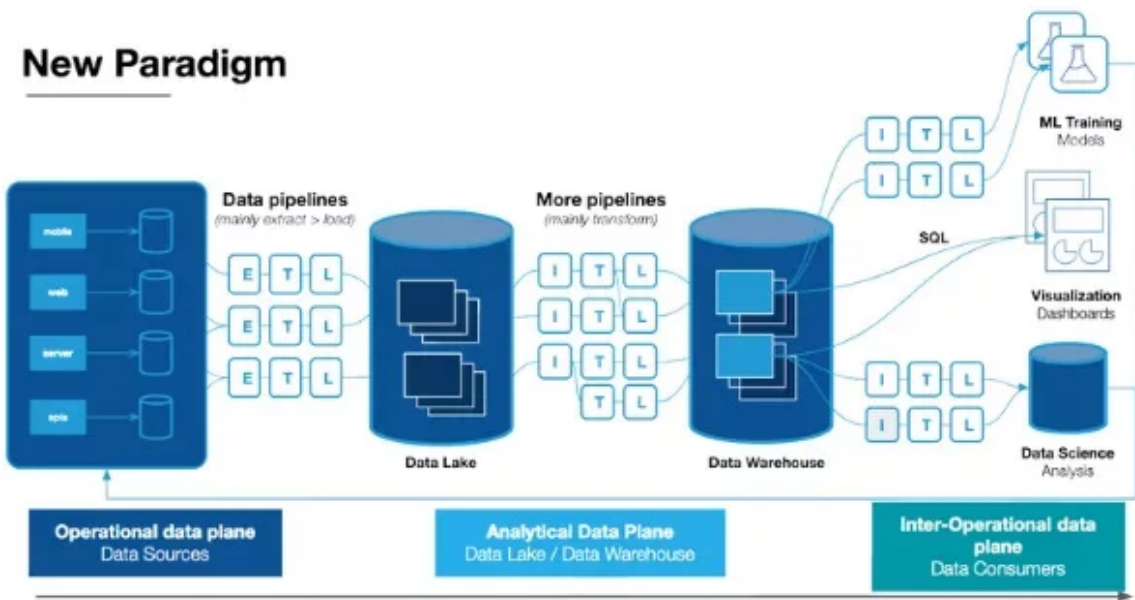
Download Today   ✕

Data Discovery
Data Reliability   **Q Search**
Case Studies
Announcements

Data Platforms, Data Reliability   Updated August 30, 2022

How to Treat Your Data As a Product

Share article   ↗   in   🐦

Your company wants to "treat data like a product." Great! What does that mean?

For the past few decades, most companies have kept data in an organizational silo.

Analytics teams served business units, and even as data became more crucial to decision-making and product roadmaps, the teams in charge of data pipelines were treated more like plumbers and less like partners.

Cookie Settings

entire ecosystem as a fully formed element of the company tech stack.

And the most forward-thinking teams are adopting a new paradigm: treating data like a product. This is a hot topic in the data community right now, and in recent months, we've had the privilege of discussing data-as-a-product with several industry leaders—and uncovering their real-world takeaways on what it looks like to bring this new approach to life on a daily basis.

How industry leaders define data as a product

To start, let's hear directly from a few data leaders on what it even means to treat data as a product. (The following quotes were edited for length and clarity.)

Chad Sanderson, Convoy

Here's what Chad Sanderson, Head of Product, Data Platform, at Seattle-based freight marketplace Convoy, had to say at a recent **event on building reliable data products:**

"In general, there are two schools of thought that are still developing around what it means to treat data as a product: in the first, you have an external or internal product or service that generates data, meaning that the data—including the entire pipeline—is part of the product. So it must be subject to the same level of rigor as the application code. It needs SLAs, it needs good testing, it needs good monitoring and documentation and so forth.

In the second approach, you think of the output of any codebase that's serving a customer as a product. For example, when you think about a data warehouse, it's really just a codebase—primarily composed of SQL—that's serving internal customers like other analysts, data scientists, and product managers who are using that data to go and make business decisions. Therefore, anything that's pushed to a "production data environment" that the company can access is a product. So if you're using a dashboarding tool like Mode or Metabase, and you're writing SQL and pushing that dashboard to a public environment where other people can access it, that is also a product. And so it must also be subject to the same level of rigor as any other product.

These are both pretty different approaches than the way we've thought about data over the past 10 or 20 years, where it's been more of a silo with data scientists and analysts that are going off and answering independent questions."

Atul Gupte, Uber

In a conversation with the Monte Carlo team, **Atul Gupte, former Product Manager for Uber's Data Platform Team**, goes a step further and describes the role of a data product manager.

- Who needs this data?

- Where is this data flowing to/from?

- What purpose does this data serve?

- Is there a way to make it easier to work with/access this data?

- Is this data compliant and/or actionable?

- How can we make data more useful to more people at the company, faster?

Data product managers answer these questions by building internal tooling and platforms for employees.

Some data product managers are beholden to data analysts and data scientists, others to operations teams, software teams, or—in the case of larger companies—executives. They usually come from backgrounds like traditional B2B product management, internal tooling product management, data analysis, or back-end engineering. But instead of working with what we traditionally define as 'customers' (i.e., an individual buying or consuming your software), you're working with "data consumers"—in other words, employees using products that make sense of your company's data."

Justin Gage, Retool

And as **Justin Gage**, data leader at Retool, shared on Medium, the concept of data as a product can help clarify the question of what data teams do, and what mandate they should focus on carrying out.

"Data as a Product (DaaP) is the simplest model to understand: the job of the data team is to provide the *data* that the company needs, for whatever purpose, be it making decisions, building personalized products, or detecting fraud. This might just sound like data *engineering*, but it's not: Data Scientists also provide data as a product, just packaged in a different way.

In the DaaP model, company data is viewed as a product, and the data team's role is to provide that *data* to the company in ways that facilitate good decision-making and application building. Some important characteristics of this model:

- Data has an SLA (figuratively) from the entire data team, not just data engineering

- Data flow is unidirectional, from the data team to the company

- Domain expertise has limited benefits for data team members"

making and other key functions, and applying standards of rigor like SLAs.

How teams can apply the data-as-a-product approach

From our conversations with these leaders and several others, we've identified five key ways modern data teams can apply this approach to their own organizations.

1) Gain stakeholder alignment early – and often

When data is your product, your internal customers are also your stakeholders. Make it a priority to partner with your key data consumers as you map out your own data product roadmap, develop SLAs, and begin treating data as a product.

This means putting on your product manager hat—or, as Atul suggested above, having a role dedicated to data product management—to fully understand the needs, concerns, and motivations of your internal customers. You'll want to have a clear grasp on who uses your data and how, and for what purposes. This will help you understand what types of data products you need to build to meet those needs.

This understanding also helps you adopt data storytelling. Software, product, and UX teams use the **practice of storytelling** to share the context of their work through different perspectives that will help stakeholders understand its value based on what matters most to them. And you'll be working to convince your stakeholders that data should be prioritized, and to justify the investments required to treat data as a product.

For example, Chad Sanderson describes storytelling as an invaluable tool when it comes to persuading stakeholders to invest in data infrastructure over flashier machine learning models or new features that promise to generate millions of dollars.

"It takes a lot of work to turn that narrative around and put data quality at an equal footing…in some cases, I've addressed this issue by focusing on the story of our company and the story of our data scientists and analysts. When they're building models, are those models actually trustworthy? How much do we trust the data that makes up the foundation of the machine learning that drives our business decisions? So I try to tell a really good story, a tight narrative with clear examples from the company that are actually tying into something that generates revenue. So at the end of the day we can say, look, improving our data quality is probably not going to have an observable, direct, measurable impact on revenue, but there is a clear qualitative understanding that it is going to benefit us financially in the long term based on all of these use cases."

2) Apply a product management mindset

Another key step is to apply a product management mindset to how you build, monitor, and measure data products.

[Jessica Cherny](#), Lead Data Analyst at Ironclad, described her company's agile-inspired workflow.

"We're treating data internally as a product, and that means applying product management principles to data and the data team. So when we have a big strategic project that requires data, we create data scoping documents, just like a product manager would create a spec, with the right stakeholders. And we keep iterating with engineers and the product managers to make sure it's a cross-functional, stakeholder-aligned output—as opposed to just having data people working in a silo and not interacting with anyone."

And similar to engineering processes, data teams should be factoring in scalability and future use cases when building pipelines. According to Chad, this can represent a significant shift from how data teams have approached their work in the past. "Oftentimes, the data that actually lands in a production database is really just service-level events that get thrown in by engineers without really thinking about it. So one of the big reasons why data models get so messy as a company evolves is that we're usually focused on rapidly building services first and thinking critically about data second. And this idea of data as a product is kind of a continuum shift to start to change that."

[Kyle Shannon](#), Senior Data & Analytics Engineer at SeatGeek, shared in the same webinar that his company is focusing on scalability due to the rapid growth of their data team. "We're really trying to understand how we can better onboard new people coming in and making better processes to make data more discoverable and accessible. People that have been at a company for a long time know where to go to find information, but if you're hiring 20 or 30 data team members over the year, it's really hard to say, 'Oh, just go into the Slack channels and ask questions.' It's not going to scale. So as you are building your data products, you have to document everything and make sure it's very clear—that you're removing redundancies or any issues you might find along the way."

Another product mindset to adopt is setting up KPIs aligned with your business goals before you begin building any new data product. As Chad described earlier, storytelling can help illustrate the potential benefits of investments in data quality, but most organizations will still expect mature teams to measure the financial impact of their initiatives.

Many data teams are adopting KPIs related to data quality, such as **calculating the cost of data downtime**—times when data is partial, erroneous, missing, or otherwise inaccurate—or by measuring the amount of time data team members spend troubleshooting or fixing data quality issues, rather than focusing on innovations or building new data products.

Setting baseline metrics for your data will help quantify the impact of your data initiatives over time. Just ensure these metrics are applied consistently across use cases, particularly if you have a central **data platform**.

users need to have the ability to self-serve and meet their own data needs. Self-service tooling that empowers non-technical teams to access data allows your data team to focus on innovative projects that add value, rather than functioning as an on-demand service to fulfill ad hoc requests.

Self-serve tooling is also one of the main principles of the data mesh concept—a new approach to decentralized data architecture. Mammad Zadeh, the former VP of Engineering at Intuit for their Data Platform team, is an enthusiastic advocate of the data mesh and believes self-serve tooling is integral to both data architecture and data products.

As he suggests, "We, in the central data teams, should make sure the right self-serve infrastructure and tooling is available to both producers and consumers of data so that they can do their jobs easily. Equip them with the right tools, let them interact directly, and get out of the way."

4) Prioritize data quality & reliability

One key component of approaching data as a product is applying standards of rigor to the entire ecosystem, from ingestion to consumer-facing data deliverables. As we discussed in the context of storytelling earlier, this means prioritizing data quality and reliability throughout the data lifecycle.

Companies can assess their current state of data quality by mapping their progress against the data reliability maturity curve. Briefly, this model suggests there are four main stages of data reliability:

#1 Reactive Teams spend a majority of their time responding to fire drills and triaging data issues—resulting in a lack of progress on important initiatives, an organizational struggle to use data effectively in their product, machine learning algorithms, or business decision-making.

#2 Proactive Teams collaborate actively between engineering, data engineering, data analysts, and data scientists to develop manual checks and custom QA queries to validate their work. Examples might include validating row counts in critical stages of the pipelines or tracking time stamps to ensure data freshness. Slack messages or email alerts still pop up when things go wrong, but these teams do catch many issues through their proactive testing.

#3 Automated At this level, teams prioritize reliable, accurate data through scheduled validation queries that deliver broader coverage of pipelines. Teams use data health dashboards to view issues, troubleshoot, and provide status updates to others in the organization. Examples include tracking and storing metrics about dimensions and measures to observe trends and changes, or monitoring and enforcing schema at the ingestion stage.

#4 Scalable These teams draw on proven Dev Ops concepts to institute a staging environment, reusable components for validation, and/or hard and soft alerts for data errors. With substantial

and table to be monitored and tracked for quality.

Reaching data maturity doesn't happen overnight. But by starting to **set clear data SLAs, SLIs, and SLOs** that measure quality, you can begin to demonstrate the value of investing in automated, scalable data reliability. Examples of common data health metrics include the number of data incidents for a particular asset, time-to-detection, and time-to-resolution.

As discussed previously, it's crucial to align your data stakeholders when setting SLAs—you want to ensure you have a shared understanding of what "reliability" means to your data consumers, which assets matter most, and which potential issues require the most immediate attention.

5) Find the right team structure for your data organization

Of course, team structure makes a huge impact on how your organization interacts with data on a daily basis. Do you have a centralized data team that handles every aspect of data management and application? Or analysts embedded across business units, meeting specific needs and gaining domain expertise—but suffering from silos and lack of cohesive governance?

Different companies will require distinct approaches depending on their size and business needs, but many data leaders we've talked to have found the best outcomes with a hub and spoke model. In this structure, a centralized data platform team handles infrastructure and data quality, while decentralized, embedded analysts and engineers deal with semantic layers and apply data to the business. This model works well if your organization is growing fast and needs to move quickly, but can lead to duplication and repeated efforts on the embedded analysts' part without solid alignment with the centralized data team.

Greg Waldman, Senior Director of Business Intelligence at restaurant POS software Toast, led his team through a **five-year organizational evolution** that included switches from centralized to decentralized to hub and spoke models. He advises data leaders at growth companies to follow a key tenet of product management—stay agile. "The way I think about data teams, in a nutshell, is that you want everyone to add as much business value as possible. We've been very open to change and trying different things and understanding that what works with 200 people versus 500 people versus a thousand is not the same answer, and that's okay. It can be somewhat obvious when you've reached those inflection points and need to try something new."

For Jessica Cherny, the advantage of decentralized analysts and engineers is their ability to understand the real business need behind data requests. "I want to understand how to design a deliverable that actually serves their needs. It happened recently when I was asked by someone on a strategic initiative to get a specific set of data right away. And I was able to say, 'Wait, hold on. Do I really need to use this complex clustering method to answer this question? What is the actual need

understand the business need behind her question and how to answer that in a simple, easily
understood way."

Again, every company will have its own cultural landscape and challenges to address, but a hub and
spoke model can help growing teams move fast to meet business needs without giving up
ownership of data quality and governance.

What's next?

We just covered a lot. But that's because adopting an organizational approach of treating data like a
product isn't just a buzzworthy trend in the data industry. It's an intentional shift in mindset that
leads to meaningful outcomes: increasing data democratization and the ability to self-serve,
improving data quality so decisions can be made accurately and confidently, and scaling the overall
impact of data throughout the organization.

*Interested in learning how to treat your data platform like a product? Reach out to Mei and Book a
time to speak with us using the form below.*

Your work email*

Talk To Us!

Mei Tao

Product at Monte
Carlo

in

Sara Gates

Sara is a content
strategist and
writer at Monte
Carlo.

in

The Data Downtime Blog

Advice and best practices for busy data leaders.

MC **MONTE CARLO**

Monte Carlo

Home

Sign in

Contact us

Product

Data Observability

Detect

Resolve

Prevent

Integrations

Documentation

Company

Customers

About us

Resources

Security

We're hiring!

Legal

Terms and Conditions

Data Processing Addendum

Terms of Use

Privacy Policy

Cookie Policy

Report Issues

As businesses increasingly rely on data to power digital products and drive better decision making, it's mission-critical that this data is accurate and reliable. Monte Carlo, the data reliability company, is creator of the industry's first end-to-end Data Observability platform. Named an Enterprise Tech 30 company, a 2021 IDC Innovator, an Inc. Best Place Workplace for 2021, and a "New Relic for data" by Forbes, we've raised $236M from Accel, ICONIQ Growth, GGV Capital, Redpoint Ventures, and Salesforce Ventures. Monte Carlo works with such data-driven companies as Fox, Affirm, Vimeo, ThredUp, PagerDuty, and other leading enterprises to help them achieve trust in data.