

A Better Way to Put Your Data to Work

by Veeral Desai, Tim Fountaine, and Kayvaun Rowshankish

From the Magazine (July–August 2022)



Harry Campbell

Summary. Most companies struggle to capture the enormous potential of their data. Typically, they launch massive programs that try to meet the needs of every data end user or have individual application-development teams set up customized data pipelines that can't... [more](#)

Though every company recognizes the power of data, most struggle to unlock its full potential. The problem is that data investments must deliver near-term value and at the same time lay the groundwork for rapidly developing future uses, while data technologies evolve in unpredictable ways, new types of data emerge, and the volume of data keeps rising.

Listen to this article

Speed + | -



00:00

To hear more, [download t](#)

The experiences of two global companies illustrate how ineffective today's predominant data strategies are at managing those challenges. The first, a large Asia-Pacific bank, took the “big bang” approach, assuming it could accommodate the needs of every analytics development team and data end user in one fell swoop. It launched a massive program to build pipelines to extract all the data in its systems, clean it, and aggregate it in a data lake in the cloud, without taking much time up front to align its efforts with business use cases. After spending nearly three years to create a new platform, the bank found that only some users, such as those seeking raw historical data for ad hoc analysis, could easily use it. In addition, the critical architectural needs of many potential applications, such as real-time data feeds for personalized customer offerings, had been overlooked. As a result the program didn't generate much value for the firm.

The second company, a large North American bank, had individual teams tap into existing data sources and systems on their own and then piece together any additional technologies their business use cases required. The teams did create some value by solving challenges like improving customer segmentation for digital channels and enabling efficient risk reporting. But the overall result was a messy snarl of customized data pipelines that couldn't easily be repurposed. Every team had to start from scratch, which made digital transformation efforts painfully costly and slow.

So if neither a monolithic nor a grassroots data strategy works, what's the right approach?

We find that companies are most successful when they treat data like a product. When a firm develops a commercial product, it typically tries to create an offering that can address the needs of as many kinds of users as possible to maximize sales. Often that means developing a base product that can be customized for different users. Automakers do this by allowing customers to add a variety of special options—leather upholstery, tinted windows, anti-theft devices, and so on—to standard models. Likewise, digital apps often let users customize their dashboards, including personalizing the layout, color schemes, and content displayed, or offer different plans and pricing structures for different user needs.

Companies that treat data like a product can reduce the time it takes to implement it in new use cases by as much as 90%.

Over time companies enhance their products, adding new features (engine modifications that boost fuel economy in a car or new functionality in an app), and introduce brand-new offerings in response to user feedback, performance evaluations, and changes in the market. All the while firms seek to increase production efficiency. Wherever possible, they reuse existing processes, machinery, and components. (Automakers use a common chassis on vastly different cars, for instance, and app developers reuse blocks of code.) Treating data in much the same way helps companies balance delivering value with it today and paving the way for quickly getting more value out of it tomorrow.

In our work we've seen that companies that treat data like a product can reduce the time it takes to implement it in new use cases by as much as 90%, decrease their total ownership (technology, development, and maintenance) costs by up to 30%, and reduce their risk and data governance burden. In the pages that follow we'll describe what constitutes a data product and outline the best practices for building one.

What Is a Data Product?

A data product delivers a high-quality, ready-to-use set of data that people across an organization can easily access and apply to different business challenges. It might, for example, provide 360-degree views of customers, including all the details that a company's business units and systems collect about them: online and in-store purchasing behavior, demographic information, payment methods, their interactions with customer service, and more. Or it might provide 360-degree views of employees or a channel, like a bank's branches. Another product might enable "digital twins," using data to virtually replicate the operation of real-world assets or processes, such as critical pieces of machinery or an entire factory production line.

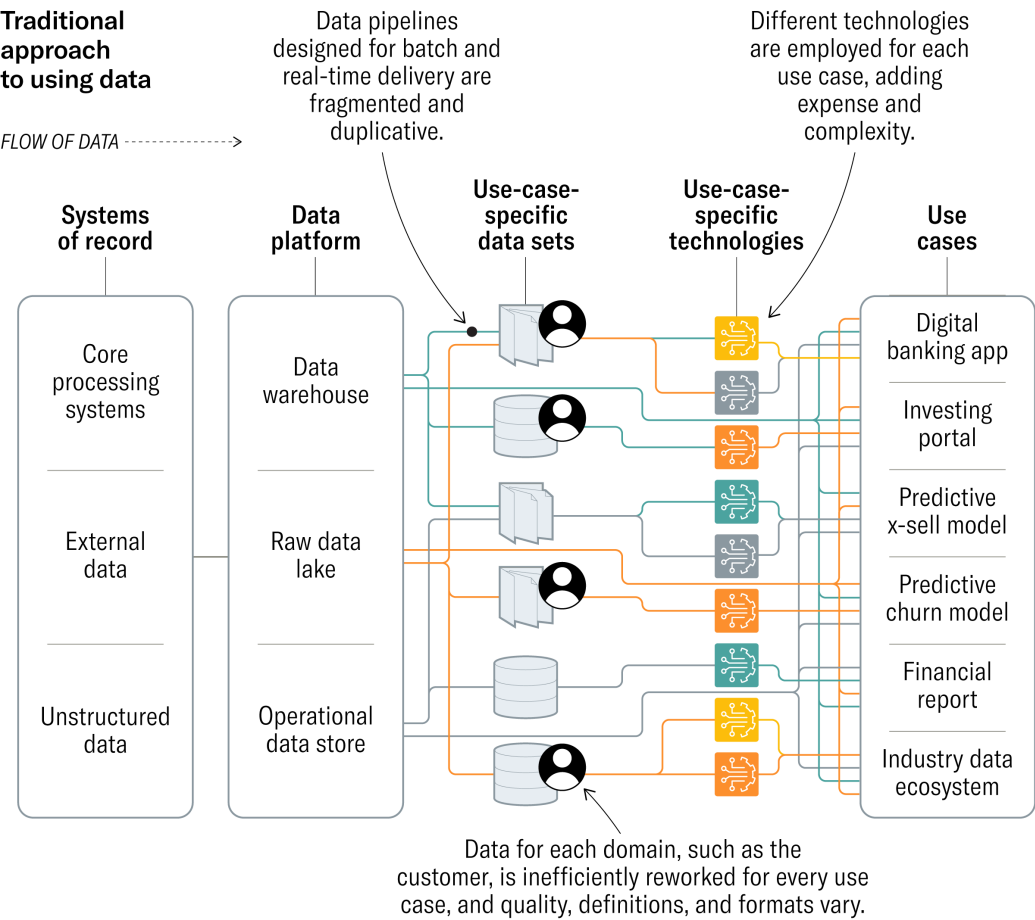
Because they have many applications, data products can generate impressive returns. At a large national bank, one customer data product has powered nearly 60 use cases—ranging from real-time scoring of credit risk to chatbots that answer customers' questions—across multiple channels. Those applications already provide

\$60 million in incremental revenue and eliminate \$40 million in losses annually. And as the product is applied to new use cases, its impact will continue to grow.

Data products sit on top of existing operational data stores, such as warehouses or lakes. The teams using them don't have to waste time searching for data, processing it into the right format, and building bespoke data sets and data pipelines (which ultimately create an architectural mess and governance challenges).

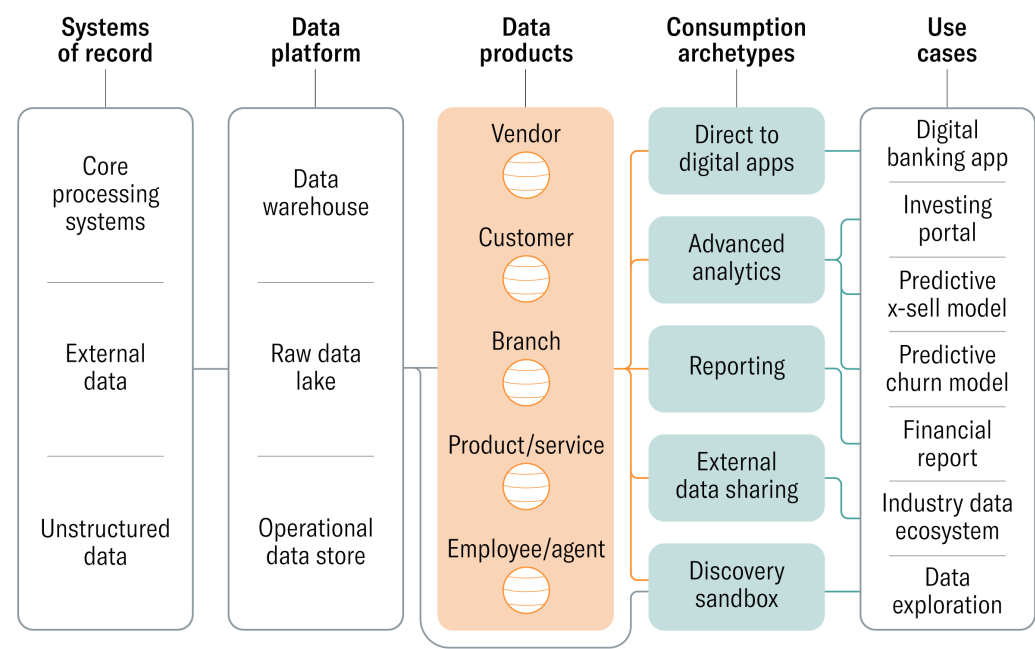
Traditional Data Consumption Versus the Data Product Model

In the traditional approach to data solutions, use case teams identify the data they need from source systems and create data sets and feeds only for their particular solutions. That leads to a lot of replicated work and a complex data architecture that’s difficult to maintain and use for new solutions.



In a data product approach, use case teams build solutions by leveraging standardized data products and wiring technologies together following consumption archetype patterns, which reduces work, simplifies the enterprise data architecture, and decreases the time it takes to realize value.

Data product approach



HBR

See more HBR charts in [Data & Visuals](#) >

Each data product supports data “consumers” with varying needs, in much the same way that a software product supports users working on computers with different operating systems. These consumers are systems, not people, and our work suggests that

organizations typically have five kinds. We call them “consumption archetypes” because they describe what the data is used for. They include:

1. Digital applications. These require specific data that is cleaned, stored in the necessary format—perhaps as individual messages in an event stream or a table of records in a data mart (a data storage area that is oriented to one topic, business function, or team)—and delivered at a particular frequency. For example, a digital app that tracks the location of a vehicle will need access in real time to event streams of GPS or sensor data. A marketing app designed to find trends in customer browsing behavior will need access to large volumes of web log data on demand (often referred to as “batch” data) from a data mart.

2. Advanced analytics systems. These too need data cleaned and delivered at a certain frequency, but it must be engineered to allow machine learning and AI systems, such as simulation and optimization engines, to process it.

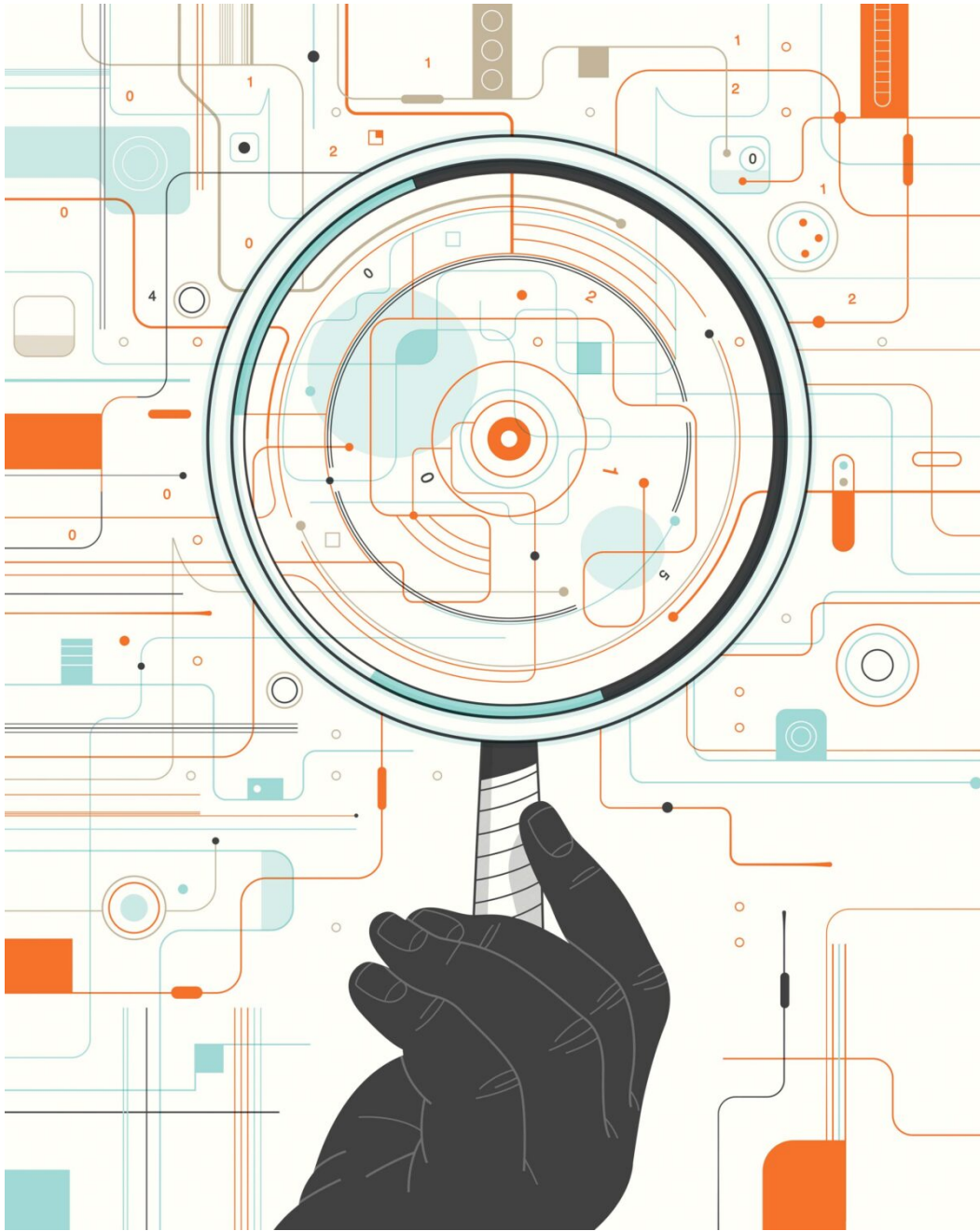
3. Reporting systems. These need highly governed data (data with clear definitions that is managed closely for quality, security, and changes) to be aggregated at a basic level and delivered in an audited form for use in dashboards or regulatory and compliance activities. Usually, the data must be delivered in batches, but companies are increasingly moving toward self-service models and intraday updates incorporating real-time feeds.

4. Discovery sandboxes. These enable ad hoc exploratory analysis of a combination of raw and aggregated data. Data scientists and data engineers frequently use these to delve into data and uncover new potential use cases.

5. External data-sharing systems. These must adhere to stringent policies and agreements about where the data sits and how it’s managed and secured. Banks use such systems to share fraud insights with one another, for example, and retailers to share data with suppliers in the hope of improving supply chains.

Each consumption archetype requires different technologies for storing, processing, and delivering data and calls for those technologies to be assembled in a specific pattern. This pattern is essentially an architectural blueprint for how the necessary technologies should fit together. For example, a pattern for a sandbox would most likely include technologies for setting up a multi-user self-service environment that can be accessed by data engineers across the company. The pattern for an advanced analytics system using real-time data feeds might include technologies for processing high volumes of unstructured data.

Like a Lego brick, a data product wired to support one or more of these consumption archetypes can be quickly snapped into any number of business applications.



Harry Campbell

Consider a mining company that created a data product providing live GPS data feeds of ore-transport-truck locations. It was designed to support all the archetypes except external data sharing for its first use case—improving ore-processing yields. The company soon discovered the product had uses far beyond that. Once it was made available more broadly in the organization, several entrepreneurial employees immediately leveraged it to eliminate bottlenecks in the mine transport system. In just three days they built a prototype of a truck-routing decision support tool that reduced queuing time and carbon emissions. If they'd had to engineer the data from scratch, it would have taken nearly three months.

As word continued to spread, employees interested in other issues that involved trucks—such as safety, maintenance, and driver scheduling—tapped into the data to find answers to thorny

questions and to build revenue-generating solutions that previously would have been impossible.

Managing and Developing Data Products

Whether they're selling sedans, software, or sneakers, most companies will have internal product managers who are dedicated to researching market needs, developing road maps of product capabilities, and designing and profitably marketing the products.

Likewise, every data product should have a designated product manager who is in charge of putting together a team of experts to build, support, and improve it over time. Both the manager and the experts should be within a data utility group that sits inside a business unit. Typically, such groups include data engineers, data architects, data modelers, data platform engineers, and site reliability engineers. Embedding them within business units gives the data product teams ready access to both the business subject-matter experts and the operational, process, legal, and risk assistance they need to develop useful and compliant data products. It also connects teams directly with feedback from users, which helps them keep improving their products and identify new uses. The first release of the customer data product at the national bank, for instance, focused on customer demographic profiles and information on transactions. Subsequent releases included data on customer interactions and on prospects, attracting significantly more data users and supporting teams developing other applications. The cost savings and incremental revenue realized by the product's early uses funded the next phases, creating a sustainable business model.

A company also needs a center of excellence to support the product teams and determine standards and best practices for building data products across the organization. For example, the center should define how teams will document data provenance, audit data use, and measure data quality, and should design the consumption archetype patterns for data product teams to use. This approach can eliminate complexity and waste. In addition, the center can be a resource for specialized talent or data experts when demand for them surges within utility groups or business-use-case teams. For example, at one telecom provider we worked with, computer vision experts, who are scarce but often in demand, sit within the central hub and are deployed to business units on request.

Like a Lego brick, a data product wired to support one or more consumption archetypes can be quickly snapped into any number of business applications.

While most companies already have some, if not all, of the talent needed to build out their utility groups and centers of excellence, many will need to deepen their bench of certain experts, particularly data engineers who can clean, transform, and aggregate data for analysis and exploration.

This was especially true for the mining company, which needed to grow its data engineering staff from three to 40 people. To fill that big gap, its leaders took a stepped approach. They hired contractors to get immediate work done and then embarked on far-reaching recruiting efforts: hosting networking events, publishing articles on LinkedIn, upgrading the skills of the software engineers already on staff, and developing internship programs with local colleges and universities. To improve retention, they created a guild for data engineers, which helped them build their skills and share best practices. The company also crafted individualized plans for data engineers that ensured those professionals had a clear growth path after joining the company.

Tracking Performance and Quality

To see whether commercial products are successes, organizations look at barometers like customer sales, retention, engagement, satisfaction, and profitability. Data products can be evaluated with commensurate metrics, such as number of active monthly users, the number of applications across the business, user satisfaction, and the return on investment for use cases.

The telecom company tracked the impact of its first data product—which provided comprehensive data on critical cellular-network equipment—in 150 use cases. They included investment decision systems, scenario-planning systems, and network optimization engines. In total they're set to produce hundreds of millions of dollars in cost savings and new revenue within three years. The company estimates that over the first 10 years the use cases will have a cumulative financial impact of \$5 billion—providing a return many times over on its initial investment.

And just as manufacturers routinely use quality assurance testing or production line inspections to make certain that their products work as promised, data product managers can ensure the quality of their offerings' data. To do so they must tightly manage data definitions (outlining, say, whether customer data includes only active customers or former and prospective customers as well), availability, and access controls. They must also work closely with employees who own the data source systems or are accountable for the data's integrity. (The latter are sometimes called "data stewards.")

Quality can suffer, for instance, when the same data is captured in different ways across different systems, resulting in duplicated entries. This was a risk with the national bank's customer data product. So its product manager worked with the stewards of the company's various customer data repositories and applications to institute a unique ID for each customer. That allowed the customer data to be seamlessly integrated into any use case or with any related data product. The product manager also partnered with the center of excellence to develop the standards and policies governing customer data across the enterprise and to monitor compliance—all of which facilitated reuse of the data product while building trust among users.

Where to Start

Leaders often ask which data products and consumption archetypes will get the highest and fastest return on investment. The answer is different for every organization.

To find the right approach for their companies, executives need to assess the feasibility and potential value of use cases in each business domain (this might be a core business process, a customer or employee journey, or a function) and group them first by the data products they require and then by the consumption archetypes involved. Categorizing the use cases like this helps leaders more efficiently sequence work and get a faster return on investment. For instance, they may end up pushing some lower-value use cases ahead if they leverage the data products and consumption archetypes of higher-value use cases.

For the executives at the national bank, this approach illuminated several priorities. First they saw that a customer data product that supported their most critical fraud-management and marketing use cases could generate tremendous value. Then they identified the kinds of data the product needed to gather first. Some of those use cases called for basic customer identifiers and reference data

(such as demographic or segmentation data) while others required comprehensive customer behavioral data. The bank also realized that the two consumption archetypes it should pursue first were a discovery sandbox and advanced analytics, which in combination would support most of the company’s priority fraud and marketing use cases.



Subscribe to our Weekly Newsletter

Managing Data Science

An eight-week newsletter on making analytics and AI work for your organization.

Sign Up

Data product decisions often involve trade-offs between impact, feasibility, and speed. Ideally, the initial target products and consumption archetypes will immediately apply to high-value use cases and a long pipeline of others, as the telecom provider’s product for its network equipment did.

However, feasibility considerations may cause a company to adjust its approach. For example, it may make sense to build momentum first in an area of the organization that has data expertise and has gotten some traction with data products, even if that isn’t where the biggest opportunity lies. We saw this happen at the mining company. It initially chose to develop two products that supported its ore-processing plant, where use cases had already been successfully proven, the managers were enthusiastic to pursue more, the team had a lot of prepared data to work with, and experts with deep expertise were available to help.

...

Most leaders today are making major efforts to turn data into a source of competitive advantage. But those initiatives can quickly fall flat if organizations don’t ensure that the hard work they do today is reusable tomorrow. Companies that manage their data like a product will find themselves with a significant market edge in the coming years, thanks to the increases in speed and flexibility and the new opportunities that approach can unlock.

A version of this article appeared in the July–August 2022 issue of *Harvard Business Review*.