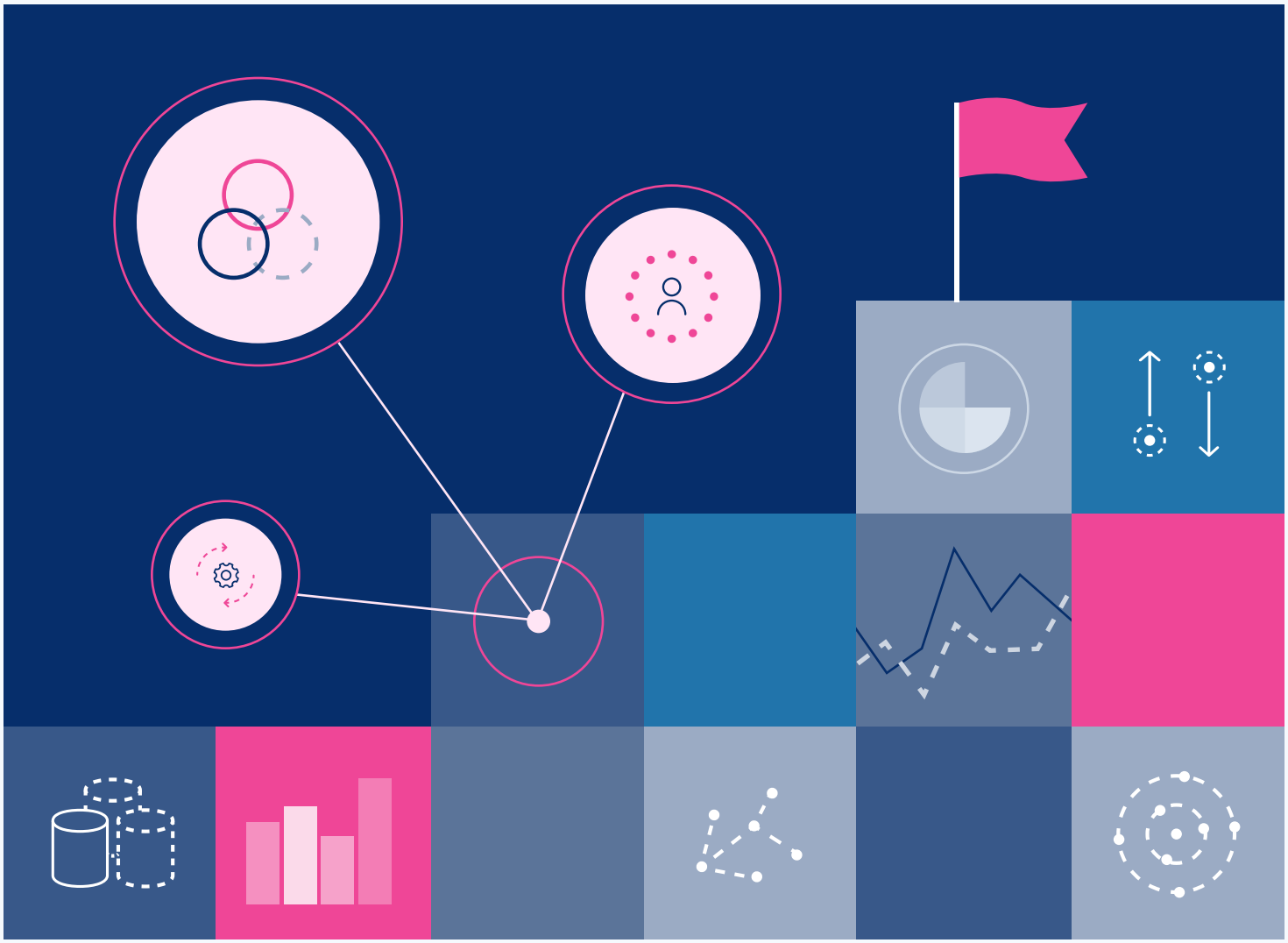
THINKDATA WORKS

GUIDE

Optimizing the Cost of Your Data Program

Learn the three pillars of cost-efficient data management, so you can reduce spend and reclaim lost time.



INTRODUCTION

There's no easy time to run a business, but the last few years have proven especially challenging. Supply chain turmoil, shifting consumer demand, a tight labor market, and rising energy prices have pushed organizations to adapt. Now, as inflation drives costs even higher, the need to spend smartly and prioritize ROI is very real.

Organizations that have weathered — and even thrived through — disruption have used data to their advantage to drive strategic improvements and bottom-line growth.

A recent survey of data executives shows the vast majority of organizations are continuing to increase their data investments, despite potential economic uncertainty. In 2022, 87.8% reported growth in their data investments and, looking ahead to 2023, 83.9% say they'll continue on this trajectory.[1]

> **87.8%** **of companies increased their data investments in 2022**

Organizations know that expanding and strengthening their data program can give them a critical edge. The challenge? Many struggle to build the right capabilities, tools, and organizational culture to deliver business value from their data investments.

If this challenge sounds familiar, then it's time to evaluate whether the cost of running your data program could be outweighing the return — and what to do about it.

In this guide, we break down the three pillars of a cost-effective data solution: **people, tools, and processes**. By building on this foundation, you can make strategic choices about your data investments that help you optimize spend, improve outcomes, and reclaim lost time.

Let's get started.

PILLAR 1

People

With any type of large-scale business transformation, the human element is one of the most complex and challenging to address. While it may sound simple, people make organizations function. For any new initiative to work, their success and satisfaction need to be at the core.

When it comes to growing your data investments, the impact on your people is significant — affecting day-to-day roles and responsibilities, training and development, organizational culture, and more.

For example, as your data program expands, the burden on your team can too, especially if they're spending a lot of time connecting to disparate data sources, managing those connections, or dealing with data quality issues. These points of friction don't just impact your team's efficiency — they can also hurt the employee experience.

Supporting cultural change while optimizing costs

For many organizations, the response to increased data management demands has been to hire more talent. But given today's rising labor costs, that solution isn't a sustainable one. 2022 saw unparalleled salary increases — as high as 15% — for data science and AI professionals, and those costs will continue to climb in coming years.[2]

DATA PROFESSIONALS' SALARIES ARE ON THE RISE[3]

$100k	$150k	$175k	$200k
Data Engineer	Data Scientist	Data Governance	Data Manager

Beyond filling talent gaps, what many organizations are still working on, in tandem with increasing their data investments, is building a data culture. The shift required to get there — to become a true data-driven organization — doesn't happen overnight. But there are some key areas you can address to optimize the "people" pillar while driving cost efficiencies in your data program.

1. Improve data discovery

One of the best ways to support a strong data culture is simply by making it easier for people to discover and use data. Given the competitive advantage you can unlock with the right data, it's not

surprising that so many organizations (46%) want a better way to source appropriate data for their use case.[4]

A data catalog gives meaning and context to data for everyone across the organization — technical and non-technical users alike. By enhancing search and discovery, a central platform puts the data people need at their fingertips, improving data trust and confidence and encouraging a culture shift toward being more data driven.

Most importantly, a catalog also improves the cost-efficiency of your strategy. A recent Forrester study found that organizations can drive, on average, $2.7 million in savings by enabling and accelerating data discovery.[5] In large part, that's because easier data discovery equates to fewer duplicative purchases. When it's easy to find and access the right data, people make the most of the data investments you have and the organization avoids unnecessary spending.

2. Empower everyone to do their best work

Knowledge workers waste up to 50% of their time dealing with mundane data-quality issues.[6] Whether they're looking for, accessing, purchasing, or cleansing data, these time-consuming tasks reduce the business value of your data investments. And they hurt the employee experience, too.

When people have to navigate a patchwork data solution with multiple, overlapping tools and processes, it adds unnecessary frustration to their day that can fuel dissatisfaction and even burnout — something 97% of data professionals report experiencing in their jobs.[7] Given the troubling numbers around chronic workplace stress, it's not surprising that data scientists stay in their roles for an average of only 1.7 years. Among the reasons for leaving, faulty data pipelines, finding and correcting data issues, unrealistic corporate expectations, and a culture of being "shamed and blamed" are key factors.[8]

One solution to the costly challenge of high turnover is helping more people and departments get more out of your growing data program. Unlike a piecemeal approach, an easy-to-use catalog platform shares data across the organization, enabling all departments to make the most of data. A centralized solution also keeps data specialists happy by allowing them to experiment more — focusing on high-impact efforts rather than menial tasks, grunt work, and duplicative effort.

3. Simplify learning and development

As you expand your data program, investing in learning and development (L&D) becomes a necessity. Good training takes time, money, and resources — and ideally you see high return on your L&D investments when your people have the knowledge and skills needed to deliver business value with data. Of course, when your data investments are distributed across various sources, access methods, and management tools, L&D costs start to add up.

One effective way to find cost efficiencies in your training program is to identify places where you can streamline the user experience. How would training requirements change if any data, no matter its source, was accessible through a standard central hub? If that were the case, a new project wouldn't mean a new system every time. Forrester's Total Economic Impact report reveals $286,085 in savings when organizations use a data catalog, shortening the onboarding of new analysts by at least 50%.[9] By using a central platform, you can focus training resources on developing new skills, not learning to navigate a tangled web of unique tools. This makes things easier for both new employees and those moving into new or expanded roles.

PILLAR 2

Tools

Strategically using data is key to going from reactive to predictive, which explains why many organizations have ramped up their data investments in recent years. The challenge with taking in more and more data from various sources and vendors is that many companies end up bolting on new pipelines and management tools ad hoc.

To be fair, this patchwork approach has been tough to avoid given the need to continually respond to new business challenges. But it's worth considering: When was the last time you stepped back, took a hard look at your tools, and asked, "what if there's a better way?"

Assessing your current toolset

The truth is, more data isn't always better if you don't have efficient tools to manage your investments. Poor data quality is responsible for an average of $15M per year in losses.[10] And, as we explored in the previous section, it wastes your team's time, too.

It's possible you're spending much more than you think trying to stay on top of separate data management tools. Here are some of the most common ways organizations can improve cost efficiency in their data tech stack:

- **Finding and reducing redundancies in spend**, which are often caused by overlapping data products, efforts, and tech purchases
- **Breaking down barriers to visibility** that make it tough to see the data and tools you have access to
- **Closing gaps in data management** to strengthen data quality
- **Simplifying data access and sharing** across your team while maintaining security
- **Identifying opportunities to consolidate** by loading several tasks into one platform rather than distributing them across individual solutions

Make your data management tools work for you (not the other way around)

By equipping your team with tools that automate and streamline routine data management activities, you can slash tedious, manual effort and focus more time and energy on high-impact tasks. In place of multiple, individual solutions, a centralized data catalog optimizes the "tools" pillar of your data program in the following ways:

1. Better connect to high-quality data

Almost half of organizations purchase data from five or more sources and 22% spend over $500K on external data. While data from outside sources can be a valuable asset to your business, they can also add greater complexity. As you bring in more external data, a centralized data platform becomes even more valuable to running a cost-efficient program and driving high ROI.

With a catalog built to handle a variety of sources, you can manage your assets more efficiently and use them more impactfully. Using a single platform, you can represent data any way you need — as metadata, as virtualized data, or fully ingested through an ETL into the warehouse infrastructure of your choosing.

2. Streamline costs for warehousing (hot and cold storage)

Data storage is important to consider when optimizing the cost of your data strategy. Fees for data storage, operations, and transfer can add up quickly and it's easy to think they're just a cost of doing business; but a data catalog platform can help reduce those costs dramatically.

More than a directory for your data lakes, data warehouses, and other repositories, a platform approach to data cataloging can bring that to the next level. A central hub lets you manage connections to unique warehouses to manage hot, cold, and archival storage via the same control plane.

By partitioning your storage in this way, you can serve up critical, high-utility data quickly on high performance infrastructure, while keeping lower-priority data in cheaper warehouses. No matter what, all data is readily available and treated as first-class, even if it's below the highest tier of importance.

3. Save time on data management and governance

Today's data professionals can manage an average of eight data feeds before it becomes a full-time job.[11] And more than a third (39%) of data scientists' time is spent on menial tasks like preparing and cleansing data.[12] That's not exactly satisfying work — and it's not the best use of your team's skills.

As you plan to increase your data investments, it's a good time to strategize how best to support your data team and make efficient use of their time. After all, the larger your data program grows, the larger their workload gets — and the more important it is to streamline day-to-day tasks.

Rather than managing data across multiple systems, a single catalog platform reduces the number of moving parts your data team needs to stay on top of. That translates to fewer points of potential failure and fewer instances of redundant or duplicative work. With a data catalog, you can also auto-discover, classify, and tag sensitive data assets at scale, enable granular access controls, and keep a reliable audit trail to comply with privacy regulations. A central solution for data governance can boost data scientists' and engineers' productivity by 30% and help you avoid costly GDPR fines — a challenge that has amounted to €2.76B in losses for companies as of February 2023.[13]

PILLAR 3

Processes

Is bringing in external data part of your strategy? Compared to companies that don't use external data, those that use more than seven data sources have nearly:

14x	2x	37%
fixed-asset turnover	**market capitalization**	**more revenue per employee**[14]

That's great news for them. But it may leave you wondering, "How do I get there? What data are these companies tapping into that I'm not?" In reality, the specific data sources aren't the differentiating factor — the process is. The reason these companies are so successful with their strategy is that they can bring in data from more sources more easily, and that means less downtime, more insights, and better decision-making.

Streamline how you access and manage data

When working with separate data sources and tools, your internal processes can become muddled and time-consuming. And, especially for large organizations that are expanding their data investments, legacy operating practices and processes can pose a challenge to becoming more data-driven.

Rooting out process inefficiencies and streamlining your approach can help you realize the full value of your data investments. With the right data tools to support cost- and time-efficient processes, you can accelerate your ability to make strategic decisions — and that can make your organization more agile, competitive, and profitable.

32% **of organizations say they're deterred by a lack of process for using external data**

~40% **of organizations have a formal strategy in place to find and integrate external data**

A central repository helps you drastically improve the "process" pillar of your data management strategy, making your internal workflows easier, less resource-intensive, and more intuitive.

1. Standardize finding and using data

Are your business users continually tapping the data engineers to get access to the data they need? Or hitting up data scientists to help them interpret it? Does every project feel like starting from scratch every time?

As data volumes increase, manual, ad-hoc processes simply aren't sustainable. Your people need direct access to high-quality data for lots of important use cases, from marketing campaigns to demand forecasting. Expediting the process by which they find and access data is key to ensuring it delivers the highest possible value for your business.

In place of ad-hoc tasks for every new dataset or employee, a data catalog provides a single control plane and access method. With a consistent, streamlined approach to connecting people and data, you save time and resources for both your IT team and data consumers. According to Forrester, a data catalog can unlock productivity improvements for business users that amount to an average of $584,182 in annual savings.

2. Maintain a single source of truth

For organizations planning to increase their use of data, relying on analysts to manage data quality and consistency isn't feasible. In some cases, multiple data consumers may even be copying one dataset and addressing data quality issues themselves, resulting in repetitive, duplicative effort that wastes resources and increases costs. On top of costing the organization time, inaccurate, outdated, and otherwise poor-quality data can cost you opportunities and potential revenue.

When all employees access and use data from a common source, it leads to consistent analysis, more

accurate results, and greater process efficiency. A data catalog helps you maintain a single source of truth, ensuring your business decisions are always based on the best possible information.

3. Understand and realize the true value of data

The ultimate function of your data program is to deliver business value — that is, to help you make better decisions that are based on data, rather than memory or intuition alone.

Whether it's understanding customers, improving their experience, building better products, or branching into new markets, companies are using data all the time to be more agile and competitive. And despite the investment in more and more data every year, 78% of companies cannot quantify the business value of that data.[15] That means they could be spending money on data that isn't actually helping them drive revenue or operate their business better.

350% **or higher ROI for Organizations using a data catalog[16]**

That brings us to the final way in which a data catalog can help your organization run a more cost-efficient data program. By allowing you to monitor activity associated with data's health, consumption, and use, a catalog gives you a picture of data across its lifecycle within your organization — and that translates to a clear view of data's value. You can easily compare your spending against activity to understand which assets are critical and which are costing you more than they're worth.

10% **increase in data accessibility** **=** **$65M+** **additional net income for a typical Fortune 1000 company[17]**

Putting it all together

There's plenty of evidence to prove that a data catalog is key to getting the most business value out of your data investments. But is it best to build or buy your solution? Let's break it down.

	Build	Buy
Impact on People	• Mostly built for technical users • Labor-intensive option that requires approx. 5-7 data engineers to support ongoing maintenance (and even more for initial development and implementation) • Training needs to be provided by an in-house team	• Supports access and collaboration for tech and business users • Requires no internal resources for fully managed services • Enablement and training for your team • Regular collection of feedback to continuously improve UX
Impact on Tools	• Susceptible to falling behind on latest developments and standards • Could negatively impact functionality and compliance, and prevent you from seeing maximum ROI in data	• Product undergoes continuous improvement to keep up with industry best practices and standards • Provider is driven to stay competitive and drive innovative solutions
Impact on Processes	• Launch in 8-12 months • Requires in-house support and troubleshooting • Doesn't necessarily empower non-technical users to easily access and use data	• Ready to use out of the box • Comes with dedicated support and solutions team to ensure streamlined processes and optimal use
Impact on Costs	• Save on software licenses • Spend 2-3 times more in ongoing tech resources (maintenance, updates and improvements, edge cases) • Many salaries now divert to planning, building, fixing, and improving your catalog	• Invest more initially in an off-the-shelf product • Spend less money over time and never worry about ongoing costs for maintenance, updates, improvements, etc. • No resources pulled from the core function of internal teams
Estimated Adoption Costs	**$1M**	Starting at **$30k**

Unlock the value of data to grow your business

It takes time to step back and critically examine where your data program is working — and not working — for your business. But it's time well spent. The process allows you to understand redundancies, gaps, and opportunities when it comes to supporting your people, tools, and processes.

Based on your analysis, a centralized data catalog could be key to unlocking the value of data to grow your business. To explore the potential impact for your organization, try our **Data Catalog ROI Calculator** or **book a consultation with one of our data experts.**

About ThinkData Works

ThinkData Works unlocks the value of your data to grow your business. Connect to any source, catalog your assets, and deliver data to the people and applications that need it most while retaining visibility and control. Our unified cloud platform cuts overhead, fuels innovation, and drives revenue growth.

Sources

1. NewVantage Partners, Data and Analytics Leadership Annual Executive Survey 2023
2. Burtch Works
3. Burtch Works
4. Anaconda, State of Data Science 2021
5. Forrester, Total Economic Impact of a Data Catalog
6. Capgemini, How Data Quality Can Hurt Your Data Science Programme... If You're Not Careful
7. Wakefield Research
8. 365 Data Science Ltd.
9. Forrester, Total Economic Impact Report
10. Gartner, How to Create a Business Case for Data Quality Improvement
11. Data professional at a major financial institution (ThinkData Works client)
12. Anaconda, State of Data Science 2021
13. McKinsey, Designing Data Governance that Delivers Value
14. Capgemini, Data Sharing Masters
15. Capgemini, Data Sharing Masters
16. Forrester, Total Economic Impact Report
17. Forrester, Better Data Quality Equals Higher Marketing ROI