

## GUIDE

# Calculating the **ROI** of a Data Catalog

5 steps to getting more out of your data program



Use this guide in combination with our [Data Catalog ROI Calculator](#)

## INTRODUCTION

It's no secret that data is the great differentiator for today's organizations—and a critical factor in bottom-line growth. The challenge? **Companies are investing more and more in data without the correct tools** to manage, analyze, and use it to drive better business.

If you want to understand the true cost of your data investments and how to get more out of them, this guide walks you through the process step by step. Keep reading to see how a centralized data catalog platform helps you optimize spend, improve project outcomes, and reclaim lost time.

Use this guide in combination with our [Data Catalog ROI Calculator](#)



# Set your goals and priorities

In the race to get ahead of the competition, it's not uncommon for organizations to put their foot on the gas and accelerate their data investments without planning a route or destination. Now that you're taking stock of your data program, it's a good time to establish some clarity around what you want to achieve.

Like any strategic project, the best place to start is by articulating the challenges you want to solve, then choosing a solution that tackles them head on. Getting crystal clear on your goals makes it possible to measure the real business impact of your data investments.

## 6 common needs of data-enabled teams



### Knowing what data assets are available

only **18%** of companies have **adopted a data catalog** to a high degree<sup>1</sup>



### Improving team productivity

Without a governance solution, data teams are **30% less productive**<sup>2</sup>



### Ensuring data quality & monitoring

The average organization **loses \$15M** annually due to low quality data<sup>3</sup>



### Knowing how data is being used

**78%** of companies cannot quantify the **value of data**<sup>4</sup>



### Getting more out of data

**91.7%** of companies are increasing their **investments in data**<sup>5</sup>



### Meeting compliance regulations

**€1.67B in GDPR fines** passed down as of August 2022<sup>6</sup>

## QUESTIONS TO ASK YOUR TEAM

- What's the business impact of your data program today?
- What's getting in the way of using data to inform more of your decisions?
- What does success look like for your data team? For your business leaders?

## Analyze talent costs

As your data investments increase, the burden on your team can too, especially if they're spending the bulk of their time connecting to disparate data sources and managing those connections.

For many organizations, the response has been to hire more talent. But given today's rising labor costs, that remedy comes with unsustainable overhead that cuts sharply into your bottom line.

### The squeeze of a tight labor market

Companies are vying to attract and retain data professionals in the midst of a talent shortage, inflation, and a looming recession. 2022 saw unparalleled salary increases at all job levels for data science and AI, and those costs will continue to climb in coming years.<sup>7</sup>

#### TYPICAL SALARIES FOR A DATA TEAM

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

**15% increase in data scientists' 2022 median base salary over 2021<sup>8</sup>**

### Increasing efficiency, not headcount

When you consider the size and cost of your data team—and the imperative to increase both as you expand your data program—the need to find efficiencies and enhance productivity is clear.

A centralized platform gives you the opportunity to do both, letting you streamline and automate more tasks, so your team can focus on high-impact efforts rather than menial tasks, grunt work, and duplicative effort.

#### QUESTIONS TO ASK YOUR TEAM

- Calculate the size and cost of your data team, from data scientists, engineers, managers, governance, to analysts. How much will it cost to keep things running the way they are?
- If you plan to expand your data operation, are you also prepared to hire more talent?
- What has to change to keep labor costs manageable while you scale up data investments?

## Audit existing solutions

Few organizations have the luxury of taking a measured approach to data. With disruptive global events accelerating the pace of change, companies have to be ready to pivot or perish.

Data is key to being proactive instead of reactive. But to get ahead, it's likely you've had to take in more and more data from various sources and vendors, bolting on new data management tools ad hoc. This can get in the way of the important practice of stepping back and evaluating your overall approach.

### Streamlining spend

By taking a critical look at purchased data management solutions, you can identify the things you're already doing that may be costing more than they should:

- **Redundancies in spend** — Where do data products, efforts, and tools overlap?
- **Barriers to visibility** — What are you not seeing when it comes to the data and tools you have access to?
- **Gaps in data management** — Where is data quality suffering because you don't have the right tools?
- **Opportunities to streamline and connect** — How would better data connections and management across fewer tools reduce time, effort, and costs?

350%

or higher ROI in organizations using a data catalog<sup>9</sup>

10%

increase in data accessibility results in

\$65M+

additional net income for a typical Fortune 1000 company<sup>10</sup>

## Annual data management costs

Function	Without a solution	Individual solutions	Data Catalog Platform
Storage	N/A	\$270 per TB	\$270 per TB
Governance	20% of total user salaries	\$250K plus \$5K per user	Included
Virtualization	N/A	\$68.5K	Included
Transformation	20% of data practitioner salaries	\$1K per data practitioner	Included
Quality & Monitoring	10% of data practitioner salaries	\$2.5K per data practitioner	Included
<b>Estimated total</b> (30-person team with salaries and tech)	<b>\$6M</b>	<b>\$5.2M</b>	<b>\$2.8M</b>

### QUESTIONS TO ASK YOUR TEAM

- What are you already doing that's costing more than it should?
- Where are you duplicating spend or efforts?
- Where are there gaps in your organization's workflows that could be streamlined or automated?

## Scope out external data needs

By increasing access to, and usage of, the right external data, companies can supercharge their insights and enhance their competitive advantage. Compared to organizations who don't use external data for decision-making, those that use more than seven data sources have approximately:

14x

the fixed-asset turnover

2x

the market capitalization

37%

more revenue per employee<sup>9</sup>

### Challenges of external data

Despite the need for insights driven by external data, a recent survey showed that only 40% of organizations have a formal strategy in place to find and integrate it.

That's a problem, since external data both costs more to acquire and introduces greater process complexity. Organizations aren't just spending big money on data from multiple providers, they're struggling to manage it effectively.

**External data is a pillar of modern business — but a costly one**

5+

data sources purchased by  
almost 50% of organizations

22%

of companies spend over  
\$500K/year on external data



**This doesn't include overhead costs for procurement, legal, security, and compliance**

## Complexities of external data

The top three deterrents to using external data are:

63%

difficult data preparation and integration

34%

lack of strategies to address usage and risk

32%

lack of process for data usage

## Managing increased demand

If you're planning to use more external data, it's important to think about how you'll handle the added demands that come along with it, so you can make the most of your investments.

The more external data you're using, the greater the benefits you'll see from a centralized data platform. If you're using a majority of external data, you can manage it more efficiently and use it more impactfully by accessing it through a catalog built to handle a variety of sources.

### QUESTIONS TO ASK YOUR TEAM

- What projects are held back by access constraints? What could you be doing differently or better if you could tap into new sources of data?
- How can you increase consumption of external data efficiently and cost effectively?
- Could you double the data you're using without doubling the size of your team?
- Do you have an organization-wide picture of data access? What are the barriers to achieving that?

## Identify where your time goes

You could have access to all the data in the world, but without the right tools to manage your investments efficiently and effectively, you won't see the results you need. In fact, it could be costing you more than you think. Organizations say poor data quality is responsible for an average of **\$15M** per year in losses.<sup>3</sup> And knowledge workers waste up to **50%** of their time dealing with mundane data-quality issues.<sup>11</sup>

As your data investments increase, is your data team struggling to keep up with data quality and monitoring tasks?

8



The number of data feeds a data professional can manage before it becomes their full-time job<sup>12</sup>

### Stages of data management

**SOURCING****PREP****ORGANIZING****ANALYSIS & USE**

When it comes to these four areas, balance across disciplines is key. For example, it might seem like a good sign that your team is spending 85% of their time in analysis. However, that could also be a symptom of poor data prep and cleansing. These problems can critically delay your ability to put data to use, and call into question the quality and reliability of your data and pipeline.

# 46%

of organizations want a better way to source the right data for their use case

# 39%

of data scientists' time is spent on prep and cleansing data<sup>13</sup>

## Freeing up time

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 higher-order tasks. A centralized data catalog helps you eliminate duplicative work, understand usage, and better connect all of your teams to high-quality data.

### QUESTIONS TO ASK YOUR TEAM

- What activities take up most of your team's time?
- What steps aren't being executed properly?
- What's getting in the way of data being used and applied in a timely manner?
- Where can streamlined tools free up capacity?

# Build a scalable data strategy to drive revenue

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 data team and business users.

Based on your analysis, a centralized data catalog could be key to unlocking the value of data to grow your business. For a full analysis of the potential impact for your organization, [book a comprehensive ROI review](#) with 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 while retaining visibility and control. Our unified cloud platform cuts overhead, fuels innovation, and drives revenue growth.

### Sources

1. Eckerson, "Trends in Data Catalogs: The Promise and Reality"
2. McKinsey, Designing Data Governance that Delivers Value
3. Gartner, How to Create a Business Case for Data Quality Improvement
4. Capgemini, Data Sharing Masters
5. NewVantage Partners, Data and AI Leadership Executive Survey, 2022
6. GDPR Enforcement Tracker, <https://www.enforcementtracker.com/>
7. Burtch Works, The Burtch Works Salary Report, 2022 Edition: Data Science & AI Professionals
8. Burtch Works
9. Forrester, Total Economic Impact Report
10. Forrester, Better Data Quality Equals Higher Marketing ROI
11. Capgemini, How Data Quality Can Hurt Your Data Science Programme... If You're Not Careful
12. Data professional at a major financial institution (ThinkData Works client)
13. Anaconda, State of Data Science 2021