

# The enterprise data infrastructure benchmark report 2026

A global benchmark of spend and its impact on AI readiness and ROI

## Introduction

**Enterprises spend an average of \$29.3 million per year on data programs — and \$2.2 million of that goes to keeping data pipelines running.**

Data has become one of the largest line items in enterprise technology budgets, driven by the need to power analytics, AI, and real-time decision-making. Yet despite unprecedented investment, most organizations struggle to translate spend into impact. Pipeline failures, downtime, and manual operations quietly consume millions of dollars each year — eroding productivity, delaying AI initiatives, and limiting returns.

To understand the true state of enterprise data infrastructure, Fivetran surveyed more than 500 senior data and technology leaders from organizations with over 5,000 employees across industries and global regions. The findings reveal that the problem is not underinvestment but architecture. Most enterprises continue to rely on closed, brittle, and labor-intensive data integration that cannot scale with growing data volumes, sources, or AI-driven use cases.

As environments grow more complex, fragmented and tightly coupled systems break more often, require constant human intervention, and restrict access to data across teams. The result is a widening gap between organizations that treat data integration as an operational burden and those that view it as a foundational architectural layer designed for openness, interoperability, and trusted access.

This report benchmarks the true cost of data operations across spend, reliability, engineering effort, and ROI — and shows how architectural choices at the integration layer determine whether data investments accelerate analytics and AI or quietly cap their potential.

## Executive summary

Annual data budgets average more than \$29 million per organization, and a significant share of that spend is lost to pipeline failures, downtime, and the headcount required to maintain increasingly complex data environments.

Below are the 3 most important insights from this year's study:

### **1. Operational inefficiency — not lack of investment — is slowing analytics and AI at scale.**

Data teams dedicate 53% of engineering time to maintenance, with \$2.2 million per year is spent on pipeline upkeep by full-time engineers. Legacy and DIY pipelines break 30-47% more

often, resulting in 60 hours of downtime per month. Data leaders estimate the business impact of downtime at \$49,600 per hour — putting nearly \$3 million in potential business value at risk each month, or more than \$36 million annually. These impacts are most pronounced in environments where pipelines are tightly coupled and require frequent manual intervention as they scale.

## **2. Integration spend is higher than expected, yet ROI remains low.**

Enterprises allocate 14% of their total data budget (~\$4.2 million) to data integration, yet only 27% of organizations report that their data investments exceed ROI expectations. Much of this spend is absorbed by manual, fragmented integration approaches that become increasingly difficult to manage as data volumes, sources, and use cases expand.

With an average of 328 pipelines and as many as 35-60 full-time engineers maintaining them, organizations bear higher costs due to manual, fragmented processes that struggle to scale efficiently.

## **3. Pipeline modernization delivers lower costs, greater reliability, and higher ROI.**

By adopting standardized and fully managed pipelines, organizations secure the reliability and cost efficiency required to build AI-ready environments — a high-performance cohort that delivers faster recovery times and is nearly 2x as likely to exceed ROI targets (45% vs. 27%). As maintenance overhead decreases, teams are able to redirect effort toward analytics, AI, governance, and other initiatives that depend on consistent, shared, and trusted access to reliable data.

# **Chapter 1: Underperforming pipelines are limiting ROI**

Enterprise investment in data has reached new heights, driven by ambitious plans for AI, real-time analytics, and personalized customer experiences. On average, large organizations now spend \$29.3 million annually on data, reflecting the central role data plays in competitiveness and innovation. Yet despite this level of investment, 73% of organizations report their data initiatives fall short of expectations.

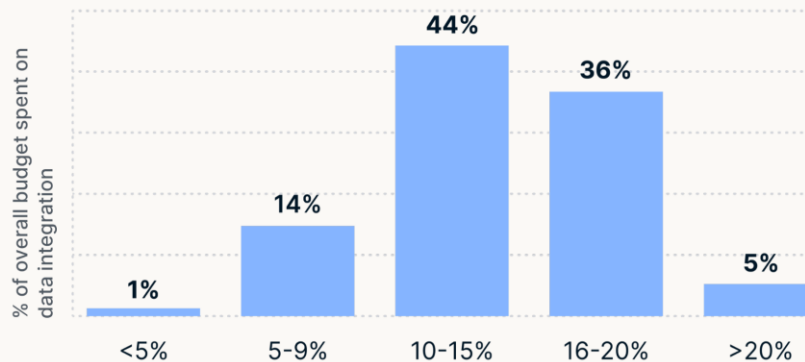
A significant portion of this spend is tied directly to data integration, including:

- 14% of the total data budget (~\$4.2 million) — including tools, infrastructure, and internal staff on data movement, ingestion, and preparation
- \$500,000+ per month on cloud ingest and compute costs
- \$2.2 million per year in engineering labor spent specifically on pipeline maintenance — a burden that is significantly higher among teams relying on DIY or legacy data integration

## % of overall budget spent on data integration

By annual data budget

Average % spent on data integration	\$1M-\$4.9M	\$5M-\$9.9M	\$10M-\$49.9M	\$50M or more
	10%	13%	15%	16%



Avg. estimated spend: **\$4.2M**

As overall data budgets grow, the proportion of those funds required to maintain data integration steadily increases — reaching 16% for the largest enterprises.

These investments reflect increasingly urgent business priorities. Teams rely on data to power AI models, deliver personalized customer experiences, and support real-time decision-making. However, operational bottlenecks, particularly fragile pipelines and labor-intensive workflows, continue to slow progress. As expectations rise, so does the pressure to improve the reliability and efficiency of the infrastructure already in place.

“73% of enterprise data initiatives fail to meet expectations despite \$29.3 million in annual data spend.”

## AI ambitions outrun data operations

While investment is rising, nearly 62% of organizations report their data maturity as “low,” revealing how far enterprises still have to go to achieve fully self-service and predictive analytics.

Our findings show that core operational capabilities — reliability, automation, observability, and standardized integration — have not kept pace with strategic ambitions. Enterprises need to

scale AI by democratizing data access and reducing decision latency, yet the underlying systems are often too manual or brittle to support these goals.

Pain points include:

- Pipeline breaks and downtime undermine data freshness
- Maintenance demands consume more than half of engineering resources
- Legacy and DIY systems drive higher costs and slower recovery times

These challenges grow exponentially at scale. Large enterprises have an average of 328 to over 400 pipelines, making incremental fixes and manual processes unsustainable and locking teams into pipeline maintenance rather than innovative projects.

Enterprises that have modernized their data integration layer with automated, fully managed pipelines are progressing toward AI and advanced analytics. Those still relying on legacy or DIY approaches are falling behind with infrastructure that cannot scale.

“62% of enterprises still report low data maturity, as fragile pipelines and manual operations consume 53% of engineering time.”

## **Chapter 2: Benchmarking the cost of fragile, maintenance-heavy pipelines**

Despite record levels of data investment, the foundation supporting enterprise analytics and AI remains brittle. Most organizations rely on DIY pipelines that are increasingly complex, difficult to scale, and expensive to maintain — creating a drag on productivity, reliability, and strategic execution.

The benchmarks below illustrate how deeply pipeline fragility affects cost, operations, and business outcomes.

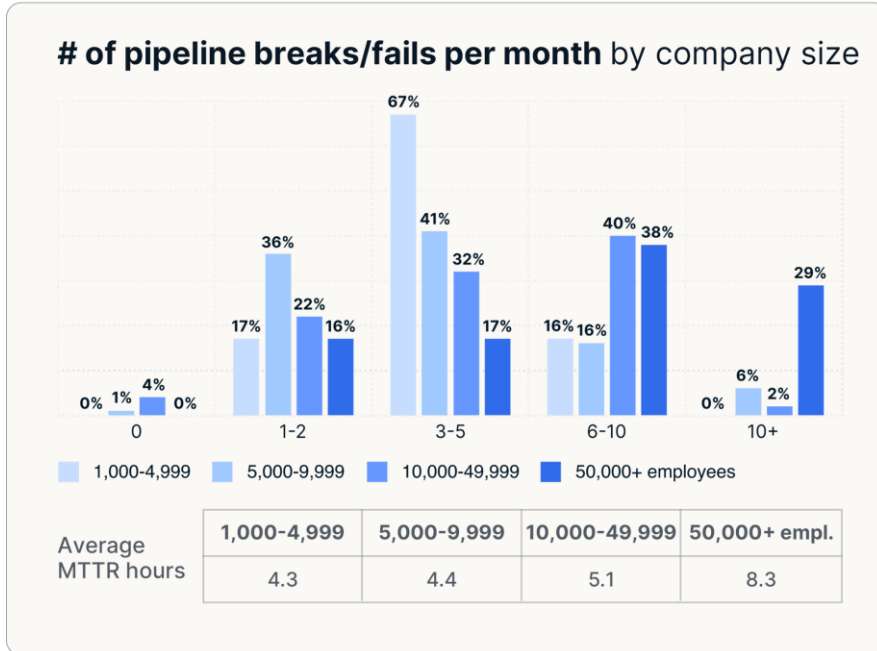
### **Pipeline reliability and downtime**

Enterprise data operations are failing at a rate that meaningfully disrupts operations — even as many organizations are still early in operationalizing data for AI.

Organizations experience:

- 4.7 pipeline breaks per month on average
- 60.4 hours of downtime every month
- Up to 8.3 breaks per month in the largest enterprises

While AI ambition is high, only a minority of enterprises report predictive or AI models in production, and data readiness and reliability gaps continue to delay analytics and AI delivery.



Downtime is costly and escalates with scale:

- **\$49,600 per hour** in estimated business impact from data downtime
- Rising to **\$75,200 per hour** of downtime in large enterprises

These disruptions degrade AI models, harm dashboard accuracy, and slow decision-making. Leaders across industries report that unreliable data integration routinely delays analytics and AI initiatives, creating inefficiencies that ripple through the business, with nearly 30% of organizations facing analytics and AI project delivery delays of a month or more due to pipeline maintenance or downtime issues.

## Engineering time, headcount, and hidden labor costs

Data integration reliability shows up most clearly in how much effort teams spend keeping pipelines running. When integrations aren't reliable, engineering time shifts from innovation to maintenance.

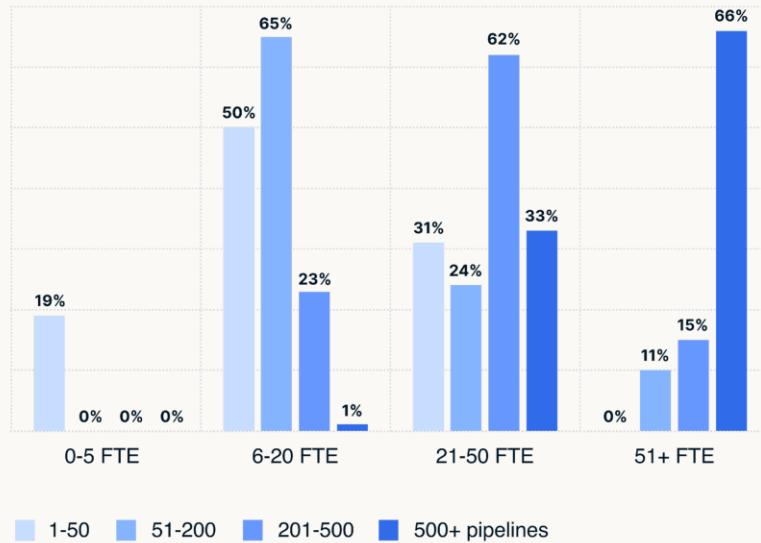
Organizations report:

- **53% of engineering time** spent on pipeline maintenance
- **35 full-time engineers** managing an average of 328 pipelines
- **50+ full-time engineers** managing 500+ pipelines in large enterprises

## # of pipelines with dedicated FTEs

by # of active data sources/pipelines maintained

Avg. # of pipelines maintained: 328



Legacy and DIY integration approaches worsen the operational load:

- **30-47%** more frequent pipeline breaks
- 2-4 additional hours required for repairs
- **25% higher cost per pipeline** (legacy \$1,900 vs. fully managed ELT \$1,600)

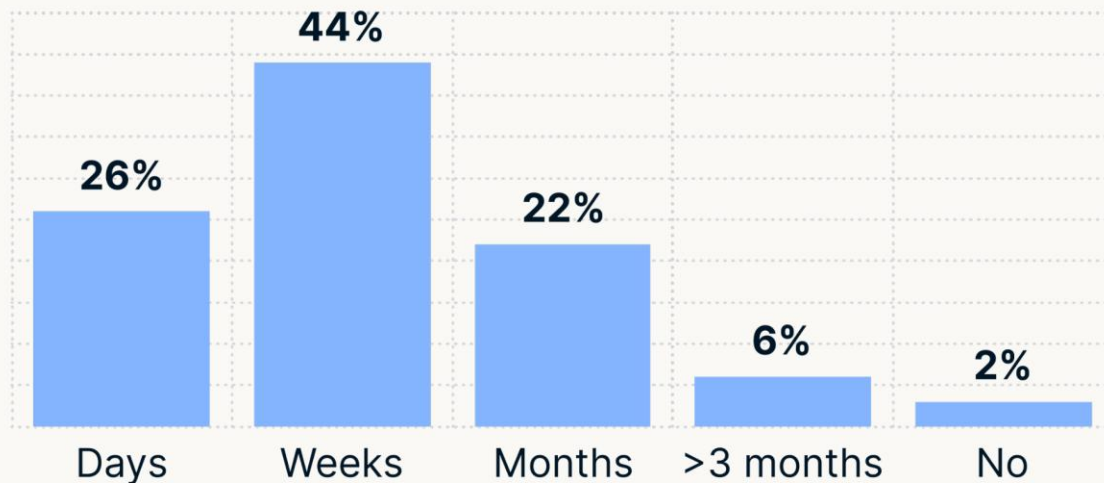
High full-time engineers-to-pipeline ratios, intensive manual intervention, and limited automation reduce operational agility and keep engineering teams in reactive mode.

## Business impact of integration failures

Unreliable data integration directly impacts enterprise performance:

- **97%** of enterprises report disruptions to AI or analytics initiatives
- **70%** of enterprises report negative impacts on customer personalization or cost-reduction projects
- **\$36 million-\$54 million** estimated annual business impact (i.e., lost revenue potential) from lack of timely access to fresh data

## Analytics or AI project delays caused by pipeline maintenance or downtime



When organizations depend on stale or incomplete data, AI models underperform or fail to deploy, personalization loses accuracy or effectiveness, strategic programs slow down due to uncertainty or lack of trusted insights, and data teams are pulled into reactive firefighting instead of driving innovation.

“The hidden cost of data integration failures: \$2.2 million in annual maintenance, 60+ hours of downtime per month, delayed AI initiatives, and reduced capacity for innovation.”

## Chapter 3: The benefits of modernizing data management

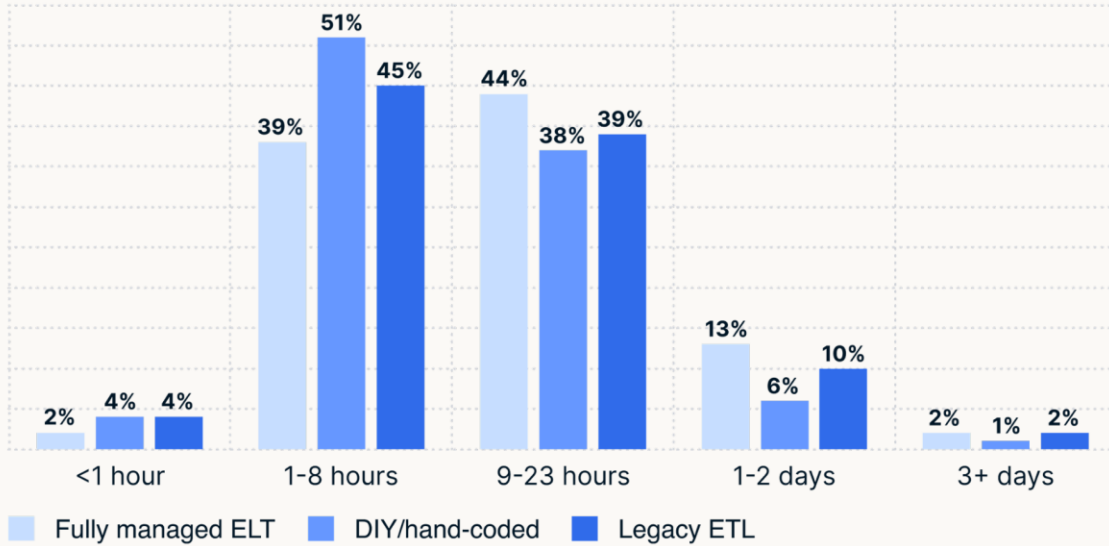
After years of incremental fixes, mounting technical debt, and rising maintenance overhead, enterprises are recognizing that legacy and DIY integration approaches can no longer support modern data management requirements.

In contrast, fully managed and automated ELT has emerged as the foundation for AI, lowering total cost of ownership, stabilizing data operations, freeing data teams to focus on innovation, and accelerating time to value.

### Lower costs and improved reliability

The cost difference between legacy and modern data integration approaches is significant and widening. Organizations using legacy ETL spend \$1,900 per pipeline, compared to \$1,600 with fully managed ETL. The difference adds up quickly at enterprise scale, where hundreds or even thousands of pipelines must be maintained.

## Median time to resolve (MTTR) by dominant pipeline management approach



Average MTTR hrs.	11.3	15.5	13.4
\$ equivalent	\$550k	\$784k	\$688k

Legacy systems also break far more frequently. Enterprises relying on custom scripts or homegrown ETL solutions report 30-47% more pipeline failures and longer recovery times, with incidents typically taking 13–16 hours to resolve. In contrast, organizations that modernize with fully managed ELT to achieve AI readiness reduce recovery times to just 11 hours. By reducing both break frequency and time to recovery, fully managed ELT improves operational predictability and materially lowers risk.

The result is a measurable improvement in uptime and reliability. With fewer failures and faster recovery, downstream analytics and AI systems have more consistent access to fresh data. Engineers can rely on their pipelines, business users can trust their data, and leaders can rely on timely insights to guide decisions.

“\$300 saved per pipeline annually with automated data integration — a 6-figure savings at enterprise scale.”

### Reclaiming engineering capacity

One of the most transformative impacts of modernization is the shift in how data teams allocate their time. In legacy environments, engineering teams spend 53% of their time on pipeline maintenance and reactive support. Modern data integration doubles the time teams have to work on strategic initiatives that drive business value.

With automated, fully managed pipelines:

- Engineers spend less time diagnosing failures and more time enabling the business
- AI initiatives and predictive modeling advance more quickly
- Governance, compliance, and real-time analytics programs gain momentum
- Self-service data access becomes broader and safer, reducing dependency on hard-to-find experts

Replacing brittle custom scripts with standardized, automated data integration reduces dependence on specialized tribal knowledge. Instead of relying on domain experts to troubleshoot failures or maintain complex workflows, teams operate within a predictable operating model — improving agility and enabling scale without proportional headcount growth.

As operational burdens decline, enterprises redirect capacity to higher-value initiatives. High-performing data teams report increased focus on:

- Predictive analytics (69%)
- Self-service analytics (65%)
- Governance and compliance (59%)
- AI innovation (56%)

## Higher ROI from more reliable data

Investments in fully managed, automated ELT deliver measurable business returns. By establishing the stable foundation necessary for AI-ready environments, organizations using fully managed ELT are nearly 2x as likely to exceed their ROI targets (45% vs. 27%), reinforcing the connection between operational discipline and financial performance.

Three factors drive this amplification effect:

- **Repeatability:** Automated pipelines reduce variability and minimize error, ensuring consistent and standardized data delivery
- **Scalability:** Fully managed ELT scales with data volumes and sources without adding engineering effort
- **Reliability:** Predictable data flows improve insight quality, strengthen model performance, and speed decision-making

Organizations with mature data operations also demonstrate higher operational effectiveness, including greater adoption of predictive models, broader self-service analytics, and tighter alignment between data teams and business stakeholders. By reducing time spent on infrastructure, enterprises unlock the capacity needed to innovate rapidly and at scale.

“45% exceed ROI targets with fully managed ELT vs. 27% using DIY or legacy approaches.”

## Key implications for data leaders

**Organizations that modernize their data operations outperform peers on cost efficiency, reliability, and AI readiness.**

The data shows that the greatest gains come not from incremental fixes, but from moving away from fragile, maintenance-heavy integrations toward operating models that are easier to manage, reuse, and scale as data environments grow. Based on this benchmark, the path forward is clear. To reduce the total cost of data, increase ROI, and accelerate AI readiness, data leaders should:

## **1. Evaluate total cost holistically**

Look beyond tooling costs and account for the full operational burden of data integration, including engineering time, downtime, break-fix cycles, and downstream business impact. Hidden costs — such as \$49,600 per hour of downtime and \$2.2 million in annual maintenance spend — often outweigh the price of the tool itself, making legacy and DIY approaches far more expensive than they appear.

## **2. Prioritize automation to reduce maintenance toil**

With 53% of engineering time consumed by integration maintenance, automation is the fastest way to reclaim capacity. Reducing manual intervention nearly doubles the time available for higher-value engineering work, without increasing headcount.

## **3. Standardize data integration to reduce operational risk**

Fragmented, custom integrations introduce variability and risk as data environments expand. Legacy and DIY approaches break 30-47% more often and cost 25% more per pipeline to maintain. Standardized integration improves reliability, consistency, and scalability as complexity grows.

## **4. Strengthen data foundations before scaling AI**

AI initiatives depend on fresh, consistent, and well-governed data. Without predictable data delivery, even well-funded AI programs stall or underperform. Building a strong data foundation ensures AI investments can be supported reliably in production.

## **5. Align spend to business outcomes — not infrastructure maintenance**

Shift investment away from maintaining brittle systems toward platforms that accelerate analytics, improve data quality, and support AI readiness. By adopting fully managed ELT, organizations establish the reliable foundation necessary for AI-ready environments and are nearly 2x as likely to exceed ROI expectations, demonstrating how operational maturity compounds returns across the data ecosystem.

Data budgets alone do not create value. Organizations that pair investment with integration practices designed for scale, operational reliability, and broad access to data are the ones achieving higher ROI, faster AI progress, and sustained competitive advantage.

For data leaders, the opportunity is clear: shift investment away from fragile, maintenance-heavy integration and toward modern data foundations built for reliable analytics and scalable AI.

## Methodology & demographics

This report is based on a global survey of 500 senior data and technology leaders conducted in Q4 2025 at a 95% confidence level ( $\pm 4.4\%$  margin of error) across the United States, United Kingdom, EMEA, and APAC. The survey examined 4 core areas of enterprise data operations: operating model and scale, cost and maintenance, data capability, access, and flexibility, and budget, ROI, and AI readiness. Findings were analyzed to identify global patterns in pipeline reliability, operational maturity, integration spending, and the impact of modernization on ROI.

Company size skewed toward larger organizations, with 70% from companies with 5,000-9,999 employees, 23% from 10,000-49,999 employees, and 5% from enterprises with 50,000+ employees.

The survey captured a broad mix of industries, including financial services (24%), manufacturing (20%), technology (20%), retail/CPG (19%), healthcare (15%), and hospitality (3%). Respondents held senior roles such as VP of Data Engineering/Data Integration/Data Analytics (26%), VP of IT/IS/Enterprise Systems/Technology (21%), CIO (16%), CTO (16%), CDO (16%), and Chief Digital Officer (5%).