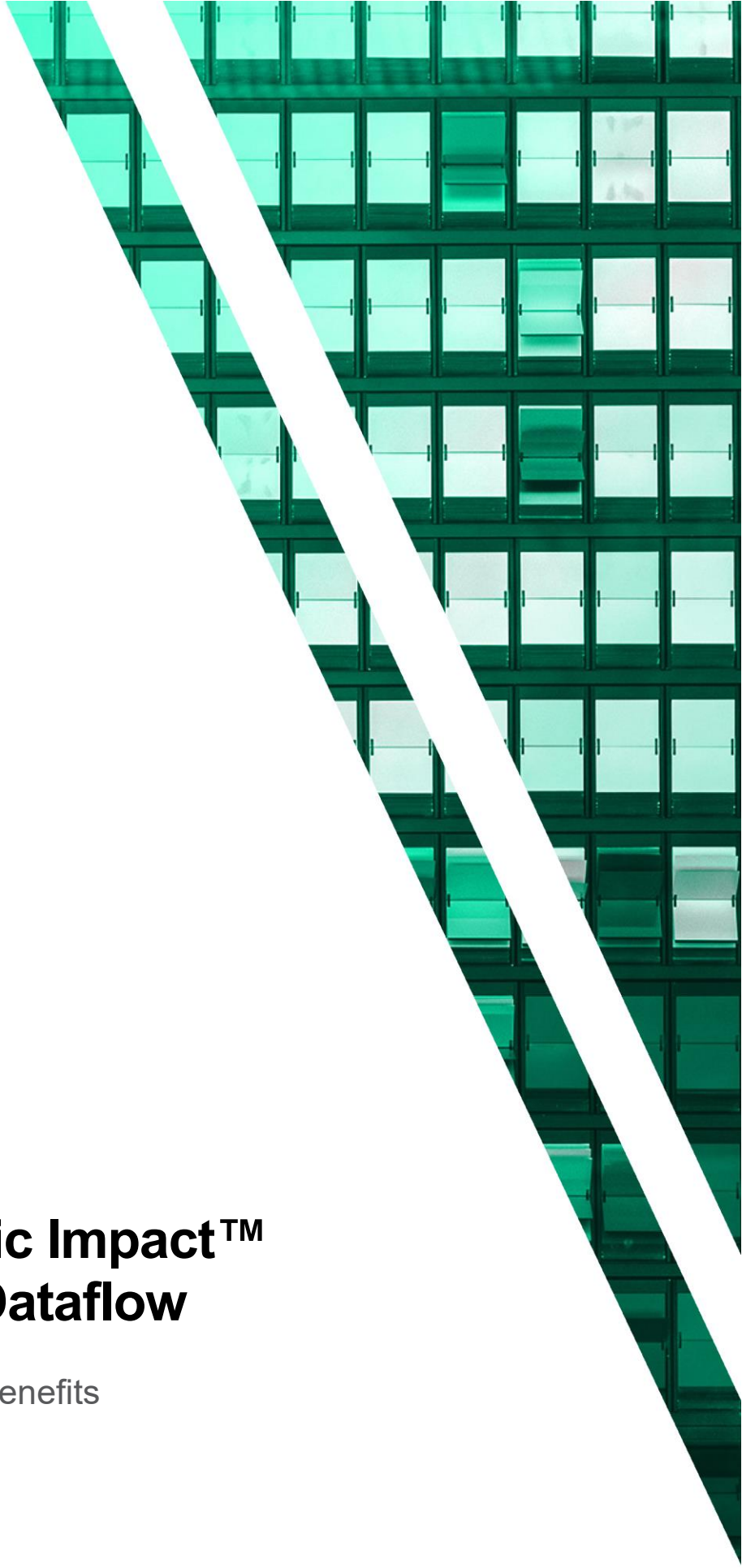
FORRESTER®



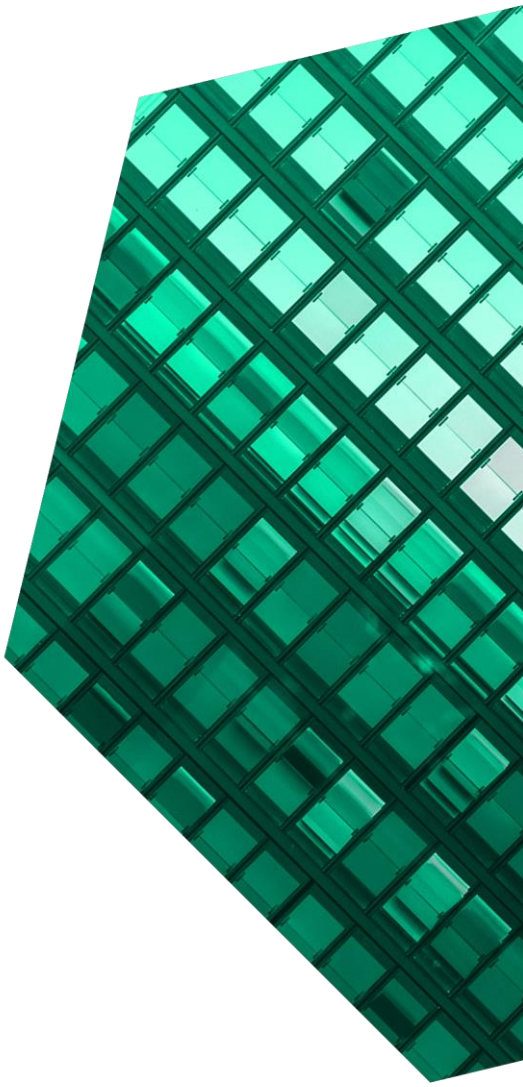
The Total Economic Impact™ Of Google Cloud Dataflow

Cost Savings And Business Benefits Enabled By Dataflow

FEBRUARY 2022

Table Of Contents

Consulting Team: Benjamin Brown
Kara Luk



ABOUT FORRESTER CONSULTING

Forrester Consulting provides independent and objective research-based consulting to help leaders succeed in their organizations. For more information, visit forrester.com/consulting.

Executive Summary

Organizations must turn raw, real-time data into actionable insights to win, serve, and retain customers. Cloud-native streaming data platforms provide critical scale, time-to-value, and speed to act rapidly on insights while also releasing organizations from operational burden and excess overhead costs. Interviewed decision-makers turned to Dataflow — Google Cloud's autoscaling, serverless managed service for both streaming and batch data processing — to empower their insights-driven business transformations.

Google Cloud Dataflow is a cloud-native managed service for both streaming and batch data processing and analytics. The service automatically scales underlying resources up and down, eliminating operational overhead. Dataflow offers a full range of data processing and analytical capabilities with performance, availability, and resiliency built in.

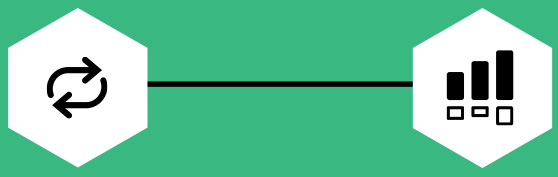
Developers can choose among a variety of programming languages to define and execute data workflows. Dataflow also seamlessly integrates with other Google Cloud Platform and open source technologies to maximize value and applicability to a wide variety of use cases.

Google commissioned Forrester Consulting to conduct a Total Economic Impact™ study examining the potential ROI that organizations may realize by deploying Google Cloud Dataflow.[1] The purpose of this study is to provide readers with a framework to evaluate the potential financial impact of Dataflow on their organizations.

Data and financial analysis. Forrester interviewed seven decision-makers from five organizations with experience using Dataflow to better understand the investment's benefits, costs, flexibility, and risks. Forrester aggregated customer data into a composite organization with a representative financial analysis.

Customer journey. Customers migrated to Dataflow primarily from legacy on-premises solutions and alternative cloud infrastructure that were fractured across regions and business units. These organizations struggled with excess costs and operational labor to support the environments. They could not scale adequately or enable real-time insights, tarnishing customer experience (CX) and hindering business growth.

KEY STATISTICS



Return on investment (ROI)
171%

Net present value (NPV)
$5.01M

Decision-makers invested in Dataflow to:

- Enable new revenue drivers such as recapturing missed sales, offering product suggestions, and executing relevancy marketing.
- Implement anomaly detection and fraud detection to protect systems and reduce revenue loss.
- Improve customer experience by enabling features such as real-time data access during support calls.
- Drive better global consistency and governance by standardizing on Dataflow for batch and stream processing across compute locations worldwide
- Reduce technology costs and operational labor.

- Shift-left teams to conduct more valuable work and improve employee experience (EX).
- Improve performance, resiliency, and trust.
- Generate insights and value from data in real time using streaming.
- Improve and drive business growth.
- Empower and expand insights-enabled innovation and decision-making.

KEY FINDINGS

Quantified benefits. Three-year risk-adjusted present value (PV) quantified benefits for the composite organization include:

- **Drove business growth through CX and new data-enabled offerings, increasing operating profit by $2.5 million.** Dataflow enabled real-time streaming use cases, improved data enrichment, encouraged data exploration, improved performance and resiliency, reduced errors, increased trust, and eliminated barriers to scale. As a result, organizations provided customers with more accurate, relevant, and in-the-moment data-backed services and insights — boosting CX, creating new revenue streams, and improving acquisition, retention, and enrichment.
- **Reduced infrastructure costs by 50% for streaming workloads and by 10% for batch workloads, saving $2.8 million.** Dataflow's serverless autoscaling and discrete control of job needs, scheduling, and regions eliminated overhead and optimized technology spending. Consolidating global data processing solutions to Dataflow further eliminated excess costs while ensuring performance, resilience, and governance across environments.
- **Boosted data engineer efficiency by 55%, recapturing $1.5 million in labor.** Dataflow streamlined workflows with code reusability, dynamic templates, and the simplicity of a managed service. Engineers trusted pipelines to run correctly and adhere to governance. Data engineers avoided laborious issue-monitoring and remediation tasks that were common in the legacy environments such as poor performance, lack of availability, and failed jobs. Teams valued the language flexibility and open source base.

> **"Overall, Dataflow is an amazing system in how it can handle the volume of data with guarantees and assured results."**
>
> *Senior software engineer, B2B technology*

- **Eliminated administrative overhead and toil, reallocating $847,000 in labor.** As a cloud-native managed service, all administration tasks such as provisioning, scaling, and updates are automatically handled by Google Cloud. Teams no longer need to manage servers and related software for legacy data processing solutions. Admins also streamlined processes for setting up data sources, adding pipelines, and enforcing governance. Organizations reallocated systems administrators to conduct more valuable work, expanding the usage of data for insights.

> **"Dataflow is integral to accelerating time-to-market, decreasing time-to-production, reducing time to figure out how to use data for use cases, focusing time on value-add tasks, streamlining ingestion, and reducing total cost of ownership."**
>
> *Lead technical architect, CPG*

- **Saved business operations costs for support teams and data end users, recapturing $263,000 in labor.** Dataflow improved the speed, quality, reliability, and ease of access to data for

insights for general business users, saving time and empowering users to drive better data-backed outcomes. It also reduced support inquiry volume while automating manual job creation.

Qualitative benefits. Dataflow enabled the following unquantified benefits for organizations:

- Improved reliability, enhanced performance, and reduced latency.
- Accelerated time-to-market.
- Improved data accessibility and usage, driving more data-backed decision-making.
- Enhanced employee experience, improved skill value, and increased organizations' abilities to attract, hire, and retain employees.
- Improved customer experience and outcomes.
- Provided pricing transparency, predictability, and scalability/elasticity.
- Bettered the ability to track, prove, and ensure multiregion compliance.
- Reduced risk of lock-in with open source.

"The jobs we migrated from on-premises to Google were a double-whammy: We got more features and we paid less, which is great. … The amount we saved justified the investment that we had to put in to do it."

Senior software engineer, media

Flexibility. Customers are evaluating the potential to use Dataflow to:

- Tackle new business opportunities not possible with legacy solutions.
- Enable new artificial intelligence (AI) and machine learning (ML) use cases.
- Achieve further savings by using Dataflow in tandem with other Google Cloud and open source technologies.
- Expand savings to other business units.

"It's already been proven that we are getting more business [with Dataflow] because we can turn around results faster for customers."

VP of technology, financial services technology

Costs. Three-year risk-adjusted PV quantified costs for the composite organization include:

- **Dataflow consumption costs of $2.2 million.** Costs are based on the amount of data processed and specific demands for how, where, and when data needs to be processed. Cost scales directly with usage, avoiding overhead.
- **Planning, implementation, and training costs of $517,000.** Dataflow implementations were often a small part of six- to 24-month full cloud migrations, typically requiring labor equivalent to approximately three FTEs. Implementation costs should be much lower if an organization already has established cloud infrastructure or if a multi- or hybrid cloud strategy is employed.
- **Ongoing management costs of $256,000.** Dataflow is a managed service and requires very minimal administration; however, a relatively small amount of tuning and optimizing is important to drive business benefits and savings.

Results. Forrester's risk-adjusted financial analysis for a composite organization shows benefits of $7.95 million versus costs of $2.94 million over three years, adding up to a net NPV of $5.01 million, a payback period of less than six months, and an ROI of 171%.

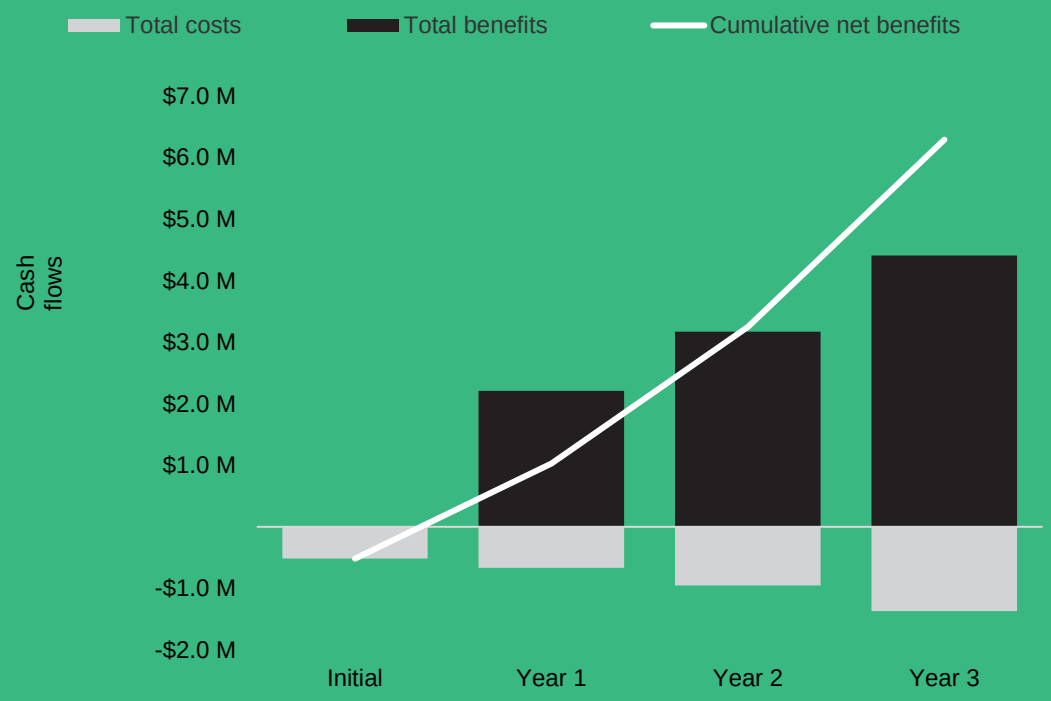
ROI
171%

BENEFITS PV
$7.95M

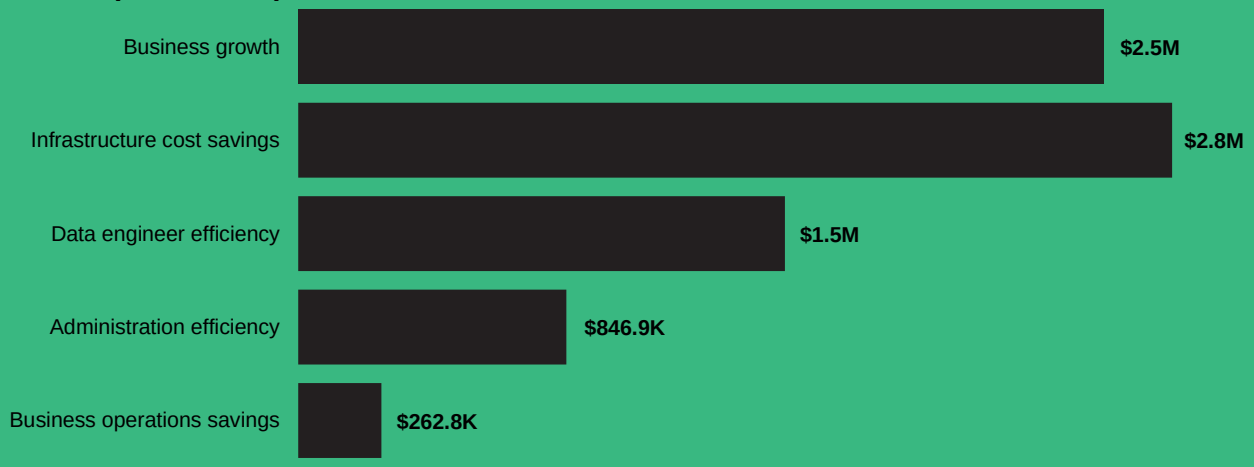
NPV
$5.01M

PAYBACK
<6 months

Cash Flow Chart (Risk-Adjusted)



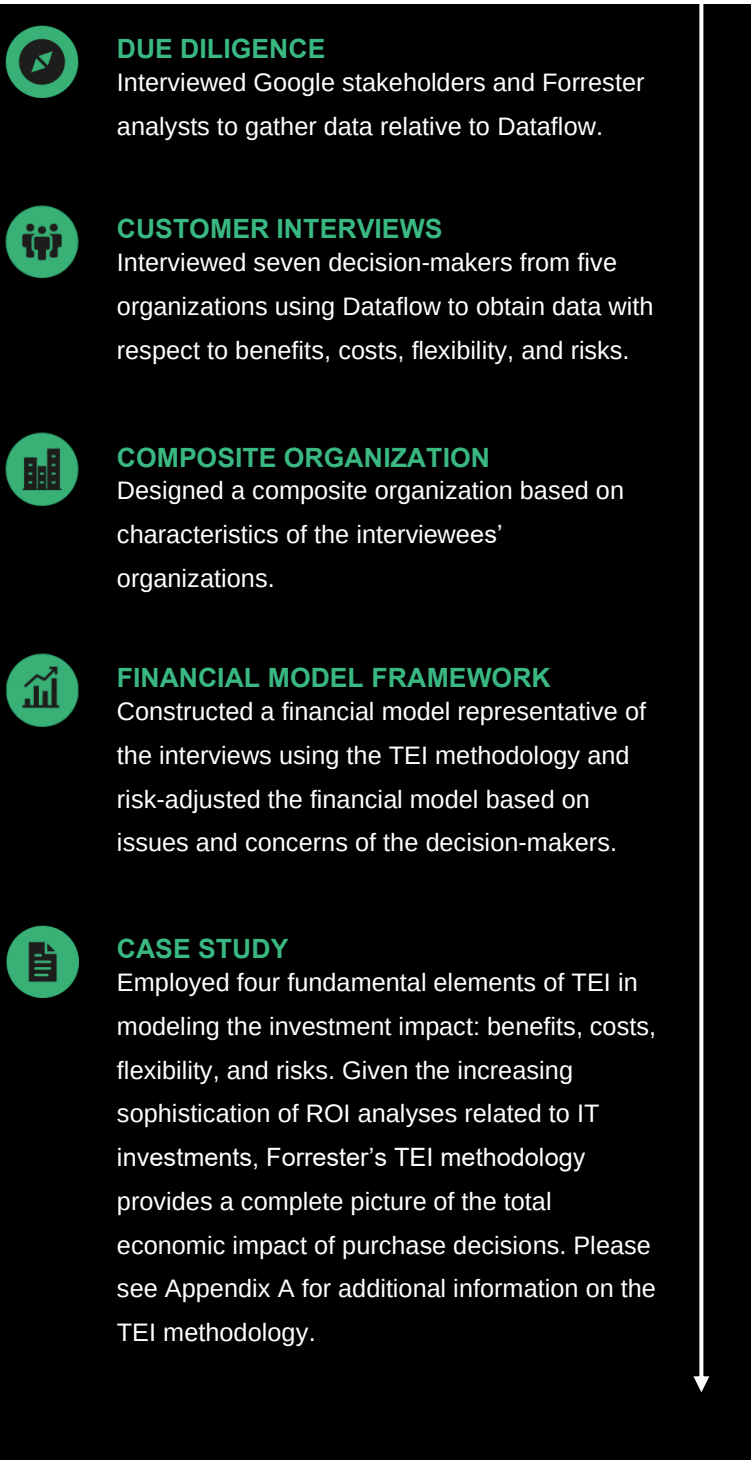
Benefits (Three-Year)



TEI FRAMEWORK AND METHODOLOGY

From the information provided in the interviews, Forrester constructed a Total Economic Impact™ framework for those organizations considering an investment in Google Cloud Dataflow.

The objective of the framework is to identify the cost, benefit, flexibility, and risk factors that affect the investment decision. Forrester took a multistep approach to evaluate the impact that Dataflow can have on an organization.



The Google Cloud Dataflow Customer Journey

Drivers leading to investments in Google Cloud Dataflow

MARKET OVERVIEW

Turning raw, real-time data into actionable insights is increasingly critical to firms' ability to win, serve, and retain customers. Organizations are processing ever-larger volumes of data, both at rest and in motion, yet all this data is only valuable insofar as they can actually do something with it.[2]

Barriers facing organizations today. Unfortunately, more than half of organizations find their inability to manage data volume to be very or extremely challenging.[3] The troubling truth for many is that the pace of innovation and availability of data has outstripped their ability to keep up because:

- **On-premises environments have restricted flexibility and innovation.** Capital expenditures lock firms when they need to be flexible.[4] Meanwhile, internal staff can't keep pace with change requirements or implement new innovations at scale fast enough.[5] The gap only grows as technologies' changes multiply faster than skills resources can keep up.[6]
- **Organizational and technological siloes increase the insight-to-action gap.[7]** More data should mean better insights, but lack of visibility, siloed data, and different data formats hinder insight generation and the ability to act. Only 43% of firms have a cross-platform view of their data, while 57% are challenged by siloed data that is managed by different teams.[8]
- **Legacy technologies cannot act in real time.** Traditional architectures are failing to meet new business requirements, especially around high-speed data streaming, real-time analytics, large volumes of messy and complex data sets, and self-service.[9] Organizations that still fully rely on on-premises infrastructure and batch processes are unable to return insights in the moment.

Organizations must form an insights-driven business transformation. Breaking through the barriers of legacy technologies and siloes is critical: As organizations accumulate more data, the time and effort needed to act on data multiplies.[10] And as years go by, delayed action becomes more and more disruptive as firms realize they suddenly need to scramble to catch up with other organizations that are achieving exponential change.[11]

To accelerate their journey to become insights-driven, organizations need an enterprisewide insights-driven business transformation that orchestrates multiple enabling technologies, platforms, solutions, and professional services.[12] Solutions must be simple, agile, integrated, cost-effective, and highly automated.[13]

> **"We are conducting a concerted cloud migration campaign that's all about data, quality assurance, and speed."**
>
> *Director of data projects, financial services*

> **"Customers move quickly, so we must operate at the speed of the customer to give them a relevant and engaging experience."**
>
> *Director of data projects, financial services*

Streaming data platforms enable the speed that organizations need. Speed is an essential requirement for today's enterprises when it comes to data. Enterprises succeed with immediate insights, smarter decisions, and speed-of-light application

integration.[14] Real-time context or observations of consumer behaviors and environments drive better decisions and outcomes for customers than do historical data or best guesses from artificial intelligence.[15]

Organizations therefore must learn to retrieve the right data at the right time. What happened yesterday happened yesterday — it's too late. Yet enterprise architectures that evolved from the predigital world, where data hopped slowly from databases to data warehouses for eventual rearview-mirror analytics, are untenable.[16] Instead, technology leaders can turn to streaming data platforms to become fast digital data enterprises.[17]

The benefits of a unified streaming and batch architecture compound over time. Streaming analytics is an essential capability in today's digital world to gain real-time actionable insights. Likewise, organizations must also have flexible, high-performance batch environments to analyze historical data for building machine learning models, business intelligence, and advanced analytics. A unified streaming and batch data platform gives organizations the flexibility to define either workload in the same programming model, run it on the same infrastructure, and manage it from a single operational management tool.

Streaming data platforms are enterprise software that connects to multiple, disparate live data sources (streaming data) to expose and/or deliver to subscribing applications, analyzes to gain real-time insights, and/or recombines for continuous integration to downstream data repositories or applications.

Key uses for streaming solutions. Technology leaders can use streaming data platforms to deliver data from one-to-many applications, analyze data in real time, and process data continuously.[18] It's easy, fast, and scalable, and it unifies architectures by enabling every application to become a source and sink for streaming data.[19] Streaming data platforms further help organizations:

- Replace yesterday's latency-laden technology with modern streaming data flow platforms to leapfrog legacy approaches and standardize how data is delivered at scale. [20]
- Enhance the value of data streams for analytics use cases such as continuous intelligence and dashboards, IoT applications, and real-time ML and AI. [21]
- Standardize application interconnectivity to become more real-time, connect applications and services with ease, scale efficiently, and drive agility to meet new and changing business requirements faster and more cheaply. [22]
- Strengthen real-time anomaly detection and fraud prevention by using AI and ML capabilities in conjunction with streaming analytics.
- Deploy real-time recommendation engines to enhance user experiences with personalized offerings, information, or content at the right time.
- Run extract, transport, and load (ETL) and extract, load, and transform (ELT) pipelines in real time.

Selecting the right streaming solution. Streaming data platform decisions must also consider the broader architecture including cloud providers, data warehouses and storage, messaging, analytics, and AI. These solutions must together provide ease of use, richness of capabilities, and scalability. Architectures must scale to handle maximum data throughput and concurrency, all while supporting high-availability features and enabling serverless

scale-down to minimize costs. Ideally, streaming solutions should be able to self-monitor and scale by automatically provisioning and deprovisioning resources to meet the load.[23]

Selecting the right streaming solution. Streaming data platform decisions must also consider the broader architecture including cloud providers, data warehouses and storage, messaging, analytics, and AI. These solutions must together provide ease of use, richness of capabilities, and scalability. Architectures must scale to handle maximum data throughput and concurrency, all while supporting high-availability features and enabling serverless scale-down to minimize costs. Ideally, streaming solutions should be able to self-monitor and scale by automatically provisioning and deprovisioning resources to meet the load.[24]

Introduction to Google Cloud Dataflow. Dataflow is a cloud-native managed service for both streaming and batch data processing and analytics. The platform is serverless and automatically scales with elasticity, eliminating operational overhead. It offers the ability to perform simple and complex operations with performance, availability, and resiliency. Developers can write using a variety of languages to build jobs in open source Apache Beam, providing flexibility and ease of use while ensuring application portability. Dataflow is often used in tandem with the Google Cloud Platform (GCP) portfolio such as Cloud Storage, BigQuery, Bigtable, and Pub/Sub.[25]

Interviewed Decision-Makers

Interviewees	Industry	Region	Annual Revenue	Dataflow Consumption
• Senior software engineer • Software engineer	B2B technology	North America	Less than $1 billion	1,500 daily jobs More than 100 TiB per month
• Lead technical architect • Software engineer	Consumer packaged goods (CPG)	Global, based in North America	More than $20 billion	Hundreds of daily jobs 150 TiB per month
VP of technology	Financial services	Global, based in North America	$1 billion to $10 billion	More than 500 daily jobs More than 500 TiB per month
Director of data projects	Financial services	Europe	More than $20 billion	Thousands of daily jobs More than 100 PiB per month
Senior software engineer	Media	Global, based in North America	$1 billion to $10 billion	Thousands of daily jobs More than 500 TiB per month

DATA COLLECTION

Forrester interviewed seven decision-makers from five organizations that utilize Dataflow. These organizations operate in a range of industries and regions and employ 200 to 50,000 FTEs.

Four interviewed customers migrated to Dataflow from on-premises infrastructure and tools, while one migrated from an alternative cloud architecture. Some workloads were migrated while others were rearchitected to utilize new streaming capabilities.

Organizations also typically changed data storage, warehousing, analytics, and AI/ML not long before or after the investment in Dataflow. Most customers adopted Dataflow as part of broader migrations to the cloud, primarily in tandem with other Google Cloud Platform services and open source technologies.

All customers use Dataflow extensively and continue to expand their environments — consuming many petabytes of data per year for thousands of jobs. Customers used both Dataflow's batch and streaming capabilities. Today, streaming jobs accounted for up to 50% of data volume processed by customers, but decision-makers noted that they expect to invest even more in streaming over the coming years.

Voice Of The Customer

"Dataflow is elastic. You pay per second for your compute resources. You pay per byte of data processed. If you're not running a job, you're not paying anything. There's no VMs just sitting there idle. It autoscales jobs and VMs up and down. It's a great model. There aren't any hidden costs because the billing is transparent in Google's documentation, and it's easy to extrapolate out as long as you are careful."

— Senior software engineer, media

"Dataflow provided us with ETL replacement that opened limitless potential use cases and enabled us to do smarter data enhancement while data remains in motion."

— Director of data projects, financial services

"The more we started to play around with Dataflow, the more we noticed how well it scaled and it just kind of fit into place."

— Lead technical architect, CPG

KEY CHALLENGES

The interviewees noted how their organizations struggled with common challenges, including:

- **Excess technology costs that could not scale.** Legacy tools, processes, and on-premises infrastructure were far too expensive to meet business needs, forcing businesses to be overly selective in their use of data. On-premises costs were "always on," incurring cost regardless of usage, but simultaneously could not scale up further at critical moments.
- **Excess labor to support legacy environment.** Organizations had to support too many data tools across disparate, siloed data centers. Setting up, maintaining, and upgrading disparate on-premises infrastructure and tools required significant labor with a drudging EX. Every new workload or scale request incurred excess time, making discussions of scale also a discussion of administrative labor. Teams required niche skills to work on legacy tools, and expertise often was not interchangeable.
- **Difficulty enforcing security, privacy, and compliance.** Ensuring governance over multiple solutions and siloed data centers posed major challenges for organizations trying to protect their environment and ensure compliance. Proprietary formats made it difficult to inspect data.
- **Processing, analyzing, and acting on data was too slow — hurting decision-making, CX, and business performance.** Infrastructure could not scale without significant labor, while data work was stuck using batch processes and complex ETL routines. Business performance and customer experience suffered without the ability to provide value in real time. Business decisions either were not made using objective data and insights at all or relied too heavily on historical trends or small sample sizes.

Voice Of The Customer

"We were spending too much time supporting too many different platforms across many geographies within the company, and the talent was not interchangeable."

— VP of technology, financial services

"After [facing some security issues], we needed to redesign, back up, and increase data security for applications. We needed global data storage and governance, and we moved to Google Cloud Platform."

— Anonymous interviewee

"Our biggest challenge was scale. We were having difficulty scaling up to meet new load demands. Also, the operational burden of these systems was nontrivial. Another challenge was that everything was in proprietary formats that were then difficult to inspect, debug, and analyze."

— Senior software engineer, media

INVESTMENT OBJECTIVES

Interviewed decision-makers investigated changing their data architecture to:

- Unify data processing across the organization, including across regional barriers.
- Reduce technology costs.
- Streamline operational labor overhead.
- Shift-left developers, IT administrators, and data engineers to conduct more valuable work and improve EX.
- Enforce governance and gain visibility across regional and business unit siloes.
- Decrease time-to-production for data projects and reduce barriers for data teams to test concepts and use data.
- Generate insights and value from data in real time using streaming.
- Enhance capabilities for data processing, analytics, and AI/ML.
- Make better, objective data-backed decisions based on objective data and insights instead of subjective guessing.
- Improve CX with better, more accurate, more dependable, and faster insights and services based on data.
- Minimize business risk with consumption-based pricing and open source architecture.
- Drive cloud migration.

"With Dataflow, Google does the scaling and capacity conservations for us … Dataflow is a first-class citizen for GCP with a lot of support."

VP of technology, financial services

SOLUTION REQUIREMENTS

Decision-makers selected Google Cloud Dataflow because Dataflow:

- Allows developers to code workflows using a variety of popular languages, enabling ease of use and flexibility and empowering staff with a variety of backgrounds to learn Dataflow.
- Provides both batch and streaming capabilities with the same architecture.
- Is a managed service that eliminates the need for internal teams to manage installs, upgrades, maintenance, and provisioning.
- Autoscales from nearly zero as a serverless service to meet maximum demand, without labor or overhead costs.
- Seamlessly integrates with other cloud-native services in Google Cloud Platform.
- Enables portability as companies build open source Apache Beam jobs.

"There are three tenets of our new data platform including Dataflow: accelerate time-to-value, grow by trying new data concepts, and reduce total cost of ownership."

Director of data projects, financial services

"Dataflow fit both batch ETL and streaming needs, could run as a persistent or nonpersistent resource, had rigor and reliability, and had great adaptability."

Director of data projects, financial services

COMPOSITE ORGANIZATION

Forrester aggregated findings from the customer interviews with research data and industry metrics to design a composite organization and an associated ROI analysis that is representative of interviewees' experiences with Dataflow.

Composite description. The composite organization has the following characteristics:

- Is a business with global operations that is based in North America.
- Generates $5 billion in annual revenue with an operating profit margin of 10%.[26]
- Employs 20 data engineers at $125,000 in fully burdened annual salary.

Legacy solution. Prior to Dataflow, the composite organization's data processing and analytics had the following characteristics:

- Data processing and analytics tools ran in on-premises infrastructure.
- Data processing and analytics tools were siloed by region and business unit.
- All data processing was done using batch processes only.
- A team of three infrastructure administrators managed the legacy environment, earning $85,000 in fully burdened annual salary.

Deployment characteristics. Forrester's Dataflow TEI analysis for the composite organization is based on the following representative assumptions:

- The composite conducts a complete migration of its data storage, processing, and analytics environment to Google Cloud, including migrating to Dataflow for all data processing. The cloud migration takes 12 months, split into two six-month phases for planning and deployment.

Key assumptions

- **Global business based in North America with $5B in annual revenue**
- **Replaces siloed on-premises data tools with Dataflow**
- **Uses both streaming and batch workflows in Dataflow**
- **Processes 2 to 4.5 PiB of data per month with Dataflow**

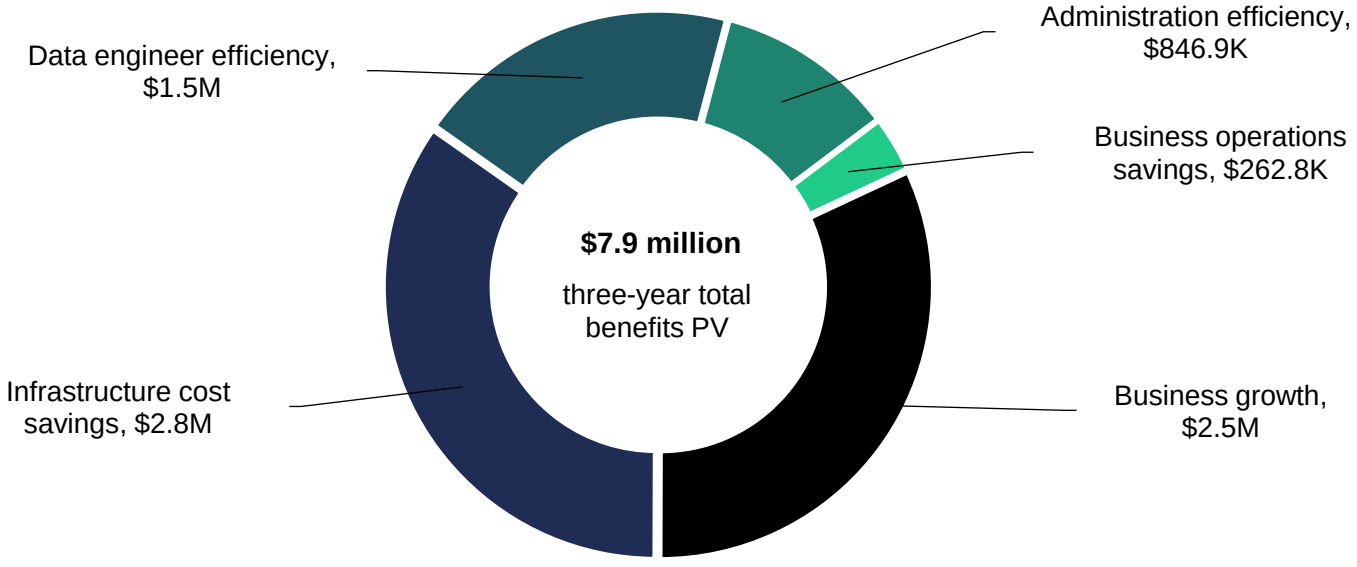
- The migrated environment processes 2 pebibytes (PiB) of data per month with Dataflow. Initially, 25% of data processing jobs are rearchitected as streaming jobs (0.5 PiB) while the other 75% are migrated as batch workflows (1.5 PiB).
- Over the next two years, the composite continues to convert batch jobs to streaming and build more data processing workflows. By Year 3, the composite processes 4.5 PiB of data per month with Dataflow at an even 50% split for streaming and batch workflows (2.25 PiB each).

Analysis Of Benefits

Quantified benefit data as applied to the composite

Total Benefits						
Ref.	**Benefit**	**Year 1**	**Year 2**	**Year 3**	**Total**	**Present Value**
Atr	Business growth	$528,000	$1,056,000	$1,584,000	$3,168,000	$2,542,810
Btr	Infrastructure cost savings	$660,387	$1,050,605	$1,716,934	$3,427,925	$2,758,578
Ctr	Data engineer efficiency	$617,760	$617,760	$617,760	$1,853,280	$1,536,278
Dtr	Administration efficiency	$340,560	$340,560	$340,560	$1,021,680	$846,922
Etr	Business operations savings	$64,770	$108,460	$152,150	$325,380	$262,831
	Total benefits (risk-adjusted)	$2,211,477	$3,173,385	$4,411,404	$9,796,265	$7,947,419

Three-Year, Risk-Adjusted Total Benefits



BUSINESS GROWTH

Data is only valuable when used; Dataflow enabled organizations to vastly improve the ways in which they used their data and the outcomes of that usage.

- Dataflow improved data enrichment, connecting more sources and enriching data to make it more meaningful and useful.
- Dataflow streaming enabled real-time usage of data in motion, vastly increasing the value of the use cases with immediate relevancy.
- Dataflow enabled teams to explore, test, and deploy new use cases for data more quickly and easily. These new use cases drove additional value for customers, employees, and margins.
- Dataflow improved resiliency, availability, and scale — ensuring that data was always ready for whatever uses were demanded of it.
- Dataflow reduced errors and provided greater visibility to actions taken, improving trust in the accuracy of insights.

These factors enabled businesses to improve the speed, quality, reliability, and accessibility for existing data workflows while also enabling teams to increase the value of workflows with streaming and create entirely new use cases for data. Dataflow also democratized and encouraged the use of data, extending its value throughout the enterprise. It is crucial to have immediate, accurate, and trustworthy data for the myriad of potential data use cases from personalization to marketing and from financial decision-making to business operations. Through these improvements, Dataflow enabled businesses to protect current business (retention), drive larger and stickier customer relationships (enrichment), gain new customers (acquisition), and grow top-line revenue.

"It's already been proven that we are getting more business [with Dataflow] because we can turn around results faster for customers."

VP of technology, financial services

"We realized that by improving analytics, we could drive revenue."

Senior software engineer, media

Evidence and data. Interviewed decision-makers shared how Dataflow benefited their organizations:

- **A financial services company increased the speed and relevancy of marketing and customer service, recapturing missed sales worth several percent of total revenue.** Marketing that was too slow failed to meet buyers at their time of need and generate revenue. Customer service was stuck using outdated data during calls, with follow-up action taking days, hurting CX and causing customer churn.
- **A media company improved CX and data quality for customers and business partners, driving retention and upselling.** Outdated data, slow response times, poor performance, and incomplete or inaccurate data hurt CX and muted customer spending. With Dataflow, the company now provided data to customers in real time with streaming that was also more accurate, reliable, and complete. The company used to resort to sampling approximately 5% of data, but with Dataflow it processed 100% of the data to provide better insights. Customers and business partners became happier, reducing complaints, preventing churn, and driving upsell.

- **A financial services company improved CX and increased revenue with Dataflow.** The company was able to deploy new data jobs for customers in 3 hours instead of 10 days and process data for customers in near real time. This significantly improved CX, business outcomes for customers, and revenue from customers (via retention and enrichment).

"When we provide data to our customers and partners with Dataflow, we are much more confident in those numbers and can provide accurate data within a minute. Our customers and partners have taken note and commented on this. It's reduced complaints and prevented churn."

Senior software engineer, media

Modeling and assumptions. Forrester modeled the impact for the composite organization assuming:

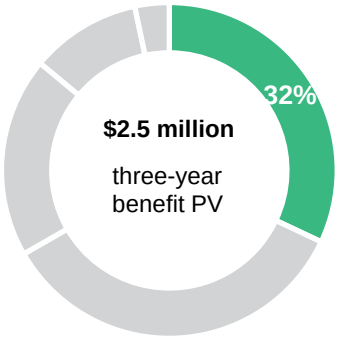
- The composite earns $5 billion in annual revenue at a 10% operating profit margin from the business units that deploy Dataflow.
- Before Dataflow, the composite lost customers and missed out on customer acquisition due to providing data or services based on data that was inadequate, incorrect, or too slow to be relevant. The value of missed business opportunities was 4% of annual revenue.
- The composite's new cloud data ecosystem improves CX and enhances offerings, recapturing 10% to 30% of missed opportunities.
- Dataflow is a critical driver of recaptured revenue opportunities but is not the only factor. Only 33% of revenue growth is attributed to Dataflow.

Risks. The expected financial impact is subject to risks and variation based on several factors that may reduce or slow recognition of this benefit, including:

- Scale, breadth, and user- or business-focus of new and migrated use cases for Dataflow.
- Level of ability to provide fast, relevant, and high-quality data services in legacy environments and the degree to which Dataflow transforms CX.
- Size of relevant business revenue opportunities and the associated profit margin.
- Potential for external market or business conditions to affect revenue.
- Ability to measure and attribute metric impact.

Results. To account for these risks, Forrester adjusted this benefit downward by 20%, yielding a three-year, risk-adjusted total PV of $2.5 million.

Business Growth



Business Growth					
Ref.	**Metric**	**Source**	**Year 1**	**Year 2**	**Year 3**
A1	Annual revenue for business segments leveraging data processed with Dataflow	Composite	$5,000,000,000	$5,000,000,000	$5,000,000,000
A2	Typical percentage of business lost or missed due to inadequate, incorrect, or too-slow data output	Interview data	4%	4%	4%
A3	Reduction in lost business from data-driven CX enhancements and new data-driven use cases	Interview data	10%	20%	30%
A4	Percentage of revenue capture attributable specifically to Dataflow	Forrester	33%	33%	33%
A5	Increased revenue attributable to Dataflow from protected and new business	A1*A2*A3*A4	$6,600,000	$13,200,000	$19,800,000
A6	Operating profit margin	NYU Stern School of Business	10%	10%	10%
At	Business growth	A5*A6	$660,000	$1,320,000	$1,980,000
	Risk adjustment	↓20%			
Atr	Business growth (risk-adjusted)		$528,000	$1,056,000	$1,584,000
	Three-year total: $3,168,000			**Three-year present value: $2,542,810**	

INFRASTRUCTURE COST SAVINGS

Dataflow significantly reduced the technology costs for processing data while also providing additional scalability, speed, and capabilities:

- Dataflow's serverless autoscaling was the primary driver of savings — neither Dataflow batch nor streaming jobs incur costs while idle.
- Organizations could further reduce costs by leveraging cheaper times, cheaper regions, and Flex scheduling in Google Cloud.
- Dataflow delivered better performance, resiliency, and governance, helping organizations eliminate duplicate jobs and processing steps.
- Consolidating global data processing to Dataflow for batch and streaming reduced excess costs that were caused by redundant technologies and infrastructure for different regions and teams.

Migrating to Dataflow batch and streaming reduces three-year, risk-adjusted present value technology costs for data processing by **18.7%** for the composite organization.

Migrating batch workloads from on-premises environments to Dataflow batch jobs reduced workload cost by an average of 10%, while rearchitecting workflows as Dataflow streaming jobs slashed workload costs by 50% through optimizing data consumption and leveraging autoscaling to the utmost. As organizations move to real-time streaming, the benefits of using Dataflow's unified batch and streaming model to standardize data processing workloads compound. Meanwhile, Dataflow jobs also provided greater value than their legacy equivalents due to the speed, resilience, performance, and added capabilities available with the solution. Dataflow also saved teams from having to resort to common control measures to stay within resource constraints such as regimenting job scheduling, reducing job frequency, or using samples

rather than full data sets — all of which were hurting the ability to use data for value.

Evidence and data. Interviewed decision-makers shared how Dataflow benefited their organizations:

- **The CPG company slashed tech costs from "millions of dollars" to only $120,000 per year.** Major drivers included scalability, timing flexibility, efficient consumption, and avoiding licenses with open source. The lead technical architect shared how Flex enabled savings: "Dataflow Flex lets us dynamically initialize different transforms based on configuration of the new job versus past jobs, rather than making templates for every possible condition. We can consolidate labor effort and optimize our compute to run jobs a lot faster and save money."
- **A financial services firm reduced net costs by 30% to 35% with Dataflow, saving millions of dollars per year.** The director of data projects shared: "Item by item, Dataflow is cheaper and drives significant cost savings. It lets us avoid spending on unneeded resources and shift workloads to times with lower-cost compute." They continued: "Cloud isn't inherently cheaper, but it is more adaptable, with more levers to control costs, clearing out waste like duplication and driving significant total cost savings."
- **The media company reduced per-job costs by 10% to 15%.** Jobs used to run twice to maintain availability, and now each runs once, with further savings from consumption efficiency, elasticity, headroom, Shuffle, and Flex.
- **Another financial services firm is cutting costs as it consolidates many regions.** The broader the deployment, the more savings, because costs were much lower with Dataflow according to the VP of technology: "In one of our experiments with Dataflow, we were able to prove that we could process 1 billion records, taking them from a file all the way through multiple Dataflow jobs, for less than $6.50."

> **"Our costs with our cloud data platform using Dataflow are just a fraction of the costs we faced before. Now we only pay for cloud infrastructure consumption because the open source base helps us avoid licensing costs. We spend about $120,000 per year with Dataflow, but we'd be spending millions with our old technologies."**
>
> *Lead technical architect, CPG*

Modeling and assumptions. Forrester modeled the impact for the composite organization with the following base case:

- The composite organization incurs costs for Dataflow consumption as calculated in Cost Table F: Dataflow Consumption.
- Batch jobs cost an average of 10% less in Dataflow than in the legacy infrastructure.
- Streaming jobs cost an average of 50% less than their closest equivalent batch processes running in legacy infrastructure.
- The composite avoids these additional costs by deploying Dataflow, as the data processing was required regardless of solution.

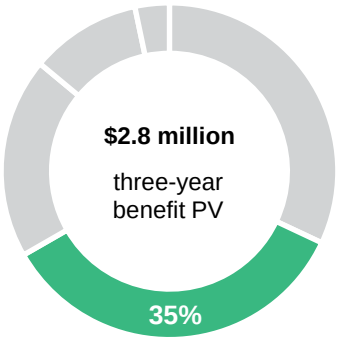
Risks. The expected financial impact is subject to risks and variation based on several factors that may reduce or slow recognition of this benefit, including:

- Size and speed of Dataflow deployment.
- Data volume, resource needs, regions, times, and workflow steps per specific job. Cost-saving techniques such as lower-cost times and regions or FlexRS are important for maximum savings.

- Actual legacy costs and architecture, and the speed and ability to decommission environments.
  - Migrating from other cloud data processing solutions may result in less technology cost savings than migrating from on-premises.
  - No technology cost savings would be realized for greenfield Dataflow usage rather than as a replacement for legacy tools.

Results. To account for these risks, Forrester adjusted this benefit downward by 10%, yielding a three-year, risk-adjusted total PV (discounted at 10%) of $2.8 million.

Infrastructure Cost Savings



Infrastructure Cost Savings

Ref.	Metric	Source	Year 1	Year 2	Year 3
B1	Annual Dataflow batch cost	F4*12	$378,528	$504,708	$567,792
B2	Savings for batch jobs using Dataflow instead of legacy solutions	Interview data	10%	10%	10%
B3	Avoided costs for batch jobs with Dataflow instead of legacy solutions (rounded)	B1/(1-B2)	$420,587	$560,787	$630,880
B4	Annual Dataflow streaming cost	F6*12	$156,588	$303,276	$638,412
B5	Savings for redesigning legacy batch jobs as streaming Dataflow jobs	Interview data	50%	50%	50%
B6	Avoided costs by redesigning legacy batch jobs with Dataflow streaming	B4/(1-B5)	$313,176	$606,552	$1,276,824
Bt	Infrastructure cost savings	B3+B6	$733,763	$1,167,339	$1,907,704
	Risk adjustment	↓10%			
Btr	Infrastructure cost savings (risk-adjusted)		$660,387	$1,050,605	$1,716,934
	Three-year total: $3,427,925			**Three-year present value: $2,758,578**	

DATA ENGINEER EFFICIENCY

Dataflow significantly streamlined data engineer workflows by enabling code reusability, dynamic templates, and the simplicity of a managed service. Engineers could trust that pipelines would run correctly and follow organizational governance. Data engineers also avoided the laborious tasks of monitoring and fixing issues that were common in the legacy environments such as poor performance, lacking availability, and failed jobs. Teams valued the flexibility and ease of selecting between a variety of languages to utilize Dataflow in their workflows.

Evidence and data. Interviewed decision-makers shared how Dataflow benefited their organizations:

- **A financial services company more than doubled data engineer capacity.** The director of data projects shared: “We've liberated 50% to

60% of [data engineering resources] with Dataflow. These teams that used to waste significant time reacquiring and retransforming can now reuse work and avoid duplication. They wasted substantial effort investigating the sources and transformations for data to prove adherence to compliance requirements. Now we have trust and assurance in the data."

- **The B2B technology company accelerated pipeline development with Dataflow.** The software engineer shared: "Once you learn Dataflow, you can write pipelines very fast. The magic happens and you don't have to worry about issues like crashing or disappearing data."
- **The CPG company used code reusability to slash time to build data jobs from days and weeks to minutes for data engineers.** The lead technical architect shared: "There was no code reusability before our new data solution based on Dataflow. You had to write every new thing from scratch and go through red tape like change review boards, approvals, and other minutiae. We had to reinvent the wheel every time. Now, building a new job has gone from days and weeks to minutes at the most."
- **A financial services firm sharply reduced the time to fulfill data requests, freeing data engineer capacity.** Engineers can now handle more pipelines, constituents, and analyses — directly benefitting revenue-producing customers. The VP of technology shared: "[Data engineers] were spending 40 to 50 hours to fulfill requests for one of our large customers. When they migrated to GCP and Dataflow, they could now fulfill them in 2 hours. This was a huge win."

Dataflow reduced data engineering labor by **55%** for pipeline creation and management, increasing team capacity and accelerating fulfillment.

"Apache Beam in Dataflow is the backbone of our new open source based data platform. All our data acquisition flows through Dataflow, and we've constructed our Dataflow jobs to be generic and dynamic so we can run all our normal and flex templates the same way with a few standard design patterns and templates."

Lead technical architect, CPG

Modeling and assumptions. Forrester modeled the impact for the composite organization assuming:

- The composite employs 20 data engineers that use Dataflow to write and update data jobs.
- Engineers save 55% of their time with pipeline creation efficiencies in Dataflow.
- Conservatively, only 50% of time saved for salaried engineers is recaptured for added value.

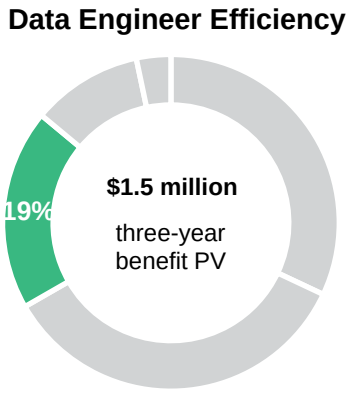
"Before Dataflow, no one wanted to write streaming jobs because it was too difficult and frustrating. No one did it at all. Now, 40% to 50% of all new jobs being built are streaming. Dataflow and the Google Cloud ecosystem have made data more accessible and easily usable. We are making more data-driven decisions where our teams build dashboards, analyze data, and operate off the findings. Dataflow powers a good chunk of it."

Senior software engineer, media

Risks. The expected financial impact is subject to risks and variation based on several factors that may reduce or slow recognition of this benefit, including:

- Number of data engineers and their salaries.
- Unique business requirements and complexities that may reduce potential time savings.
- Actual labor for the legacy environment and the speed and ability to decommission environments.
  - Migrating from other cloud data processing solutions may result in less labor savings than migrating from on-premises.
  - No labor savings would be realized for greenfield Dataflow usage rather than as a replacement for legacy environments.

Results. To account for these risks, Forrester adjusted this benefit downward by 10%, yielding a three-year, risk-adjusted total PV of $1.5 million.



Data Engineer Efficiency

Ref.	Metric	Source	Year 1	Year 2	Year 3
C1	Number of data engineers	Composite	20	20	20
C2	Percentage of time saved for engineers by switching to Dataflow	Interview data	55%	55%	55%
C3	Productivity recapture rate for salaried employees	Forrester	50%	50%	50%
C4	Total hours recaptured for added business value	C1*2,080*C2*C3	11,440	11,440	11,440
C5	Fully burdened hourly salary for data engineers	Payscale	$60	$60	$60
Ct	Data engineer efficiency	C4*C5	$686,400	$686,400	$686,400
	Risk adjustment	↓10%			
Ctr	Data engineer efficiency (risk-adjusted)		$617,760	$617,760	$617,760
	Three-year total: $1,853,280			**Three-year present value: $1,536,278**	

ADMINISTRATION EFFICIENCY

Dataflow is a cloud-native managed service that requires very minimal management. All administration tasks such as provisioning, scaling, and updates are handled by Google Cloud. Teams no longer needed to deploy and manage servers and related systems used for legacy data processing solutions (disaster recovery, OS, storage, data software updates, etc.). Teams also streamlined the process for setting up new data sources for ingestion with cloud APIs and common development tools.

Consolidating batch and streaming into Dataflow and consolidating technologies in different regions enabled significant savings required to address the

different solutions and compliance requirements globally, especially for large companies. Most organizations also found their demand for and usage of data to be continually growing, requiring constant new deployments and provisioning to meet scale demands. This effort was no longer needed with Dataflow.

"We saved a lot of developer hours not having to maintain the old system. Shutting those off freed us to do other things."

Senior software engineer, media

Evidence and data. Interviewed decision-makers shared how Dataflow benefited their organizations:

- **The CPG company reallocated 12 admins from ingestion and management to value-add modeling and insights.** Teams previously had to write and deploy code for every task; with Dataflow, fewer people could support more sources with configurations that took only minutes. The lead technical architect explained: "Fewer people can now support more ingestion sources. The people that were working on the ingestion pipeline have been assigned to downstream projects where they're working on modeling or insights activities. They're developing models or training prediction models that provide more horsepower to get insights faster."
- **A financial services firm reallocated a four-person operations team.** Dataflow released the team from needing to manage the legacy on-premises environment, allowing the ops team to focus instead on deploying and managing other new technologies for the organization.
- **A financial services company enabled one team to support all regions where Dataflow is deployed.** Each region had its own data centers and tools with a team of five to 10 FTEs for managing the systems; with Dataflow, these local teams could be reallocated for more valuable work. The organization has deployed to three regions thus far and is working to reach many more.
- **The media company reduced complexity by rearchitecting with Dataflow for all batch and streaming.** The resulting system requires much less labor for management, including streamlined issue identification and debugging. What time is still spent on Dataflow is now dedicated to ongoing tuning and optimization, boosting benefits for the organization. The senior software engineer shared: "We save a lot of time on operational tasks, and we see it across teams as they move to Dataflow. It's a significant savings. We used to have a very large operational burden. Worse, it would be unpredictable — randomly, people would need to work extra hours without notice, which is way worse than knowing how much time we need to spend each week."

"Reliability is better in all ways with Dataflow — system reliability, data reliability, and avoiding issues while running jobs."

Senior software engineer, media

Modeling and assumptions. Forrester modeled the impact for the composite organization assuming:

- The composite reallocates four systems administrators (alternatively, IT admins) previously dedicated to managing the legacy on-premises data environments.

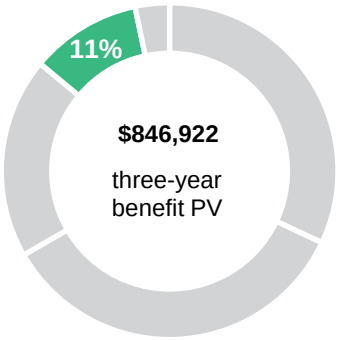
- The composite eliminates labor for scaling up legacy data solutions to keep pace with demand, saving four weeks for a developer and a data engineer once per quarter. Fifty percent of time saved is recaptured for these salaried FTEs.

Risks. The expected financial impact is subject to risks and variation based on several factors that may reduce or slow recognition of this benefit, including:

- Number of IT administrators and data engineers and their respective salaries.
- Actual labor for the legacy environment and the speed and ability to decommission environments.
  - Migrating from other cloud data processing solutions may result in less labor savings than migrating from on-premises.
  - No labor savings would be realized for greenfield Dataflow usage rather than as a replacement for legacy environments.
- Unique business requirements and complexities that may reduce potential time savings.

Results. To account for these risks, Forrester adjusted this benefit downward by 10%, yielding a three-year, risk-adjusted total PV of $847,000.

Administration Efficiency



Administration Efficiency					
Ref.	**Metric**	**Source**	**Year 1**	**Year 2**	**Year 3**
D1	Reallocated systems administrator FTEs for legacy infrastructure	Interview data	4	4	4
D2	Fully burdened annual salary for systems administrators	Payscale	$85,000	$85,000	$85,000
D3	Infrastructure administration efficiency	D1*D2	$340,000	$340,000	$340,000
D4	Avoided annual hours for developers and data engineers to scale legacy solutions to meet demand	Interview data	1,280	1,280	1,280
D5	Productivity recapture rate for salaried employees	Forrester	50%	50%	50%
D6	Fully burdened hourly salary for developers and data engineers	Payscale	$60	$60	$60
D7	Recaptured scalability efficiency	D4*D5*D6	$38,400	$38,400	$38,400
Dt	Administration efficiency	D3+D7	$378,400	$378,400	$378,400
	Risk adjustment	↓10%			
Dtr	Administration efficiency (risk-adjusted)		$340,560	$340,560	$340,560
	Three-year total: $1,021,680			**Three-year present value: $846,922**	

BUSINESS OPERATIONS SAVINGS

Dataflow improved the speed, quality, reliability, and ease of access of data workflows for general business users, saving time and empowering users to drive better data-backed outcomes. Dataflow reduced issues and requests, reducing support inquiry volume and cost. It also enabled companies to automate the preparation of data for immediate usage, saving time for data engineers and for users who can use and act on the data immediately. Dataflow also empowered data engineers to build new data jobs for business needs in a fraction of the time, driving further end-user efficiency and business value. As with business growth, Dataflow can power an immense breadth of potential use cases and users that benefit from improved data access.

Dataflow enabled real-time, automated data requests for end users and empowered support teams with real-time data for effective and efficient service.

Evidence and data. Interviewed decision-makers shared how Dataflow benefited their organizations:

- **The media company improved the speed, reliability, availability, and accuracy of data for customers — deflecting issues.** By deflecting issues, the company saved significant support and remediation labor while also protecting revenue. The senior software engineer explained, “Dataflow has certainly reduced complaints, cut down on debugging requests, and prevented churn.”
- **A financial services company enabled real-time data access during support calls and reduced customer service demand.** With hundreds of millions of customer interactions per year, empowering agents with accurate real-time data was critical to efficiency and quality of service. However, customer service was previously stuck using outdated data during calls, and follow-up action took days, hurting CX and causing customer churn.
- **The CPG company saved time for general business end users with faster access to data and insights.** Dataflow enabled teams to provide real-time data for customers through integrated self-service, allowing end users to get and explore the data they needed to do their work.

Modeling and assumptions. Forrester modeled the impact for the composite organization assuming:

- The composite builds user-facing services that eliminate the need for 500 to 1,500 unique data requests, saving business users 4 hours per request of lost productivity while waiting; salaried FTEs had a 50% productivity recapture rate.
- Before Dataflow, the composite’s support teams handled 1,000 support cases per year caused by inadequate, incorrect, or too-slow data. Deploying Dataflow deflects 30% to 50% of these support interactions, saving 4 hours of labor per issue for hourly support teams.

“Providing customer care is remarkably expensive, especially when you can't see what's going on with the customer in real time. Using legacy systems, we'd have to collect information, analyze it after the call, and then call back — driving excess costs. With Dataflow, we can now work together in real time.”

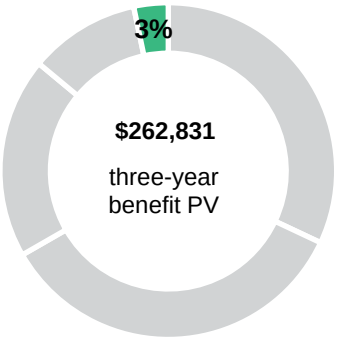
Director of data products, financial services

Risks. The expected financial impact is subject to risks and variation based on several factors that may reduce or slow recognition of this benefit, including:

- Scale, breadth, and user- or business-focus of new and migrated use cases for Dataflow.
- Level of ability to provide fast, relevant, and high-quality data services in legacy environments and the degree of transformation with Dataflow.
- Number of business and customer end users and the frequency and complexity of their data needs and support requests.
- Ability to measure and attribute metric impact.

Results. To account for these risks, Forrester adjusted this benefit downward by 15%, yielding a three-year, risk-adjusted total PV of $263,000.

Business Operations Savings



Business Operations Savings

Ref.	Metric	Source	Year 1	Year 2	Year 3
E1	Annual data requests avoided or eliminated with services powered by Dataflow for business end users	Composite	500	1,000	1,500
E2	Hours saved by business user per request with immediate data access	Interview data	4	4	4
E3	Productivity recapture rate for salaried employees	Forrester	50%	50%	50%
E4	Fully burdened hourly salary for general business end users	Payscale	$39	$39	$39
E5	Data efficiencies for end users	E1*E2*E3*E4	$39,000	$78,000	$117,000
E6	Annual customer or internal support interactions due to inadequate, incorrect, or too-slow data	Composite	1,000	1,000	1,000
E7	Reduction in support calls from performance, speed, and new data-driven CX enhancements and use cases with Dataflow	Interview data	30%	40%	50%
E8	Avoided support calls attributable to Dataflow	E6*E7	300	400	500
E9	Average support labor hours required to address an issue	Interview data	4	4	4
E10	Fully burdened hourly pay for support personnel	Payscale	$31	$31	$31
E11	Customer support savings	E8*E9*E10	$37,200	$49,600	$62,000
Et	Business operations savings	E5+E11	$76,200	$127,600	$179,000
	Risk adjustment	↓15%			
Etr	Business operations savings (risk-adjusted)		$64,770	$108,460	$152,150
	Three-year total: $325,380			**Three-year present value: $262,831**	

UNQUANTIFIED BENEFITS

Additional benefits that customers experienced but were not able to quantify include:

- **Improved reliability, enhanced performance, and reduced latency.** Speed is essential: If your business reacts with latency, you can't make proactive business decisions. Dataflow batch and streaming operations scaled to meet demand with the needed resources, ensuring performance with low latency and reliability with built-in fault tolerance. Dataflow Shuffle enabled even faster processing of data. A media company's senior software engineer shared, "Reliability is better in all ways with Dataflow — system reliability, data reliability, and avoiding issues while running jobs."
- **Accelerated time-to-market.** Faster development of new data jobs and pipelines combined with the ability to automate data requests and stream data in real time enabled companies to discover insights and act faster than before, accelerating time-to-market with business initiatives. The CPG company's lead technical architect shared: "[Dataflow] accelerated the time-to-market for value-added insights. … It removed barriers to get to data and encouraged people to try out advanced concepts to quickly yield value."
- **Improved accessibility and usage of data, driving more data-backed decision-making.** Dataflow enabled organizations to use data more quickly and easily with better quality, helping streamline decision-making processes and improve decision outcomes. The senior software engineer for a media company shared: "The ecosystem with Dataflow and BigQuery has made data more accessible and easily usable. People can query data ad hoc or use dashboards. We've seen a much larger amount of data-driven decisions where people analyze data and operate off the decisions. Dataflow is powering a lot of those analytics."
- **Enhanced EX, improved employee skill value, and increased ability to attract, hire, and retain employees.** Dataflow provided teams with language flexibility, performance, availability, and ease of use while freeing them from frustrating manual or repetitive tasks to build and manage pipelines or administer infrastructure. Employees therefore enjoyed a more balanced, enjoyable, and empowering workplace environment where they could work on more interesting work that drove more value. Skill sets applicable to Dataflow were more available and useful, helping firms more easily find employees and helping others learn knowledge that is more widely applicable to their careers. Teams also valued the opportunity to work on and contribute to open source. Ultimately, organizations could better attract, hire, and retain employees while also reducing the cost and time needed for hiring and training.
  - A financial services company's employees experienced a more enjoyable and exciting workplace with less stress, dedicating substantially more time to meaningful work while also learning more marketable skills. The VP of technology shared that they expect Dataflow to benefit employee retention and culture while reducing hiring and training costs: "There is an immediate benefit for our engineers. The work is less stressful than supporting on-prem. It's more enjoyable, and they are excited because they're learning new, marketable skills."
  - Another financial services company was able to reduce technical knowledge demands to connect data and produce insights, enabling it to reallocate workers to higher-value work and improve their employee satisfaction.

  - A media company improved user experience for data engineers and developers with Dataflow, while also enabling them to give back to the open source community as Apache Beam contributors.

- **Improved experience and outcomes for customers.** As discussed in Benefit A "Business Growth" and Benefit E "Business Operations Savings," Dataflow improved CX and enabled release of new and improved customer-facing offerings. In addition to growing revenue and reducing support costs, these same factors with Dataflow helped customers achieve better business outcomes. Customers themselves could make better data-backed decisions with easier-to-access, more accurate, and real-time data to benefit their own businesses. For example, a financial services company helped its customers improve their own financial health as the director of data projects shared: "Dataflow helps us identify risk and make decisions to help vulnerable customers quickly, so they avoid bad outcomes like collections, fines, or losses. It's about doing the right thing for our customers."
- **Provided pricing transparency, predictability, and scalability/elasticity.** Dataflow costs are clearly laid out in public documentation with all the nuanced drivers that may affect costs. Organizations can therefore discretely track the cost of their pipelines to find opportunities to control cost and better predict future cash flows. Compared with on-premises solutions, this was a huge step forward. A media company's senior software engineer explained: "We used to spend a lot buying servers and maintaining data centers — it was hard to calculate per-unit cost as a result. Now it's obvious with Google Cloud."
- **Bettered the ability to track, prove, and ensure multiregion compliance.** Dataflow provided greater visibility to pipelines with discrete control regarding when and how data was processed. Customers were able to better enforce governance while monitoring and reporting on compliance. This was especially important for global companies as they worked to consolidate solutions across regions and needed to carefully protect and control data access and protection for compliance, privacy, and security.
- **Reduced risk of lock-in with open source foundation.** Teams could use a language of their choice to get data in, minimizing barriers to entry while allowing for repurposable skills. Jobs were architected in Apache Beam, allowing portability to other runners. Portability of code and skills together reduced business risk, helping ensure that today's investment would remain valuable tomorrow with minimal technical debt. A media company's senior software engineer explained: "There's a big community around Dataflow. It was well-supported, a lot of people knew it, you could get questions answered, et cetera. And in theory you could take your jobs and run them elsewhere without completely rewriting them."

FLEXIBILITY

The value of flexibility is unique to each customer. There are multiple scenarios in which a customer might implement Dataflow and later realize additional uses and business opportunities, including:

- **Tackle new business opportunities not possible before.** Decision-makers looked to develop data use cases for new business opportunities by leveraging their new Dataflow capabilities for streaming, advanced enrichment, and ability to scale beyond legacy overhead.
- **Empower teams with creativity and ease, driving innovation.** The CPG company can now test many more ideas for prescriptive and predictive insights without risk of excess wasted cost and labor. The company's lead technical architect shared: "Dataflow helps us grow by enabling calculated risk-taking with the productivity gained by developers, data

scientists, and modelers. It removes the barriers we faced to getting data and trying things out, so we can now encourage testing out advanced concepts that could quickly yield value if proven successful."

- **Increase usage of data throughout the organization.** The media company's senior software engineer shared: "Dataflow and the Google Cloud ecosystem has made data more accessible and easily usable. We are making more data-driven decisions where our teams build dashboards, analyze data, and operate off the findings. Dataflow powers a good chunk of it."
- **Expand the use of data with reusability to tap new and previously undiscovered opportunities.** At a financial services firm, teams are working to use low-latency data with markets in real time to reduce risk, improve outcomes, and optimize productivity for internal teams and customers.
- **Enable new AI and ML use cases.** Decision-makers looked to use Dataflow in tandem with other GCP and open source services to develop and enhance AI/ML capabilities. A media company's senior software engineer shared, "One of the big projects we're working on is enabling our broader machine learning efforts to use Dataflow with BigQuery to power models."
- **Achieve further savings by using Dataflow in tandem with other technology investments.** Decision-makers hope to continue to reduce costs and improve capabilities by integrating Dataflow with additional Google Cloud and open source technologies, such as storage, BigQuery, PubSub, and TensorFlow.
- **Expand savings to other business units.** Decision-makers were actively working to extend the use of Dataflow to other teams; data is crucial to most components of any organization, and these interviewees predict significant further value could be achieved by applying the benefits of Dataflow to other business units.

Flexibility would also be quantified when evaluated as part of a specific project (described in more detail in Appendix A).

> **"The biggest reason we went with Dataflow was the managed nature of it. We were really interested in not having to worry about the operational burden of managing the instances, doing upgrades, or whatever. Google handles most of that for you with Dataflow. The ability to scale also really met our requirements, if not more. The open source nature of Dataflow was also very attractive."**
>
> *Senior software engineer, media*

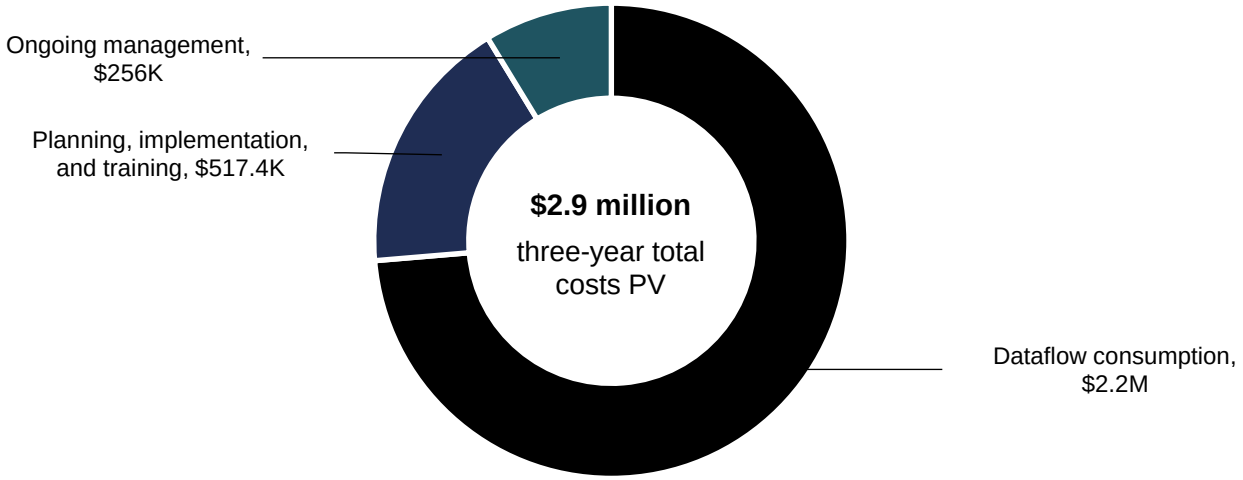
Analysis Of Costs

Quantified cost data as applied to the composite

Total Costs

Ref.	Cost	Initial	Year 1	Year 2	Year 3	Total	Present Value
Ftr	Dataflow consumption	$0	$561,872	$848,383	$1,266,514	$2,676,769	$2,163,487
Gtr	Planning, implementation, and training	$517,440	$0	$0	$0	$517,440	$517,440
Htr	Ongoing management	$0	$102,960	$102,960	$102,960	$308,880	$256,046
	Total costs (risk-adjusted)	$517,440	$664,832	$951,343	$1,369,474	$3,503,089	$2,936,973

Three-Year, Risk-Adjusted Total Costs



“You definitely need to invest time in making Dataflow work for you, but there is significantly less work than in the past, and you get a much better outcome in the end.”

— Senior software engineer, media

DATAFLOW CONSUMPTION

Google Cloud Dataflow is billed based on the amount of data processed and specific demands for how, where, and when data needs be processed. Cost scales based directly on usage, helping stay in line with value — as opposed to per-user subscriptions or especially ongoing on-premises overhead where cost does not map directly to value. Customers valued the ability to scale down to nearly zero with the potential to scale up massively when needed, avoiding idle or wasteful costs. However, leaders must closely watch data pipelines to ensure that unanticipated usage spikes do not catch them off-guard with excess costs.

Modeling and assumptions. Forrester modeled the cost for the composite organization assuming:

- The composite initially processes 2 PiB of data with Dataflow per month. Jobs are split evenly between batch and streaming; however, the batch jobs process 4x more data than streaming and therefore account for 75% of data processed
- Monthly data processed increases to 4.5 PiB by Year 3, with adoption of streaming outpacing batch to account for 50% of data by Year 3.
- Costs are calculated with the Google Cloud cost calculator based on average consumption volume and costs shared by interviewees.[27]

Risks. The expected investment is subject to risks and variation based on several factors that may increase costs or extend deployment, including:

- Costs (and benefits) of additional services used by organizations are excluded to isolate the ROI of Dataflow. Decision-makers should also consider the cost-benefit for other services such as GCP, cloud storage, BigQuery, or Pub/Sub.
- Service usage costs are affected by region, time of day, streaming versus batch, use of Flex, use of Shuffle, total volume of data processed, seconds of data processing used, number of machines, resources per machine, and beyond.
- Unique data needs and usage, such as regulatory demands, region isolation, workload duplication, and egress costs, may significantly alter costs.

Results. To account for these risks, Forrester adjusted this cost upward by 5%, yielding a three-year, risk-adjusted total PV of $2.2 million.

Dataflow Consumption

Ref.	Metric	Source	Year 1	Year 2	Year 3
F1	Total data processed with Dataflow per month, in gibibytes (GiB)	Composite	2,000,000	3,000,000	4,500,000
F2	Percentage of data processed for batch jobs	Interview data	75%	66.67%	50%
F3	Monthly data processed for batch jobs (GiB) (rounded)	F1*F2	1,500,000	2,000,000	2,250,000
F4	Estimated monthly cost for batch jobs	List pricing and interview data	$31,544	$42,059	$47,316
F5	Monthly data processed for streaming jobs (GiB)	F1-F3	500,000	1,000,000	2,250,000
F6	Estimated monthly cost for streaming jobs	List pricing and interview data	$13,049	$25,273	$53,201
Ft	Dataflow consumption	(F4+F6)*12	$535,116	$807,984	$1,206,204
	Risk adjustment	↑5%			
Ftr	Dataflow consumption (risk-adjusted)		$561,872	$848,383	$1,266,514
	Three-year total: $2,676,769			**Three-year present value: $2,163,487**	

PLANNING, IMPLEMENTATION, AND TRAINING

Decision-makers typically implemented Dataflow during or adjacent to cloud migrations, with Dataflow as one key piece of the puzzle. Implementation effort included assessing current architecture, evaluating solutions, planning the new architecture, training on Dataflow, building integrations, constructing workflows, and ultimately testing and releasing to production. Although a single Dataflow workflow could be built and released in two two-week sprints initially, and in less than 20 hours after building out standard architecture and templates, the full transformation including the Dataflow deployment subcomponent as part of cloud migrations typically required six to 24 months.

Modeling and assumptions. Forrester modeled the cost for the composite organization assuming:

- The composite adopts Dataflow as part of a one-year cloud migration of data storage, processing, and analysis from its regional on-premises environments to Google Cloud Platform.
- The composite migrates all its legacy on-premises data pipelines to Dataflow. Half of workflows are migrated as batch jobs while the other half are rearchitected as streaming.
- The migration takes one year, including six months for planning and proof of concept and six months for development, testing, and deployment.
- The Dataflow portion of the migration is led by eight data engineers who dedicate 25% of their time during planning and 50% of their time during deployment to the rollout of Dataflow specifically.
- Another 16 data engineers and developers receive two weeks of training for how to properly use Dataflow and maximize its value.

Risks. The expected investment is subject to risks and variation based on several factors that may increase costs or extend deployment, including:

- Deployment size, legacy technology landscape, target technology landscape, and the level of change needed to deploy Dataflow, especially alongside any cloud migration or adoption efforts.
- Unique organizational requirements, processes, or technology complexities that can limit or lengthen implementation, such as regional regulatory demands, specific integrations, or high data access and protection requirements.
- Size, expertise, and labor cost of internal or professional services teams for deployment.

Results. To account for these risks, Forrester adjusted this cost upward by 10%, yielding a three-year, risk-adjusted total PV of $517,000.

Planning, Implementation, And Training

Ref.	Metric	Source	Initial
G1	Planning labor hours	8 FTEs * 6 months at 25% utilization	2,080
G2	Deployment labor (hours)	8 FTEs * 6 months at 50% utilization	4,160
G3	Training labor (hours)	16 FTEs * 80	1,600
G4	Fully burdened hourly salary for data engineers	B5	$60
Gt	Planning, implementation, and training	(G1+G2+G3)*G4	$470,400
	Risk adjustment	↑10%	
Gtr	Planning, implementation, and training (risk-adjusted)		$517,440
	Three-year total: $517,440	**Three-year present value: $517,440**	

ONGOING MANAGEMENT

The Dataflow service requires minimal administration because it is a managed service. This eliminates the need for infrastructure deployments, management, maintenance, and upgrades along with software support and upgrades especially related to scale.

Nonetheless, a (reduced) level of ongoing effort was still required by organizations to continually: 1) improve and tune pipelines; 2) construct workflows and jobs to meet new user and business needs; and 3) manage technical challenges related to custom code and integrations.

Investing in improving performance and outcomes of Dataflow jobs along with expanding the use cases for it helped to ensure adoption and maximize realization of benefits, driving further business value over time as modeled in Benefit A and Benefit E.

Modeling and assumptions. Forrester modeled the cost for the composite organization assuming that three data engineers dedicate 25% of their time to improving and tuning pipelines, constructing new workflows and jobs, and managing technical challenges that may arise. This level of effort is representative of interviewees' organizations.

Risks. The expected investment is subject to risks and variation based on several factors that may increase costs or extend deployment, including:

- Size and complexity of the Dataflow deployment.
- Amount and complexity of additional workflows and existing workflow updates after launch.
- Unique business needs such as regional compliance, redundancy, and governance.
- Level of internal expertise regarding Dataflow and related languages and average salaries.

Results. To account for these risks, Forrester adjusted this cost upward by 10%, yielding a three-year, risk-adjusted total PV of $256,000.

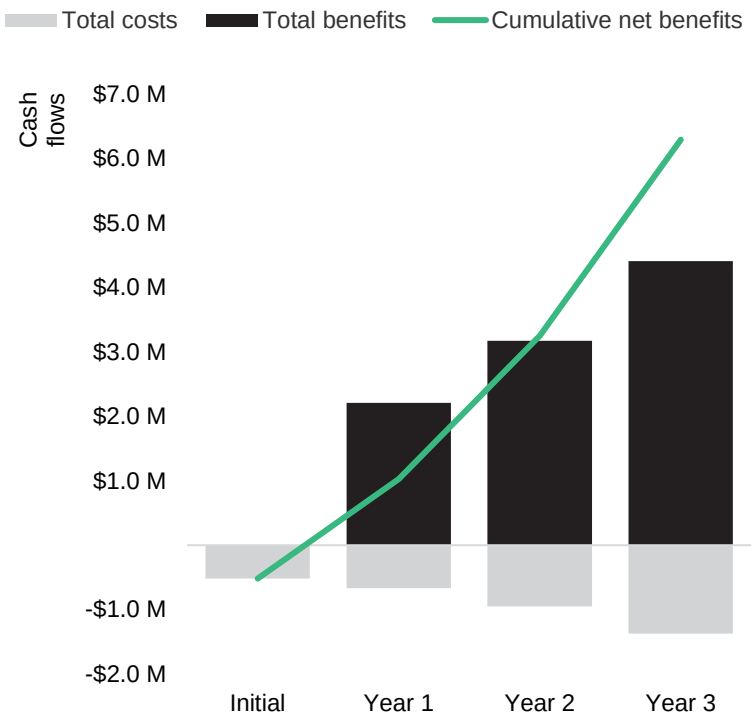
Ongoing Management

Ref.	Metric	Source	Year 1	Year 2	Year 3
H1	Ongoing labor for tuning, management, and new pipeline development with Dataflow (hours)	3 FTEs at 25% utilization	1,560	1,560	1,560
H2	Fully burdened hourly salary for data engineers	B5	$60	$60	$60
Ht	Ongoing management	H1*H2	$93,600	$93,600	$93,600
	Risk adjustment	↑10%			
Htr	Ongoing management (risk-adjusted)		$102,960	$102,960	$102,960
	Three-year total: $308,880			**Three-year present value: $256,046**	

Financial Summary

CONSOLIDATED THREE-YEAR RISK-ADJUSTED METRICS

Cash Flow Chart (Risk-Adjusted)



The financial results calculated in the Benefits and Costs sections can be used to determine the ROI, NPV, and payback period for the composite organization's investment. Forrester assumes a yearly discount rate of 10% for this analysis.

These risk-adjusted ROI, NPV, and payback period values are determined by applying risk-adjustment factors to the unadjusted results in each Benefit and Cost section.

Cash Flow Analysis (Risk-Adjusted Estimates)

	Initial	Year 1	Year 2	Year 3	Total	Present Value
Total costs	($517,440)	($664,832)	($951,343)	($1,369,474)	($3,503,089)	($2,936,973)
Total benefits	$0	$2,211,477	$3,173,385	$4,411,404	$9,796,265	$7,947,419
Net benefits	($517,440)	$1,546,645	$2,222,042	$3,041,929	$6,293,176	$5,010,446
ROI						171%
Payback (months)						<6

Appendix A: Total Economic Impact

Total Economic Impact is a methodology developed by Forrester Research that enhances a company's technology decision-making processes and assists vendors in communicating the value proposition of their products and services to clients. The TEI methodology helps companies demonstrate, justify, and realize the tangible value of IT initiatives to both senior management and other key business stakeholders.

TOTAL ECONOMIC IMPACT APPROACH

Benefits represent the value delivered to the business by the product. The TEI methodology places equal weight on the measure of benefits and the measure of costs, allowing for a full examination of the effect of the technology on the entire organization.

Costs consider all expenses necessary to deliver the proposed value, or benefits, of the product. The cost category within TEI captures incremental costs over the existing environment for ongoing costs associated with the solution.

Flexibility represents the strategic value that can be obtained for some future additional investment building on top of the initial investment already made. Having the ability to capture that benefit has a PV that can be estimated.

Risks measure the uncertainty of benefit and cost estimates given: 1) the likelihood that estimates will meet original projections and 2) the likelihood that estimates will be tracked over time. TEI risk factors are based on "triangular distribution."

The initial investment column contains costs incurred at "time 0" or at the beginning of Year 1 that are not discounted. All other cash flows are discounted using the discount rate at the end of the year. PV Sources are calculated for each total cost and benefit estimate. NPV Sources in the summary tables are the sum of the initial investment and the discounted cash flows in each year. Sums and present value Sources of the Total Benefits, Total Costs, and Cash Flow tables may not exactly add up, as some rounding may occur.

PRESENT VALUE (PV)

The present or current value of (discounted) cost and benefit estimates given at an interest rate (the discount rate). The PV of costs and benefits feed into the total NPV of cash flows.

NET PRESENT VALUE (NPV)

The present or current value of (discounted) future net cash flows given an interest rate (the discount rate). A positive project NPV normally indicates that the investment should be made, unless other projects have higher NPVs.

RETURN ON INVESTMENT (ROI)

A project's expected return in percentage terms. ROI is calculated by dividing net benefits (benefits less costs) by costs.

DISCOUNT RATE

The interest rate used in cash flow analysis to take into account the time value of money. Organizations typically use discount rates between 8% and 16%.

PAYBACK PERIOD

The breakeven point for an investment. This is the point in time at which net benefits (benefits minus costs) equal initial investment or cost.

Appendix B: Supplemental Material

“Align Strategy And Technology To Deliver Next-Gen Personalization,” Forrester Research, Inc., September 2, 2020.

“Anticipatory Experiences: The Challenges,” Forrester Research, Inc., January 20, 2021.

“Architecting Data For The Intelligent Edge,” Forrester Research, Inc., July 7, 2021.

“Consumers Don’t Want Self-Service Digital Experiences; They Want Service,” Forrester Research, Inc., August 30, 2021.

“Create A Modern Software Strategy With AI-Driven Platforms And Their Ecosystems,” Forrester Research, Inc., July 23, 2020.

“Data Mesh Makes Connected Intelligence Experiences A Reality,” Forrester Research, Inc., July 7, 2021.

“Determine Your Data’s Worth: Data Plus Use Equals Value,” Forrester Research, Inc., August 2, 2021.

“Enterprise Data Fabric Enables DataOps,” Forrester Research, Inc., August 2, 2021.

“Event-Driven Architecture And Design: Five Big Mistakes And Five Best Practices,” Forrester Research, Inc., November 10, 2020.

“How To Choose The Right Tech For Your CX Needs,” Forrester Research, Inc., June 29, 2021.

“Move Your Big Data Into The Public Cloud,” Forrester Research, Inc., June 15, 2017.

“New Technology Projection: The Total Economic Impact Of Anthos,” a commissioned study conducted by Forrester Consulting on behalf of Google, November 2019.

“Now Tech: Streaming Data Platforms, Q2 2021,” Forrester Research, Inc., April 7, 2021.

“Research Overview: Data, Analytics, And Insights,” Forrester Research, Inc., January 27, 2020.

“The Forrester Tech Tide™: Enterprise Business Insights And Analytics, Q1 2021,” Forrester Research, Inc., February 17, 2021.

“The Forrester Wave™: Cloud Data Warehouse, Q1 2021,” Forrester Research, Inc., March 24, 2021.

“The Forrester Wave™: Data Management For Analytics, Q1 2020,” Forrester Research, Inc., February 12, 2020.

“The Forrester Wave™: Streaming Analytics, Q2 2021,” Forrester Research, Inc., June 7, 2021.

“The Forrester Wave™: Streaming Analytics, Q3 2019,” Forrester Research, Inc., September 23, 2019.

“The Future of Analytics,” a commissioned study conducted by Forrester Consulting on behalf of Google, July 2020.

“The Insights Beat: Time To Reeducate Your Organization On Data,” Forrester Research, Inc., September 12, 2021.

“The Total Economic Impact Of Google Kubernetes Engine,” a commissioned study conducted by Forrester Consulting on behalf of Google, June 2021.

“Top Trends Shaping Fraud Management And Anti-Money Laundering,” Forrester Research, Inc., August 6, 2021.

Appendix C: Endnotes

[1] Total Economic Impact is a methodology developed by Forrester Research that enhances a company's technology decision-making processes and assists vendors in communicating the value proposition of their products and services to clients. The TEI methodology helps companies demonstrate, justify, and realize the tangible value of IT initiatives to both senior management and other key business stakeholders.

[2] Source: "Determine Your Data's Worth: Data Plus Use Equals Value," Forrester Research, Inc., August 2, 2021.

[3] Source: "The Future of Analytics," a commissioned study conducted by Forrester Consulting on behalf of Google, July 2020.

[4] Source: "Move Your Big Data Into The Public Cloud," Forrester Research, Inc., June 15, 2017.

[5] Source: Ibid.

[6] Source: Ibid.

[7] Source: Srividya Sridharan, "The Insights Beat: Time To Reeducate Your Organization On Data," Forrester Blogs, September 12, 2021 (https://www.forrester.com/blogs/the-insights-beat-time-to-re-educate-your-organization-on-data/).

[8] Source: "The Future of Analytics," a commissioned study conducted by Forrester Consulting on behalf of Google, July 2020.

[9] Source: "The Forrester Wave™: Data Management For Analytics, Q1 2020," Forrester Research, Inc., February 12, 2020.

[10] Source: "The Future of Analytics," a commissioned study conducted by Forrester Consulting on behalf of Google, July 2020.

[11] Source: "Move Your Big Data Into The Public Cloud," Forrester Research, Inc., June 15, 2017.

[12] Source: "The Forrester Tech Tide™: Enterprise Business Insights And Analytics, Q1 2021," Forrester Research, Inc., February 17, 2021.

[13] Source: "The Forrester Wave™: Data Management For Analytics, Q1 2020," Forrester Research, Inc., February 12, 2020.

[14] Source: "Now Tech: Streaming Data Platforms, Q2 2021," Forrester Research, Inc., April 7, 2021.

[15] Source: "Consumers Don't Want Self-Service Digital Experiences; They Want Service," Forrester Research, Inc., August 30, 2021.

[16] Source: "Now Tech: Streaming Data Platforms, Q2 2021," Forrester Research, Inc., April 7, 2021.

[17] Source: Srividya Sridharan, "The Insights Beat: Time To Reeducate Your Organization On Data," Forrester Blogs, September 12, 2021 (https://www.forrester.com/blogs/the-insights-beat-time-to-re-educate-your-organization-on-data/).

[18] Source: "Now Tech: Streaming Data Platforms, Q2 2021," Forrester Research, Inc., April 7, 2021.

[19] Source: Ibid.

[20] Source: Ibid.

[21] Source: Ibid.

[22] Source: Ibid.

[23] Source: “The Forrester Wave™: Streaming Analytics, Q2 2021,” Forrester Research, Inc., June 7, 2021.

[24] Source: Ibid.

[25] Source: Ibid.

[26] Source: Average total market pre-tax unadjusted operating margin is 9.62%. “Margins by Sector (US),” Stern School of Business at New York University, January 2021 (http://pages.stern.nyu.edu/~adamodar/New_Home_Page/datafile/margin.html)

[27] Source: Pricing computed through Google Cloud Pricing calculator, which aligns approximately with expected spend based on customer interviews.

Estimated quote derived with: Los Angeles, N1-standard-16 instance, 730 hours, 8 nodes, 500gibPD. https://cloud.google.com/products/calculator


FORRESTER®