



March 2021

Why an intelligent data cloud is key to digital transformation



Google Cloud

Table of Contents

Executive summary	1
Modern data strategies stuck in ancient data systems	2
From gap to chasm: Why companies are failing to transform data into value	3
Transforming data to intelligence to value	5
An intelligent data cloud	7
What we've learned at Google	9
Data transformation is a journey, not a destination	9
Getting started	10

Executive summary

In today's world of digital disruption, organizations in every field face threats from alternatives with better technology, better business models, better operational value chains and better customer experiences. Digital transformation is not a new phenomenon but more recently we've seen it accelerate across retail, healthcare, financial services, transportation, automotive, media and entertainment, and manufacturing. We're seeing digital innovators disrupt incumbents' business models left and right by innovating at a much faster pace. These digital innovators understand the power of unifying data across their organization to drive real transformation and value creation. They get the impact of resilient, mission critical databases, analytics and machine learning systems that can reliably run the business 24/7, and power innovation. Digital innovators are building their data clouds with a platform that's open, intelligent and trusted, and they are doing it now.

Customers powering their digital transformation with intelligent data clouds:

- 1 The Home Depot (THD) makes [400k+ associates smarter](#) across 50,000+ items stocked at over 2,000 locations by giving them visibility to the things each customer needs. THD also uses Cloud SQL, Spanner, and Bigtable for their operational use cases and AI to help locate goods using their mobile apps for in-store navigation.
- 2 The American Cancer Society used [ML](#) and compute to identify novel patterns in digital pathology images to potentially improve patient outcomes, and [analyze breast cancer images 12X faster](#). If they hadn't taken a machine learning approach, it would have taken them years instead of three months to analyze 1,700 tissue samples—even with a team of dedicated pathologists.
- 3 UPS saves up to [\\$400 million a year](#) by reducing fuel consumption by 10 million gallons per year, using BigQuery, and Spanner, to help them deliver more packages at a lower cost and serve its customers in a smarter, more agile way.
- 4 Online retailer Zulily is using the latest tools in artificial intelligence, machine learning, and cloud computing to [innovate and serve its customers with purpose](#). To give online shoppers a complete view of what a piece of apparel will look like when tried on, Zulily trains machine learning models to inspect product images and discern the information. They build solutions like this with the help of cloud-based technologies such as [AutoML Vision](#).
- 5 ANZ Bank runs bank processing [up to 250x faster](#) than before using BigQuery, Cloud SQL and Spanner.
- 6 [Priceline is responding more quickly to evolving customer needs](#) by making data-driven decisions much faster, with the help of [ML](#), [BigQuery](#) and [Google Kubernetes Engine](#).

But it's not easy to get these kinds of results and siloed, legacy systems tend to be the main culprit, requiring extensive maintenance and management that eats up the critical resources needed to capture value. From listening to our customers, we also hear that even when they do have more modern tools, they don't connect easily so the majority of their time is spent on systems engineering, leaving very little time for actual data analysis.

When enterprises fail to find ways to integrate, manage, and use their data, they are leaving a lot of value on the table—and the data value gap continues to widen as the amount of data increases. Organizations must take steps now to figure out how to close the gap and support value generation if they want to be able to adapt to the inevitable future disruption that is going to continue to define their businesses.

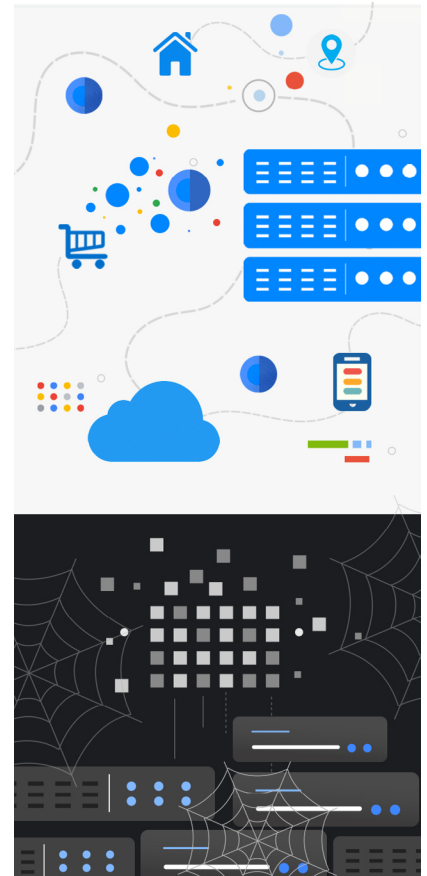
In this whitepaper, we'll explore why your business needs an intelligent data cloud to run your day-to-day operations, why data transformation is the key to unlocking more value for your business, and how Google can help.

Modern data strategies stuck in ancient data systems

Google Cloud customer [The Home Depot](#) (THD) has made a name for itself going big—big stores, big product selection, and above all big customer satisfaction. But over time, THD realized it had a problem and of course, it was big—big data. While their success has largely been data-driven over the years, THD was looking for a way to modernize its approach. They needed to better integrate the complexities in their related businesses, such as tool rental and home services. They also wanted to better empower their data analysis teams and store associates with mobile computing devices, as well as leverage ecommerce and new modern tools like artificial intelligence (AI) to meet customer needs.

Their existing on-premises data warehouse was proving too limited to handle contemporary pressures, overtaxed by the constant

When enterprises fail to find ways to integrate, manage, and use their data, they are leaving a lot of value on the table—and the data value gap continues to widen as the amount of data increases.



demand for analytics and struggling to manage the increasingly complex use cases from their data analysts. This not only drove massive growth of the data warehouse, it also created challenges in managing priorities, performance, and cost.

If THD wanted to add capacity to the environment, it required major planning, architecture, and testing effort. In one case, adding on-premises capacity took six months of planning and a three-day service outage. But the results were short-lived—within a year, capacity was again scarce, impacting performance and ability to execute all the reporting and analytics workloads required. The Home Depot also needed to modernize their operational databases in order to deploy applications faster for their teams and move away from managing resources.

These challenges resulted in no real-time access into sales, product, or shipping metrics which THD needed to optimize the customer experience, product SKUs, and more—which would ultimately help them differentiate in an industry where a seamless customer experience is everything.

Sound familiar? These challenges are by now a common story across the enterprise. Most companies, like THD, are finding that operating legacy technology while trying to deliver a modern data strategy is no longer possible.

From gap to chasm: Why companies are failing to transform data into value

So what's holding enterprises back?

At the same time, the pressure to understand, respond to, and sometimes even predict risks and opportunities against the astronomical amount of data is only growing. Every executive recognizes the massive potential of their data to drive competitive advantage and accelerate digital transformation. Done right, data intelligence can help shape delightful, personalized customer experiences, streamline business operations, better forecast demand, and drive innovative and impactful outcomes. But it requires the ability to put all that data to work and derive insights from it—otherwise, you have all the ingredients but you're cooking without the recipe. It might deliver results, but it will always fall short of the promised meal.

Unfortunately, achieving real-time data insights still remains more of a pipe dream despite the exponential leaps forward in technology over the last few decades. And instead of being rocketed to new innovative heights, many companies instead find themselves staring down a widening gap between the value they have managed to deliver and the potential value they know can be achieved.

Here's why it's so hard for organizations to convert their data into value:



Data silos block businesses from getting insights.

Data silos are pervasive across organizations in every industry. These independent datasets are a consequence of logical, physical, technical, or cultural barriers, which typically lead to fragmented systems that are unable to communicate and share information in real time. For instance, human resources, finance, and other departments may collect overlapping data, but use different systems and tools to store and manage their data, leading to inconsistencies. Data silos prevent enterprises from achieving a consolidated view of data, which makes it impossible to uncover hidden opportunities. Critically, inconsistencies can also lead to mistrust, which hurts collaboration but also keeps people from wanting to use and collaborate with data again.



On-premises infrastructure can't scale fast enough to meet data growth.

Scaling on-premises infrastructure to keep up with growing customer demand and data growth has reached an untenable level. Rigid legacy infrastructures struggle to scale fast enough to keep pace with fluctuations in data requirements. The days of overnight data operations are being replaced by the need for streaming and batch data processing, while also supporting simultaneous processing. And legacy infrastructure just isn't able to keep up. Hitting capacity limits end up slowing users down and tying up database administrators, too.



IT dependency and operational overhead for managing infrastructure is costly.

Like other on-premises systems, databases follow the old-school mode of paying for hardware and licensing costs, as well as the associated ongoing systems engineering. Updating and extending storage usually requires modifications to both hardware and software, forcing teams to waste time that would be better spent elsewhere. Furthermore, legacy BI tools are reliant on someone manually creating, running and updating reports that are frequently outdated by the time they reach your inbox.

As a result, many companies feel they are always running to keep up with their data. Instead of planning ahead, businesses are left reacting to whatever just happened. This becomes particularly troubling when unforeseen factors or disruptions occur. If COVID-19 has taught the world anything, it's that nothing is certain and the best way to prepare is to plan for change.



AI (and managing data) is complicated.

AI-powered predictive analytics can be intimidating and time-consuming. But the hardest part of AI and machine learning (ML) is data management. For instance, ML models are only as good as the data used to train them. This is the concept of “garbage in, garbage out” in action—AI doesn’t remove inaccuracies or inconsistencies, so poor data quality will in turn yield poor insights. In addition, machine learning requires collecting and labeling a massive amount of data. In some cases, data is a free byproduct of a system or product, but for many others, getting the data you need to train data science models is incredibly expensive and challenging to collect. Many organizations lack the skills necessary to manage datasets and aren’t sure where to start investing when collecting data.

Transforming data to intelligence to value



By now, it’s clear that data alone is no longer the primary competitive differentiator—it’s what organizations are able to do with that data that will matter. But according to [Gartner](#), approaching information and data as an asset is still in the early phases, which can make it a competitive advantage for organizations that invest in data transformation. Companies will require faster, forward-looking decisions to compete and data and analytics capabilities will need to become core competencies for delivering enterprise value.

There are many factors to consider when it comes to achieving data transformation. Advanced technology surrounding data means there are now better options that provide more access and easier manageability compared to the legacy solutions available in the past. Still, managing data at scale is hard even with today’s better process automation and tools to make sense of gathered data. Sometimes, more access translates to more risk, bringing a new set of challenges around the security, quality, and interpretability of data.

Companies rely on their databases and data warehouses for their most valuable information to collect, process, and analyze virtually every kind of data imaginable. They need to create data transformation pipelines from a combination of databases, data warehouses, and data lakes that can deliver real-time, actionable insights—and deliver them to the people who need them with a click of a mouse. They want to invest in AI and machine learning to help them with the heavy lifting. On top of that, businesses need to be able to run daily operations securely and with ease.

So, how do you build all of that? **The answer lies in an intelligent data cloud.**

A data cloud helps companies harness the value of their massive data volumes by providing a modern, cloud-based data infrastructure that helps them achieve a single, universal view of all their data. With on-demand compute, database storage, content delivery, and other advanced analytics capabilities, such as streaming analytics, machine learning and data science, a data cloud empowers everyone in an organization to access the data they need with confidence.

We've learned a number of lessons about what companies need to optimize data in real time for value, both from our own experiences at Google Cloud and from helping our customers solve hard problems.

An intelligent data cloud should have the following key characteristics:

- 1 **Agility.** Data and data analytics should support moving faster, not slow teams down. This means building a data foundation that offers easy access to instant compute and storage resources that will help reduce the burden on IT resources. Rather than adding more tasks, a data cloud should provide an agile, flexible data architecture that provides self-service provisioning and the ability to scale up and down as needed.
- 2 **Discoverable.** It's essential to make data discoverable and accessible so that various groups of users can properly and consistently interpret and act on data. Instead of needing to know where data is located, dealing with different formats, and having access to the right tools for each source, a modern data cloud should unify structured and unstructured data, regardless of source, to reduce complexity and make it easier to discover the data you need to make better, faster decisions.
- 3 **Intelligent.** Applying intelligent capabilities, such as AI/ML, and self-service analytics is a top priority for companies looking to save time, effort, and open up the doors to innovation. A modern data cloud should come with capabilities to automate processes and provide advanced analytics techniques such as augmented analytics and use of NLP (natural language processing), analytical applications, and data science to enable companies to spot trends and offer insights that improve business decisions.
- 4 **Open.** Today, organizations need the freedom to build what they want, where they want without the fear of vendor lock-in or unnecessary technical debt holding them back. It's important that organizations choose more open solutions, which ensure interoperability and portability for enterprise data, enabling them to adapt and evolve data management and data analytics strategies quickly.

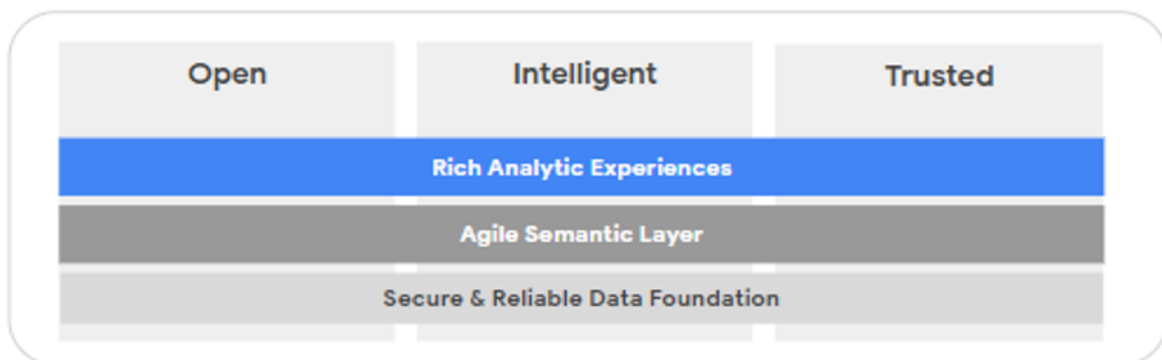
- 5 **Trusted.** One of the biggest goals of streamlined data collection and use is that it's trusted. That means building trust in data: trust that it's up-to-date, accurate, and always protected. Data comes from a huge variety of sources, but it all requires the same baseline level of protection and governance. A modern data cloud should be secure by default and leverage security capabilities that ensure compliance, redundancy, and reliability.

An intelligent data cloud

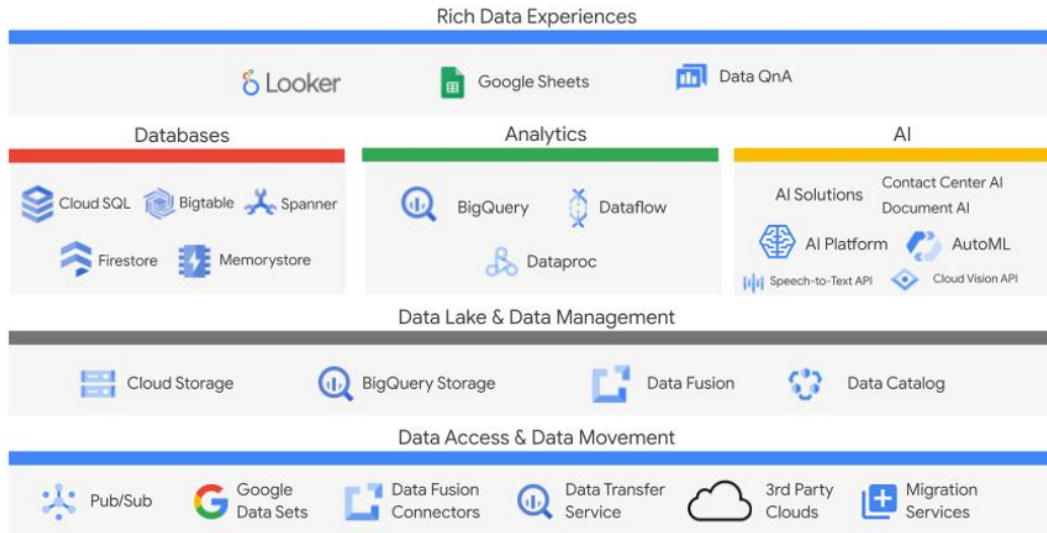
Unifying your data across databases, analytics and AI/ML systems within a single platform offers the scale, speed and security that modern businesses need. It provides employees across the business with the right information at the right time, to make more informed, smarter decisions.

For instance, if we think back to the main challenges facing The Home Depot, migrating over to Google Cloud had a huge impact. By moving from their on-premises data warehouse to BigQuery, query performance dropped from hours and days to seconds and minutes. They have also built one of the industry's most efficient replenishment systems to monitor, analyze, and act on app performance data across all of their stores and warehouses in real time. Today, THD is able to keep over 2,000 locations stocked with over 50,000 products with help from Google Cloud. They have also enabled 400,000 associates to work smarter by providing visibility to the things each customer needs, such as an item location within any local store. In short, The Home Depot unified data across their organization and made their data cloud the heart of their business.

Let's take a closer look at how it all works. Below are two diagrams. The first shows the criteria for your data cloud at a high level. The second includes some of the core technologies that make up Google Cloud's data platform and where they fit in this model.



Google's Data Platform Technologies



From the point when data is created, from any source at any speed, Google Cloud database solutions support resilient, mission critical systems that reliably run your business 24/7. [Spanner](#) provides unlimited scale, global consistency across regions and high availability up to 99.999%. [Firestore](#) continues to see rapid adoption with over 2M databases being created to power mobile, web, and IoT applications across customer environments. [Cloud SQL](#)—one of the fastest growing services at Google Cloud—offers familiar open source MySQL and PostgreSQL standard connection drivers. This lets you work with your preferred tools, and keeps you up-to-date with the latest community enhancements. All these innovations mean that you're always up and running to deliver the most transformative applications, while maintaining control of your data.

Once the data is generated, it needs to be aggregated and analyzed, and our smart analytics solutions allow companies to democratize access to all of their business data using the combination of [BigQuery](#) to run their analytics at petabyte scale, [Dataproc](#) for secure and scalable open source data science, and with self-service visualizations using [Looker](#), providing accelerated time-to-insights. Our unified platform makes it easy for our customers to get the most from their structured or unstructured data, regardless of where your data is stored.

With the leading [AI solutions](#) portfolio, Google helps businesses commit to business outcomes: saving calls, improving customer experiences, preventing fraud and increasing manufacturing efficiency—all in ways that make us a strategic partner on innovation, not a technology vendor.



And with [AI Platform](#), the component technologies in our solutions like understanding the content of images, or turning unstructured PDFs into a structured data table—are available for you to build your own solutions too.

And because our data cloud solutions are more open than any other cloud provider, our customers have the choice to run anywhere, securely, with the most robust processing and analytics engines.

What we've learned at Google

Our most valuable learnings come from our customers. And most customers are on a journey: from living in the weeds of operating infrastructure themselves to a hybrid or multicloud model, where some workloads are in the cloud and some remain on-prem, to fully managed cloud services.

We're on this journey, too. Looking back, as our analytics needs grew, we built Dremel, a giant, serverless query engine. Then, we discovered our customers wanted and needed this too, so we built BigQuery, the external version of Dremel. Similarly, we needed a framework to help with the transformation of data at scale, and that's how Dataflow and Apache Beam were born. On the database side, we built Spanner and made it available to customers for workloads that need strongly consistent reads and automatic, multi-site replication and failover (think global transactions); Bigtable is our NoSQL implementation for Search, Analytics, Maps and Gmail. And in AI we released AutoML, which automates the design of machine learning models at scale, making the process of building ML models much more accessible.

We have continued designing Google Cloud to remove the traditional constraints of scale, performance, and cost by tapping into the same data platform technologies that power Google's products used by billions of people around the world. From ingestion to processing all the way to business intelligence (BI) and AI, Google Cloud offers intelligent, data-driven solutions that drive business impact.

Data transformation is a journey, not a destination

Data-driven applications and insights will continue to be the dominant focus, helping to drive better decisions, increasing engagement, and powering nearly every process across the enterprise. But we're no longer working towards a defined end goal—the transformation is the goal.

If we've learned anything in the past year, it's that the future is hard—in some cases, impossible—to predict. Investments and imperatives are no longer things that are static or gradually shifting. They can change fast and without warning, and it's the strategic gathering, managing, and developing of information that will become the new constant.

Even once companies reach maturity, there will always be a need to innovate, adapt, and respond to continuous change. Data-driven transformation powered by Google Cloud will help you plan for change, no matter what that change brings.

Getting started

To learn more about how leading companies are powering innovation with our data solutions, come to our [Data Cloud Summit](#) on May 26, 2021. There will be sessions, demos, and live Q&As where you can discover how data can help you make smarter business decisions and solve your organization's most complex challenges.



Google Cloud