The future of data warehousing

July 2020

Google Cloud

Table of contents

Editor's note
The future of data warehousing

While data has influenced major business decisions throughout the ages, the term "data warehouse" isn't even half a century old. But it's an important concept; by centralizing data from many disparate sources, analysts can make more informed decisions. Traditionally, this has meant collecting data in on-premises infrastructure. But as data volumes have grown, these legacy systems have not been able to keep up with changing demands: They are difficult to scale, offer no support for machine learning and artificial intelligence initiatives, and are simply too expensive to maintain.

With increased advances in technology, the introduction of more cost-effective compute and storage options have changed the data warehouse landscape. Now more than ever, businesses need an agile and efficient analytics infrastructure that allows them to derive insights at a fraction of the cost of legacy systems. Modern data warehouses can give organizations the competitive edge, offering customers fast and easy access to business intelligence while seamlessly scaling to meet increased demands.

In this ebook, based on the keynote presentation featuring Forrester Analyst Michele Goetz at the 2019 Future of Your Data Warehouse digital conference, we'll look at the evolution of data warehousing and how a cloud infrastructure can offer immense benefits for the future.

Chapter 1
A new world of data

Successful businesses have used data to make informed decisions, no matter the era. It was true for Neolithic farmers when they planned the next year's harvest. It was true for 17th-century navigators whose survival depended on having decent maps. And it's true for today's most sophisticated researchers as they seek answers to problems like disease and climate change.

What's different now is that the scale of the world's data has become unfathomably large, with IDC predicting that the global datasphere will increase fivefold between 2018 and 2025[1]. That means it's no longer a question of whether you have the right data—it's how you can best understand the wealth of information that's already at your fingertips, and what you can do with it. The speed at which data is generated and the variety of data has also increased; "internet of things" sensors, web, social media, mobile apps, and more are all contributing to this new landscape.
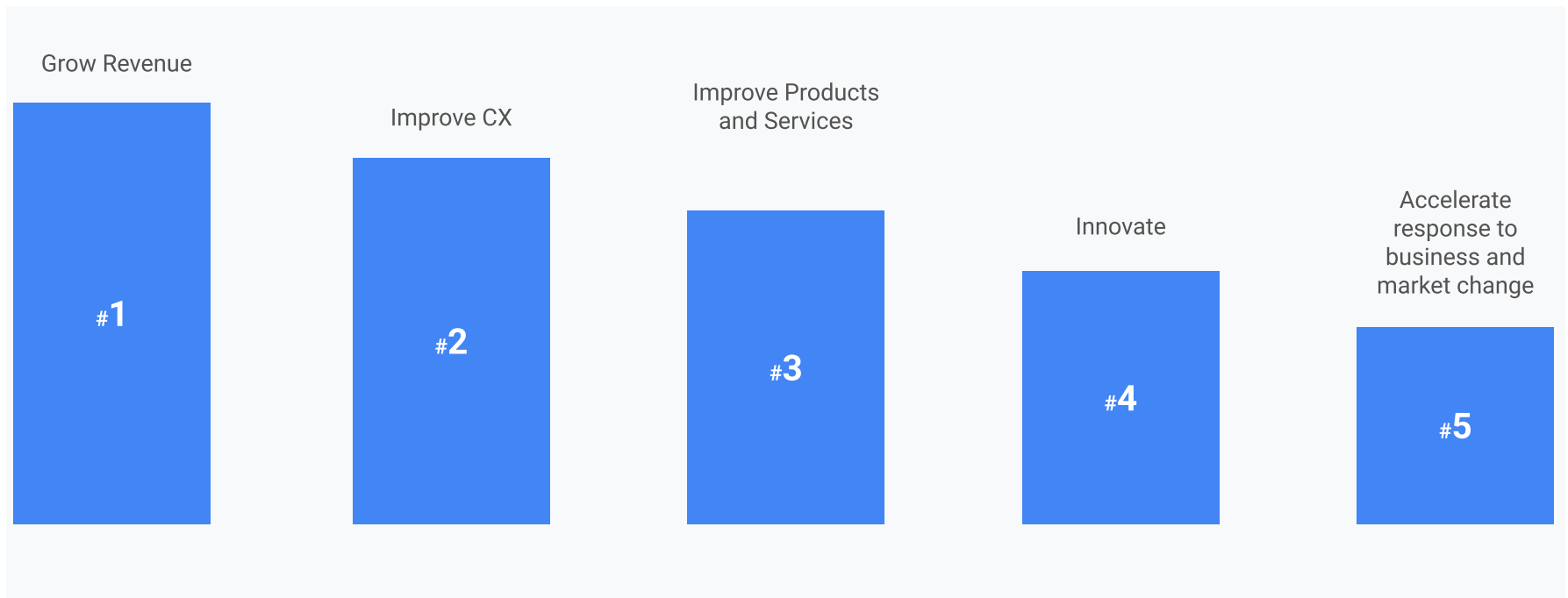
It's no surprise that the ability to turn data into valuable insights can yield better customer experiences and new opportunities, as well as products and services that keep pace with changing demands and market conditions. Growing pressure to unleash data's full potential is transforming how companies operate, and it's also transforming the roles of people within those organizations. Research from Forrester suggests that data and analytics professionals now see themselves as key enablers of critical business goals, such as growing revenue and improving the customer experience.

"We used to constantly hear the term 'big data' and it felt like such a cliche, but the reality is that data is what kills the blind spots of our business. It gives us the ability to really see what's happening in the market, see what's happening with our customers, and see what's happening within our business itself."

Michele Goetz, Principal Analyst, Forrester Research, on the keynote presentation, "The Future of Data Warehousing"

---

1   IDC whitepaper, sponsored by Seagate, Data Age 2025: The Digitization of the World: From Edge to Core, November 2018.

Google Cloud

Top priorities for data and analytics pros in 2019[2]

Grow Revenue
#1

Improve CX
#2

Improve Products
and Services
#3

Innovate
#4

Accelerate
response to
business and
market change
#5

2   Forrester Analytics Global Business Technographics® Data & Analytics Survey, 2019.

Chapter 2
The limitations of traditional data warehousing

Data professionals and business leaders alike recognize that it's not enough to simply collect and store data. To remain competitive in today's landscape, companies need to transform data into insights, quickly share those insights across the entire organization, and activate the resulting collective knowledge before it becomes stale. This takes technology—and, according to Information Age, businesses are increasingly discovering their existing tools aren't up to the task.

For many organizations, legacy data warehouses cannot keep up with changing needs as the amount of data grows and demand for insights ramps up. These typically complex and outdated architectures weren't designed to accommodate exponential data growth and the speed at which today's leading businesses move. As a result, they're underperforming—while the cost of owning, licensing, and operating them creeps up. Furthermore, they don't typically support advanced analytics capabilities and real-time insights, competitive differentiators that today's businesses are embracing to succeed and stay current. This leaves existing data systems and the people who manage them scrambling. And when an organization's data needs can't be met, it often leads to shadow IT as business teams turn to a variety of disparate and incompatible solutions, such as using unapproved software or storing data on personal devices, creating even bigger silos and new security concerns.

This is why data warehousing is becoming much more than a humble operational function kept out of sight (and mind) in a dusty server room. Indeed, it's moving to the strategic heart of the business. In a future that's likely to be heavily shaped by increased

"You have to put data warehousing in context of not only the general use cases where data is going to provide value, but the future of the way that we're going to run our business. We're really starting to transform ourselves in the way that we work."

Michele Goetz, *Principal Analyst, Forrester Research, on the keynote presentation, "The Future of Data Warehousing"*

automation, more collaboration, enhanced adaptability and agility, smarter predictions, faster reaction times, and increased awareness of ecosystems, data warehousing needs to support and even help define the future direction of a business. But none of this is possible without data warehouse modernization.

Chapter 3
Enter the cloud

Between 2017 and 2019, the amount of structured data stored in the cloud jumped from 37% to 45%, and the pace of migration is also accelerating for other types of data, according to Forrester[3]. Organizations are taking advantage of the cloud—not just as a centralized repository for their data but also as a platform for extracting and disseminating insights. Over the next 12 months, 60% of organizations will increase their reliance on data warehousing as a service by more than 5%, and 65% will expand their use of data management services in the cloud by more than 5%, Forrester has found.[4]

Businesses are finding that the solution to these legacy data warehousing challenges comes from the cloud. With cloud technology's enormous flexibility, businesses can quickly and seamlessly scale up or down their storage and compute capacity to accommodate changing dynamics. Cloud data warehouses are fully managed, increasing business agility by removing the need for employees to spend valuable time on setup, maintenance, upgrades, configuration, and capacity planning. Organizations can now quickly consolidate their siloed data and create a single source of truth across teams.

The cloud also offers modern solutions to IT issues in data management. IT teams can enhance data security using built-in access controls and other advanced features, allowing them to easily and securely share insights across the organization.

"We're taking advantage of the cloud to allow us to accelerate collection of information, to not always worry about 'is it valuable, is it not valuable' because we know that anything we can collect about the way that we operate and the way that we engage is going to be important."

Michele Goetz, *Principal Analyst, Forrester Research, on the keynote presentation, "The Future of Data Warehousing"*

---

3   Forrester Analytics Business Global Technographics® Data & Analytics Survey, 2017, 2018, 2019.
4   Forrester Analytics Business Technographics Global Data & Analytics Survey, 2019.

Many cloud solutions support advanced analytics, so it's easy for even nonexperts to start using ML to build models. This allows the nonexperts to focus on predicting the next big thing, while enabling data scientists and analysts to innovate, rather than spend time on moving data for analysis.

As a bonus, using these cloud solutions can actually come with a lower total cost of ownership versus on-prem alternatives. The cloud eliminates the capital expenditure (capex) of buying servers, networking solutions, or storage to build a data center, while also removing the associated operational expenses (opex) of server management, performance tuning, and resource provisioning.

Chapter 4
Reimagining your data warehouse

While traditional, on-premises data warehouses often cannot keep up with the performance and capacity requirements of current realities, neither can many other cloud-based data warehousing options. They may be lifted and shifted right from an on-prem infrastructure and unable to keep up with modern needs. They might not easily scale to accommodate gigabytes to petabytes of streaming data from numerous different sources, which is what's necessary to unlock the type of real-time insights businesses need to make fast decisions and smart predictions.

To get the most out of the cloud, it's important to find the right cloud solution for your data warehouse needs. With so many cloud options, however, it may be hard to know which one would be the best for your business. Let's start with the five questions you need to consider before reimagining your data warehouse.

How efficient is your infrastructure?

Traditional data warehouse platforms can bog down IT teams—and budgets—with routine maintenance and configuration tasks, and they often can't scale on demand to enable organizations to seize new business opportunities. Many managed cloud data warehouses are just as complex as on-prem solutions, with bolted-on, stand-alone components for functions like ML and geographic information systems (GIS). This greatly hampers business agility and flexibility.

To get the most out of your cloud-based data warehouse, you need a solution that is agile and efficient, and allows you to run analytics without breaking the bank.

Can you run advanced analytics at the speed your customers need?

Consider the lag time of moving data from your operational and transactional systems to your data warehouse. Does it take hours? A day or more? The time lost to these operations can often be measured in money and opportunities lost. Faster data analytics for businesses translates to more time spent on product innovation and improved customer experiences.

Modern cloud-based data warehouses support streaming analytics and have ML applications already built in, so there's no need to move data. This enables businesses to pose analytical questions and predict outcomes based on real-time data.

Are data sharing and collaboration easy?

Is your data warehouse a collaborative hub? Traditional data warehouses were not designed for easy collaboration. In a world where collaboration is at the heart of a business's processes, your data warehouse needs to seamlessly support users in accessing and sharing information.

Cloud data warehouses allow IT, business departments, subject-matter experts, and other stakeholders to securely access and share analytical insights within your organization.

Do you have security and governance you can trust?

As your business consumes more data from more places and expands analytics solutions, security and governance becomes even more crucial. Your IT security teams have to keep up with security threats and vulnerabilities in your environment.

To protect your data and reputation, modern data warehouses have built-in security measures, such as encryption by default and robust access controls. These ensure that only authorized people have access to your internal, customer, and partner data.

Can your platform handle changing business needs?

In a global, always-on world, your data warehouse needs to be available at all times. While you may be able to accommodate legacy systems going offline for scheduled maintenance, it's a lot harder to maintain ongoing operations when the unexpected happens and you need to scale up or down rapidly.

Cloud data warehouses are fully managed and offer availability and reliability service-level agreements (SLAs) to ensure that your business is always running smoothly.

Chapter 5
Introducing BigQuery

At Google, we developed our serverless, highly scalable, and cost-effective cloud data warehouse BigQuery to address all of these questions. It brings agility, advanced analytics, data sharing and collaboration, security and governance, and data diversity.

BigQuery is easy to set up and manage and doesn't require a database administrator. Your organization can quickly get up and running in seconds, and start querying gigabytes to petabytes of data with standard SQL. It has powerful security, governance, and reliability controls with a 99.9% uptime SLA and built-in protections like encryption by default, and fine-grained identity and access controls.

But with so many data warehouse options on the market, why should you choose BigQuery?

Accelerated time to value and lower TCO

BigQuery offers both long-term value and the ability to do more with less. A serverless infrastructure means that you no longer need to spend time on systems engineering efforts like performance tuning, resource provisioning, or deployment and configuration. Instead, you can focus on analysis and insights.

BigQuery at a glance

- An enterprise data warehouse

- Fully managed and serverless

- Scales up to petabytes on demand

- Agile and cost-effective to meet your needs

- Encrypted, durable, and highly available

- Real-time insights over streaming and batch data

- Built-in machine learning for predictive analytics

- In-memory analysis service for blazing-fast reporting

A recent three-year total cost of ownership (TCO) study by Enterprise Strategy Group found that companies save 52% when they migrate from an on-premises data warehouse to a cloud environment. It also compared four leading solutions and concluded that BigQuery is the most cost-effective, providing a three-year TCO between 26% and 34% lower than the other cloud data warehouses[5].

Instant insights with real-time analytics

BigQuery supports streaming data from IoT sensors, web, social media, mobile apps, and more to help businesses deliver insights quickly and easily, right inside their data warehouse. BigQuery's powerful and high-performance streaming makes your latest business data immediately available for analysis. You can also build comprehensive batch and streaming data pipelines with Pub/Sub and Dataflow, allowing you to query data in real time and know what's happening right now.

Additionally, BigQuery supports federated queries for processing data in non-native BigQuery storage systems without moving data and with increased efficiency. In particular, BigQuery's ability to query data in Cloud Storage and the ability to run Spark and Hadoop jobs on BigQuery data directly make it easy to integrate your data lake with your data warehouse.

"We migrated our enterprise data warehouse to BigQuery, and that was a huge effort, but it was a tremendous success for our company."

David Narayan, *Distinguished Engineer, The Home Depot*

---

5   ESG Publication, "The Economic Advantages of Migrating Enterprise Data Warehouse Workloads to Google BigQuery," March 2019.

Realize the power of advanced and predictive insights

The ML tools in BigQuery make advanced and predictive analytics available to more people across the organization. For example, you can:

- Operationalize ML without the need to move your data from the warehouse

- Build ML models using SQL with BigQuery ML or unlock advanced ML use cases with AI Platform and TensorFlow integrations, allowing you to train ML models on structured data in minutes

- Use the BigQuery Storage API to execute open-source data science workloads—such as Spark, Hadoop, TensorFlow, MapReduce, Pandas, and more—and reduce data movement and increase efficiency

- Run advanced geospatial analytics with BigQuery GIS to gain location-based insights

Protect your data and operate with trust

BigQuery offers robust security, governance, and reliability controls that offer high availability and a 99.9% SLA, so you can have peace of mind that your data is protected. Data is automatically replicated, restored, and backed up to ensure business continuity, and it's encrypted at rest and in transit by default. Customer-managed encryption keys provide control over your data and leverage fine-grained identity and access management with Cloud IAM. And the vertical integration of BigQuery within Google's hardware and software security stack eliminates surface-level attacks.

Access and share insights with ease

Seamlessly access and share analytical insights within your organization as datasets, queries, spreadsheets, and reports with a few clicks. Securely create public and external read-only datasets to scale knowledge sharing and collaboration with your external business stakeholders. Analyze large and complex datasets interactively with sub-second query response time by combining BigQuery BI Engine—an in-memory, column-oriented analysis service—with your favorite business intelligence (BI) tools. Easily create stunning reports and dashboards using Looker and Google Data Studio, as well as other popular BI tools like Tableau, Qlik, and more. And securely share insights at any scale.

Customer case study

StubHub: Partnered with Google to transform their data infrastructure and empower teams to unlock the power of data.

Strategy

- Replaced legacy data warehousing infrastructure with BigQuery
- Used BigQuery to establish a new data lake and set up an ETL framework
- Migrated 60% of their data to BigQuery in less than a year

Results

- Gained the ability to run complex queries in minutes, including ones that always used to time out
- Created a single source of truth for all of their data
- Laid the foundation for using BigQuery ML to automate anomaly detection

"We wanted productivity tools that would allow us to make time richer for team members in support of our customers and achieve 10x improvements rather than small gains."

Yao Morin, *Chief Data Officer, StubHub*

StubHub

Customer case study

Maisons du Monde: A data-driven approach for customer-focused retail.

Strategy

- Used BigQuery to analyze each touchpoint leading to a conversion
- Generated item recommendations that correlate to a particular product based on probability calculations, using Dataproc
- Analyzed comments automatically using Google Natural Language API and in-house designed clustering

Results

- Gained a data-driven approach to optimizing the customer journey
- Delivered more targeted customer service and enabled the personalization of marketing mail-outs and website recommendations
- Sped up the integration of customer feedback

"The system doesn't need any oversight so we can focus on improvements and innovations, as the other big advantage with Google Cloud is being able to try things out. We can experiment quickly, and it costs very little to do so."

Karim Louedec, *Chief Data Officer, Maisons du Monde*

MAISONS DU MONDE

Taking the next step

Leading organizations around the world use BigQuery to jump-start data analytics cost-effectively, make informed decisions based on real-time data, predict what's next for their customers and market, share insights securely and at scale, and protect their most sensitive data. From The Home Depot and Zulily to 20th Century Fox and HSBC, customers are accelerating their digital transformation journey with ease.

Learn more about how BigQuery can help businesses like yours turn data into value and make better decisions, faster.

Google Cloud