Google Cloud

Building a Serverless Data Warehouse

Zeke Bishop
Enterprise Customer Engineer
zekebishop@google.com

June 2021

Generating value from a future-ready data strategy

1. Data is the fuel accelerating business innovation in today's environment
2. The race is on and frontrunners have already achieved exponential business value in partnership with Google
3. Why Google Cloud is Uniquely Positioned to Deliver the Future Ready Smart Analytics Platform for your Enterprise?
4. How do we work together to quickly build a foundation on Google Cloud

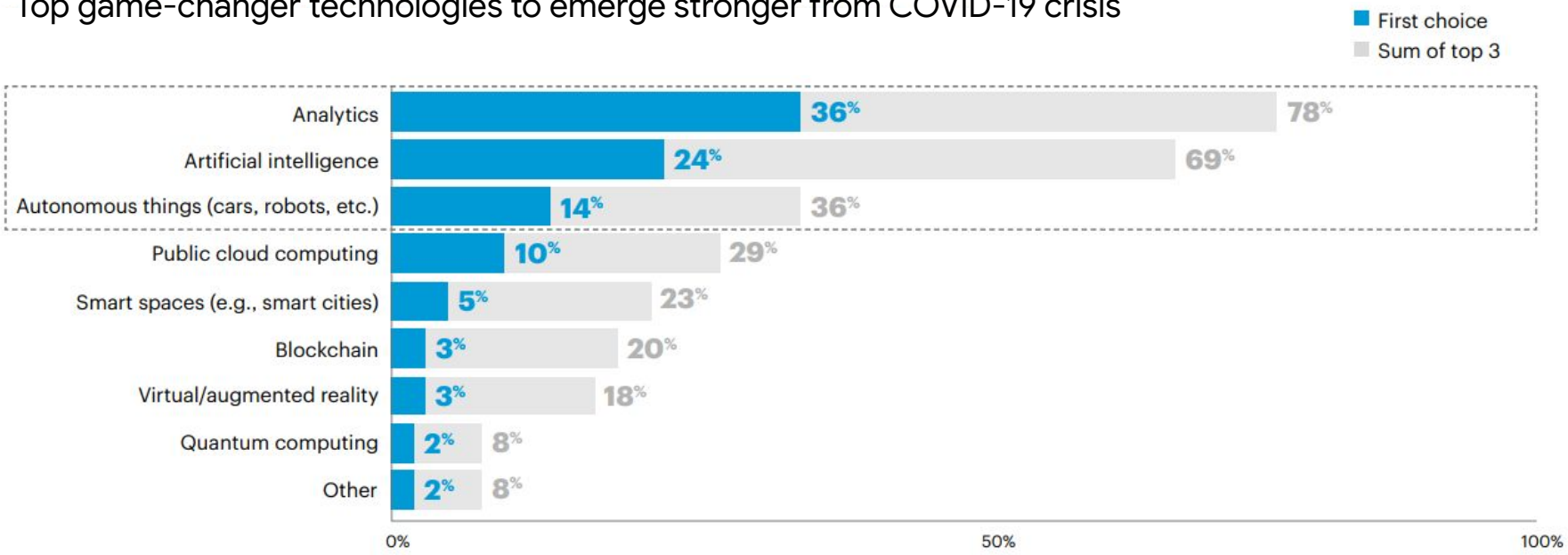
Data is the fuel accelerating business innovation in today's environment

01

The heat is on:

Boards of directors place analytics and AI as the #1 & #2 priorities

Top game-changer technologies to emerge stronger from COVID-19 crisis



n = 255, all respondent, excluding "can't say"
Q. Which will be the top 3 game-changer technologies for your industry to emerge from the COVID-19 crisis?
Source: View From the Board of Directors 2021, CIO Research, July 2020

What are we hearing from business leaders today?

Deliver **insights**, not infrastructure

Empower analytics **across** my enterprise **silos**

Embed ML and drive an end-to-end lifecycle

Understand our Industry **Context**

Technology needs to **Scale** with my enterprise

Businesses need actionable insights from their data wherever it resides

Our Approach:

- A 360 degree view of the businesses by **breaking data silos**
- To be situationally aware of, and responsive to, **real-time business events**
- A clear vision to enable **data-driven decision making** across the organization
- To gain **prescriptive & predictive** insights faster
- Simple and powerful data **security & governance**

The race is on and frontrunners have already achieved exponential business value in partnership with Google

Traditional data warehouses create bottlenecks that hinder business transformation

Cannot scale fast enough to keep up with data growth

Data is not fresh or current enough

Does not support machine learning and AI initiatives

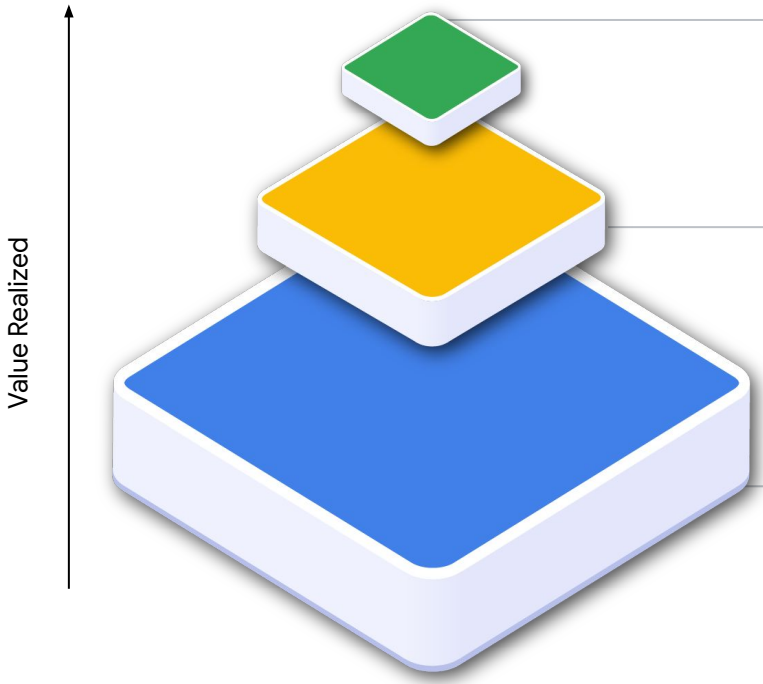
Data access restrictions inhibit collaboration

Expensive to purchase, renew, and maintain legacy systems

Common Data Opportunities facing Organizations Today

Problem Today	What you Need	Why this Matters	Google Capability	How Google is Different	Value
Too many decisions are done by guess-work	Break down data silos	The more data you bring together, the better the quality of the insights you get through AI	Unified data platform that is interoperable	• A single source of truth without centralizing data -- keep data ownership structure in place while, at the same time, breaking down data silos	**Intelligent** • Autonomous infrastructure • Faster, richer insights • Improved decision making
Throwing away fine-grained transaction and log data that is potentially very valuable	Lower TCO	Store more of the data you are already collecting	Serverless, low cost of operations - full separation of compute from store Log Analytics solution	• Pay only for what you use; partitions make unused data inexpensive • GA 360 from your web properties	**Open** • Works across clouds and on-prem • Open-source alternatives • Single source of truth
Not enough top-tier ML talent to go around	Democratize analytics and ML	Take advantage of domain expertise while upskilling your workforce	SQL interface to ML State of the art pre-built ML models	• ML without moving data • State of the art pre-built ML models	
My suppliers, partners, and customers have data I can leverage **Need to monetize data**	A data exchange	The more data you bring together, the better the insights are.	Securely share data, in raw or aggregated form	• Share data in place • You don't have to pay for their compute	**Flexible** • Simplified enterprise architecture • Interoperates with familiar technologies
Decision making is too slow	Streaming analytics	Make decisions as events happen, not once a day	Serverless stream analytics Streaming SQL Streaming insights	• The same tools that work for batch automatically work for stream	

Value from data modernization solutions is generated at three levels



Rethink the business - Transformational Impact

New products, services, and markets by combining data transformation with AI/ML, APIs, and collaboration tools leading to ***Co-Innovation / Business Transformation***

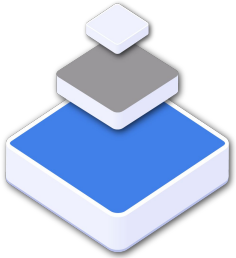
Build new capabilities - Maximize Revenue & Lower Risks

Usage of data analysis to enhance business processes, increase product penetration, improve risk management, etc. thereby improving on ***Revenue / Business Cost / Risk Drivers***

"Move & Improve" Data Warehouse Infrastructure - Lower TCO

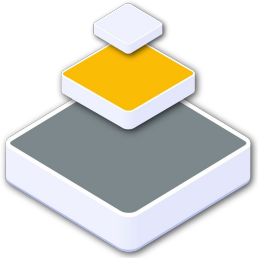
IT cost reductions due to lower storage and administration costs, improved development time, and enhanced flexibility - ***IT Total Cost (TCO) Reduction***

There are several drivers of value across the business driven by data modernization



"Move & Improve" data warehouse

- Reduced infrastructure costs
- Reduced system deployment costs
- Reduced query execution times
- Reduced data model and query development and maintenance costs

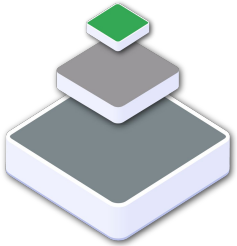


Build new capabilities - Revenue Impact

- Improved customer retention
- Reduced revenue loss due to stock-outs
- Improved marketing effectiveness
- Higher ecommerce conversion rates
- Improved app engagement

Build new capabilities - Cost and Risk Impact

- Optimized marketing costs
- Reduced subscriber / customer acquisition costs
- Reduced inventory costs
- Reduced manufacturing costs
- Improved business user productivity
- Improved IT productivity



Rethink the business - Transformation Impact

- Improved % of revenue from new products / services
- Improved new customer acquisition
- Higher return on care management investment

Customers have already begun to realize significant business benefits

"Move & Improve" data warehouse - lower TCO

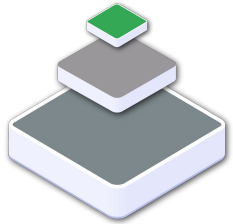
- **45% reduction** in development time **55% reduction** in maintenance costs — STATE AUTO Insurance Companies
- **20% reduction** in app testing time — PROGRESSIVE
- **$2M annual** cost reduction and **50% reduction** in analytics report processing — the YMCA

Build new capabilities - Revenue Impact

- **140% increase** in app engagement — FOX SPORTS
- Reduced credit analysis time **to 20 seconds from 5 days** — ANZ
- **4X increase in customer conversion** due to personalization — AMERICAN EAGLE

Build new capabilities - Cost and Risk Impact

- **60% increase** in barrelage per run — ABInBev
- **Stream 140M events** per day to better manage food inventory and delivery — Blue Apron
- **80% reduction** in ad serving and operations costs — Domino's

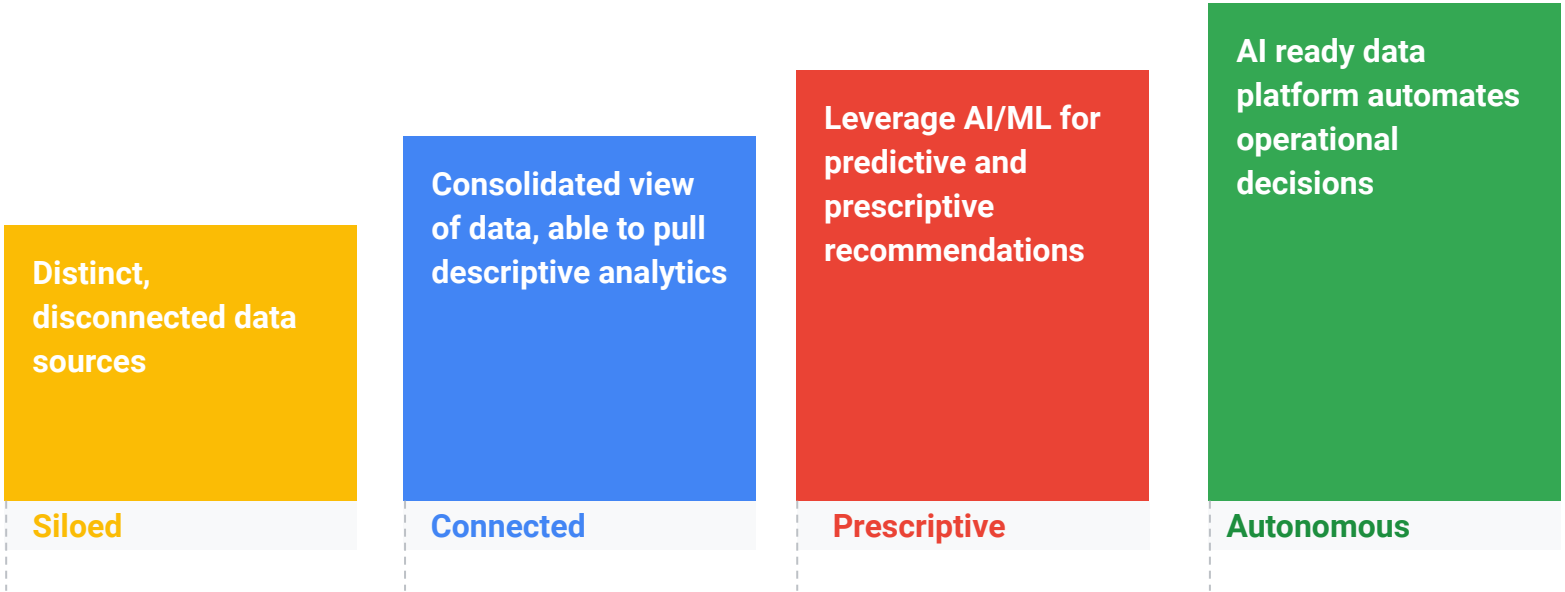


Rethink the business - Transformation Impact

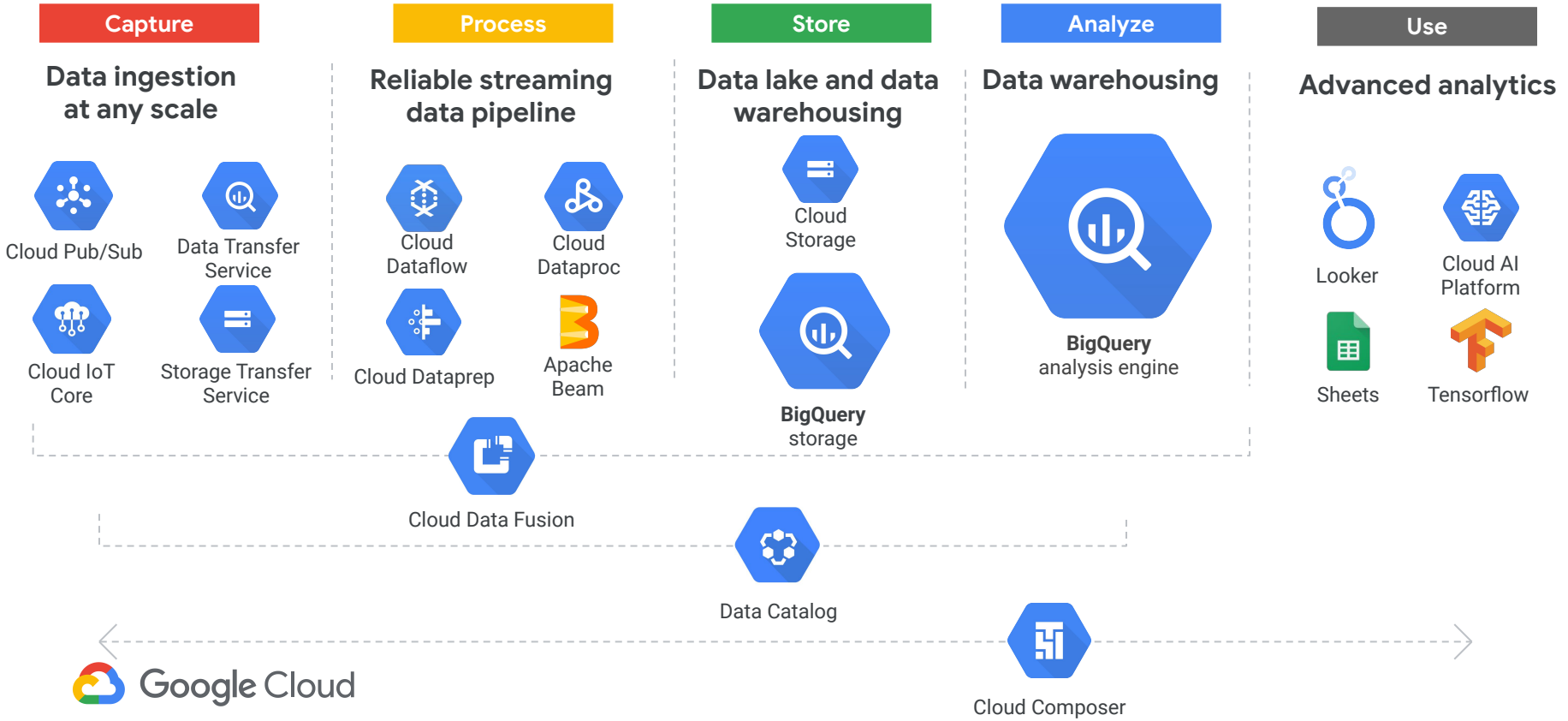
- **10X average return** on care mgmt. investment — Wellframe
- **Analyzed data from > 100M players** and solved game design challenges with AI — King
- Accelerated research time to market **by > 6 months** enhancing business model — Descartes Labs

Why Google Cloud is Uniquely Positioned to Deliver the Future Ready Smart Analytics Platform for your Enterprise?

The journey to a self-learning enterprise...



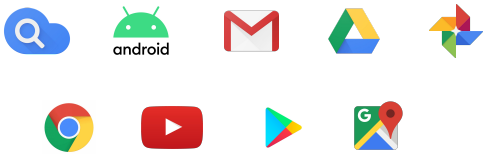
… Is fully supported by Google's Smart Analytics Platform



Why Google Cloud

Big data is in our DNA

9 products with > 1BN users



01 Performance at Scale

02 Total Cost of Ownership

03 Interoperability

04 Democratized ML/AI

05 Reliable & Secure by design

06 Real-Time Insights

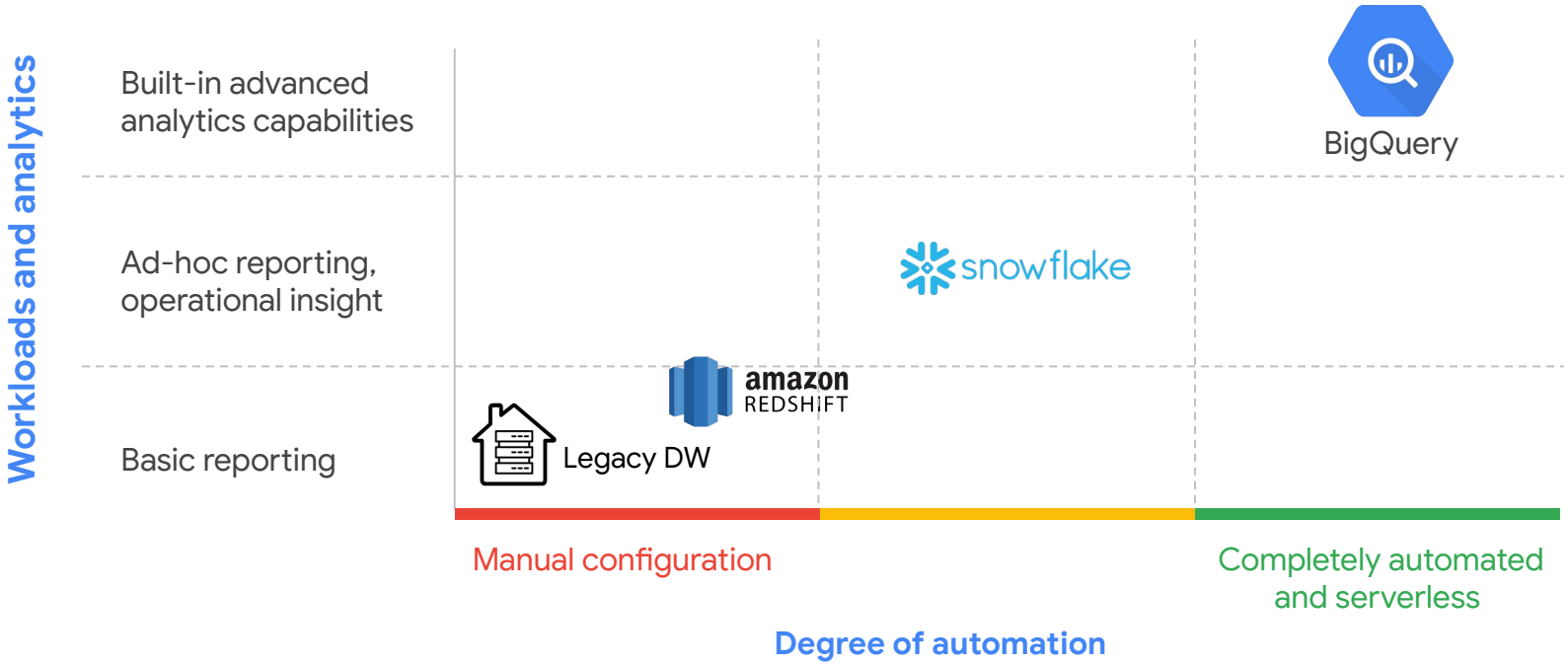
07 Usefully Multi-Cloud

08 Industry Leadership

Our Smart Analytics Platform Differentiators

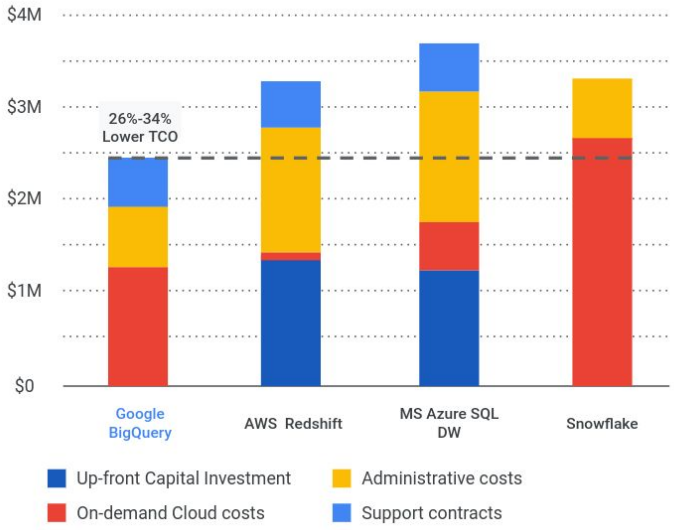
Performance at Scale

Petabyte scale, automated, and intelligent - lets your enterprise focus on delivering insights not infrastructure



Economic Value - BigQuery lowers your data warehouse TCO massively

Expected 3-Year Total Cost of Ownership



52% Lower TCO[1]
(versus on-premises)

26-34% Lower TCO[2]
(vs other Cloud DW's)

Flat-rate and variable pricing options to give customers control over TCO

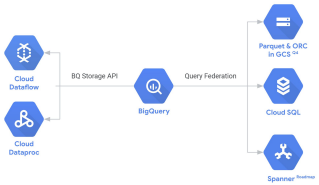
1) Migrating Enterprise Data Warehouse Workloads - ESG 2019
2) Google BigQuery vs. Alternative Cloud-based EDW Solutions - ESG 2019

Google Cloud

Interoperability & Faster Time to Insight by Breaking Down Data Silos

No ETL: Interoperability

Unified, interoperable platform across your Data Warehouse and Data Lakes with best of breed capabilities in both areas. No compromises.

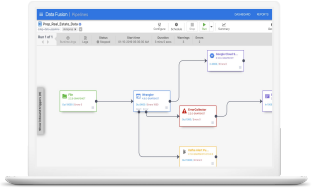


Data Transfer Service

Directly land data into BigQuery from SFDC and 100+ business apps

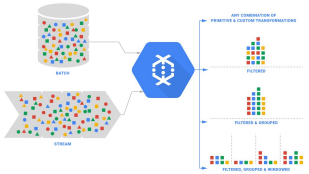
Visual ETL Pipelines

Code-free ETL and data integration across on-prem and cloud sources using Cloud Data Fusion



Code-free Streaming ETL

Dataflow Templates simplify data pipeline development and lower data latency

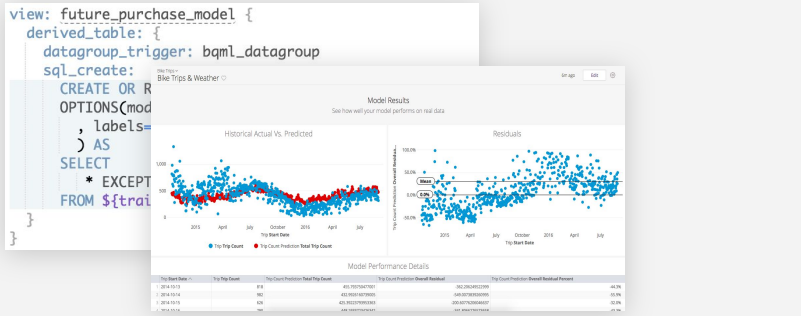


Unlocking Machine Learning for the Enterprise

Build custom ML models with standard SQL

BigQuery ML allows data analysts or data scientists to build and operationalize machine learning models, directly within BigQuery, using simple SQL. This has been game changing for organizations.

1. **Execute** ML initiatives without moving data from BigQuery
2. **Iterate** on models in SQL in BigQuery to increase development speed
3. **Automate** common ML tasks, and hyperparameter tuning



Secure and Reliable by design

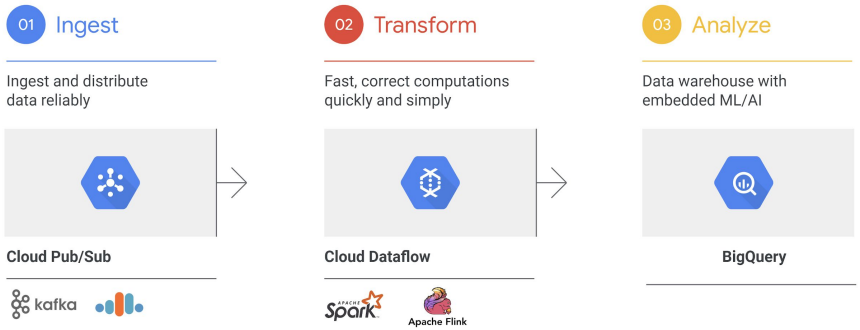
- **Reliable** with 99.99% uptime SLA
- **Maximum data durability** with data replication across multiple data centers
- **Data governance and security** with data access controls and regulatory compliance
- **Built-in data protection** with encryption, VPC service controls, and data replication

Google is one of our strategic partners. We're investing in machine learning and data capabilities with them because their cloud solution is approved to handle personal information securely and reliably.

HSBC

Richard Bates,
Global Head of HSBC's Intelligence Hub, Wealth and Personal Banking

Seamless Real-Time Insights



BigQuery **high-performance streaming** makes data immediately available

Pub/sub and Dataflow integrations allow customers to build comprehensive batch and streaming pipelines

BigQuery Streaming API increased streaming capacity by 10x

BigQuery BI Engine fast, high concurrency, in-memory analysis service for complex data sets

> *By collecting more data, we're innovating faster and making smarter decisions. Using real-time analytics on Google Cloud Platform, we were able to drive a significant increase in sales conversion in just a few weeks. Without Google Cloud Platform, we would never have been able to scale the clickstream data collection 100-fold in a fraction of the time.*

zulily

Bindu Thota, Director of Product Management

Cross Cloud Intelligent, Not just Multi-Cloud

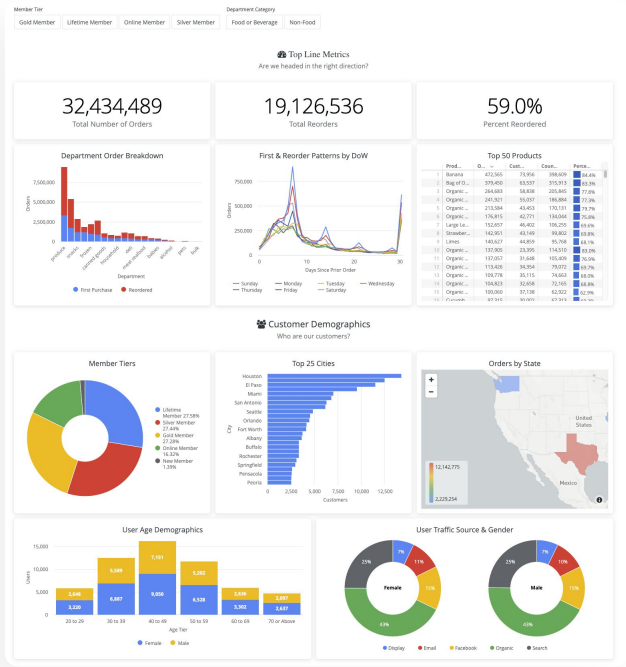
Analyze your data wherever it is, with a consistent experience, to gain insights across silos

Single pane of glass across data in AWS, Azure and GCP

Business reporting without data movement

Join streaming data with batch data across clouds

Unified workflow and query engine to improve BI integrations (Looker and Tableau)



Recognized as a Leader in Databases and Analytics

Gartner	**A LEADER** Cloud Database Management Systems
FORRESTER®	**A LEADER** Cloud Data Warehouse
FORRESTER®	**A LEADER** Data Management for Analytics
FORRESTER®	**A LEADER** Streaming Analytics
FORRESTER®	**A LEADER** Database-as-a-Service
FORRESTER®	**A LEADER** Big Data NoSQL
IDC ANALYZE THE FUTURE	**A LEADER** Data & Analytics Platforms, APeJ

Google Cloud

A Leader in Cloud Data Warehouse

Google receives **5 of 5** in 19 different criteria, such as:

- ✔ Data Ingestion
- ✔ Data Lake Integration
- ✔ ML / Data Science
- ✔ Performance
- ✔ Scalability
- ✔ Solution Roadmap
- ✔ Strategy Execution
- ✔ Customer Adoption
- ✔ Use Cases
- ✔ Partners

Google Cloud

The Forrester Wave™: Cloud Data Warehouse, Q1 2021, Noel Yuhanna

The Forrester Wave™ is a graphical representation of Forrester's call on a market and is plotted using a detailed spreadsheet with exposed scores, weightings, and comments. Forrester does not endorse any vendor, product, or service depicted in the Forrester Wave. Information is based on best available resources. Opinions reflect judgment at the time and are subject to change.

How do we work together on this journey to eventually become a “self-learning enterprise”?

How do we engage to initiate your data journey with Google?

1

Executive alignment

Set direction and expected outcome of the engagement.

2

Value discovery & analysis

Identify data use cases, value drivers, and the state of existing data platform. Identify potential Google data solutions.

3

Data roadmap & business case

Develop roadmap for data use cases linked to business impact, validated by business owners.

4

Value delivery & realization

Migrate identified use-cases in an iterative manner, test and validate migration and assess return on investment.

Identify top use cases and execute a PoC in a matter of weeks while working with Google to build a complete roadmap for your data journey

Our expertise is delivered through a robust partner ecosystem

DW Migration Tech Partners

Informatica

CompilerWorks®

DATOMETRY

striim

Global System Integrators

accenture

Infosys

wipro

Atos

Capgemini

HCL

TATA CONSULTANCY SERVICES

Cognizant

Deloitte.

Regional System Integrators

CI&T

datametica

Myers-Holum

NEUDESIC

slalom

Maven Wave

Pythian
love your data

epam

StackPros

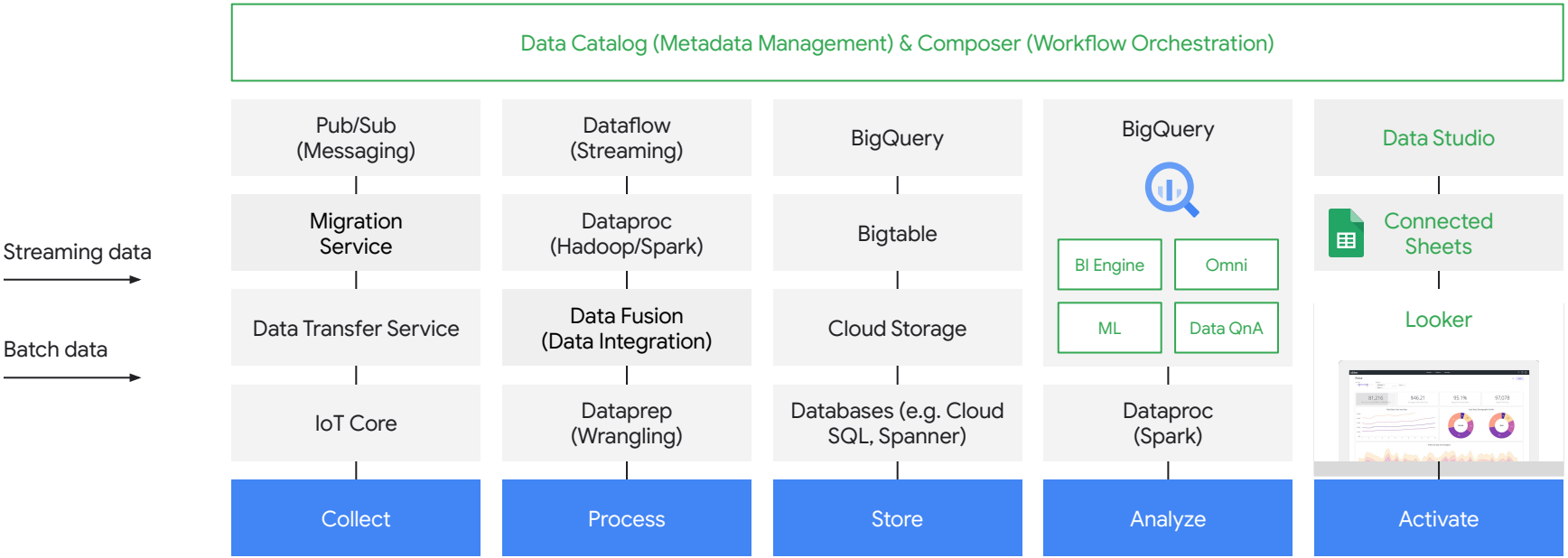
How Google's Smart Analytics Platform is Unique in the Industry

	Google BigQuery	snowflake	Legacy Solutions	
Scale	✓	Partial	✗	BigQuery is fully managed, serverless and architected for petabyte scale. While others are tied to clusters or require manual reclustering efforts BQ manages the infrastructure for you and allows your teams to focus on delivering insights
Total Cost of Ownership	✓	✗	✗	BigQuery eliminates the need for upfront investment and planning for your EDW, reduces operational and administrative expenses - all while delivering on business agility. Enterprise Strategy Group (ESG) estimated savings of 26-34% over cloud-based EDW alternatives and >40% over legacy on-premise solutions
Interoperability	✓	✗	✗	BigQuery provides a unified, interoperable best of breed platform across your Data Warehouse and Data Lakes and data integration across on-prem and cloud sources. BQ was made to tear down data silos and allow you to avoid creating new ones.
Democratized ML/AI	✓	✗	✗	BigQuery democratizes Machine Learning for the enterprise user (not just data scientists) with accessible capabilities using SQL. While allowing for more sophisticated data science teams to access the power of Google's leading edge AI technologies via Cloud AI. More than 80% of our BigQuery customers have incorporated ML into their business analysis
Reliable & Secure	✓	Partial	Partial	BigQuery offers robust security, governance and reliability that is unmatched in the industry. High availability and a 99.99% SLA, automatic data replication, restore and backup to ensure business continuity. Ability to classify and redact sensitive data, fine-grained identity and access management including access transparency so you can log each view. Data is encrypted at rest and in transit by default, and customer-managed encryption keys provide control over your data
Real-Time	✓	✗	✗	Designed to excel in IoT and other scenarios where your analysis depends on real-time streaming data as well as a BI acceleration engine for high-concurrency low-latency use cases - both are unique differentiators for Google Cloud and essential for businesses that need to make real time decisions
Usefully Multi-Cloud	✓	✗	✗	BigQuery breaks down the silos to provide a single pane of glass for all your data across multiple clouds (AWS. Azure). Most other vendors are focused on providing the same service running in 3 clouds but these are 3 silos. BigQuery breaks the silo and enables customers to analyze data across datasets
Industry Leadership	✓	?	?	Recognized industry leader by both Gartner and Forrester in Data Management and Analytics. With 9 Google products with more than a billion users running on our platform you can be sure that big data is in our DNA and we are ready to help your business build a future ready data platform

Google Cloud

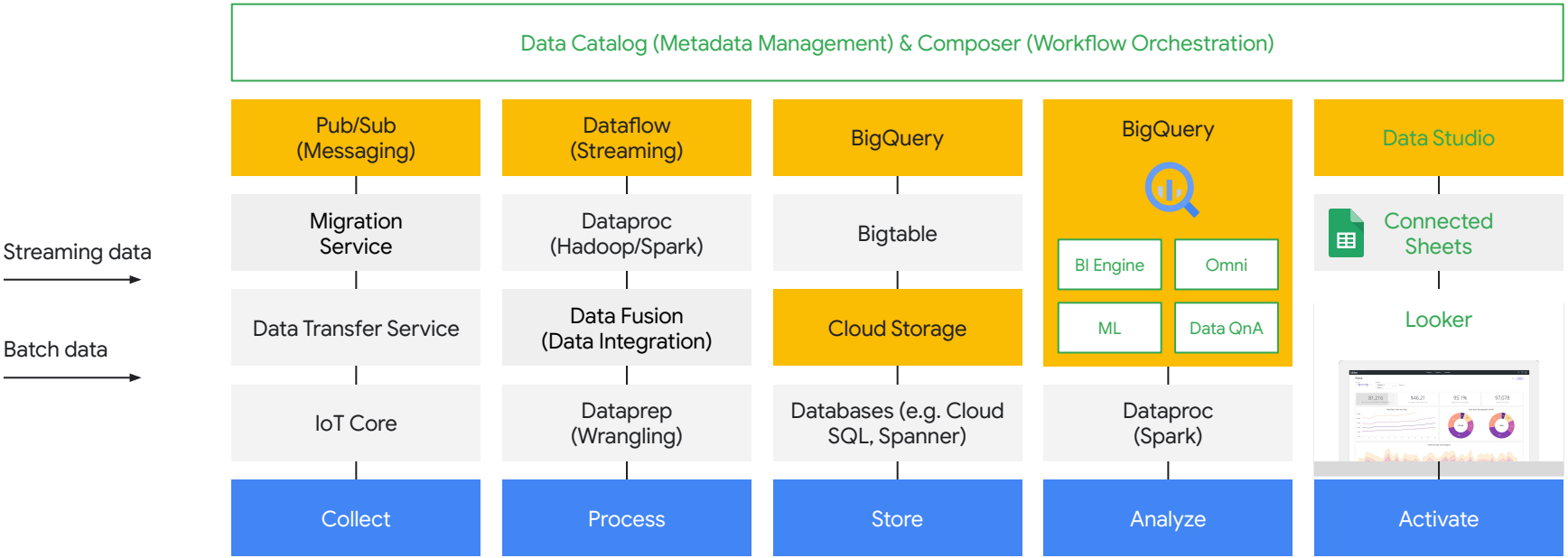
Google's Smart Analytics Platform

Open, Intelligent, Flexible



Let's Use GCP

Building Serverless Data Pipeline - https://codelabs.developers.google.com/codelabs/iot-data-pipeline#0



Thank you

Appendix

Google's Smart Analytics Platform Advantage [1 / 2]

Performance at Scale: BigQuery is fully managed, serverless and architected for Petabyte scale. While others are tied to clusters or require manual reclustering efforts BQ manages the infrastructure for you and allows your teams to focus on delivering insights

- Home Depot empowered business units to create their own BigQuery projects and reduced the dependency on IT budgeting for additional data warehousing capacity. This resulted in wider adoption of the platform and faster time to value.
- HSBC, the largest bank in Europe and #6 globally, migrated 30 years worth of data, 169 petabytes across 66 countries, to solve for the capacity constraints they faced managing their large environment on-premise

Total Cost of Ownership: BigQuery eliminates the need for upfront investment and planning for your EDW, reduces operational and administrative expenses - all while delivering on business agility. Enterprise Strategy Group (ESG) estimated savings of 26-34% over cloud-based EDW alternatives

- Metro AG, a large European B2B wholesaler, reduced infrastructure costs by more than 30% with GCP and migrating from Teradata to BigQuery
- Verizon Media chose Bigquery for their Media Analytics Warehouse to increase performance in their large scale, high concurrency environment, and lower TCO

Interoperability: BigQuery provides a unified, interoperable best of breed platform across your Data Warehouse and Data Lakes and data integration across on-prem and cloud sources. BQ was made to tear down data silos and allow you to avoid creating new ones.

- Woolworths**,** a large retailer in Australia, stood up their data infrastructure in GCP in under 5 months bringing together data across finance, supply chain and retail. They gained fast adoption of the new platform and quickly began experimenting with Machine Learning and other cloud based services to extract key insights

Democratized ML/AI: BigQuery democratizes Machine Learning for the enterprise user (not just data scientists) with accessible capabilities using SQL. While allowing for more sophisticated data science teams to access the power of Google's leading edge AI technologies via Cloud AI. More than 80% of our BigQuery customers have incorporated ML into their business analysis

- 20th Century Fox, a global movie studio, operationalized ML using simple SQL and created customer segmentation and targeting strategy for new releases in a matter of days. Their marketing analysts are able to execute on this project with ease and reduced dependencies on their lean data science team.

Google's Smart Analytics Platform Advantage [2 / 2]

Reliable & Secure by design: BigQuery offers robust security, governance and reliability that is unmatched in the industry. High availability and a 99.99% SLA, automatic data replication, restore and back-up to ensure business continuity. Ability to classify and redact sensitive data, fine-grained identity and access management including access transparency so you can log each view. Data is encrypted at rest and in transit by default, and customer-managed encryption keys provide control over your data

- Security is a top priority for HSBC, with $2.7 Trillion in assets. By managing their own encryption keys, leveraging data loss prevention technologies and Cloud Identity Access Management HSBC secures every bit of their data on BigQuery.

Real-Time Insights: Designed to excel in IoT and other scenarios where your analysis depends on real-time streaming data as well as a BI acceleration engine for high-concurrency low-latency use cases - both are unique differentiators for Google Cloud and essential for businesses that need to make real time decisions

- Zulily launches more than 9,000 SKUs every day on their online retail store. With BQ they are able to analyze the product performance and pricing data in real-time and make changes to their online merchandising dynamically. This has improved Zulily's online conversion significantly.

Usefully Multi-Cloud: BigQuery breaks down the silos to provide a single pane of glass for all your data across multiple clouds (AWS. Azure). Most other vendors are focused on providing the same service running in 3 clouds but these are 3 silos. BigQuery breaks the silo and enables customers to analyze data across datasets across all these clouds

- Broad Institute is able to analyze biomedical data stored in repositories across major public clouds right from within the familiar BigQuery interface, thus making this data available to enable search and extraction of genomic variants. Previously, running the same kind of analytics required ongoing data extraction and loading processes that created a growing technical burden. With BigQuery Omni, The Broad Institute has been able to reduce egress costs, while improving the quality of their research

Industry Leadership: Recognized industry leader by both Gartner and Forrester in Data Management and Analytics. With 9 Google products with more than a billion users running on our platform you can be sure that big data is in our DNA and we are ready to help your business build a future ready data platform