Data Engineer

Spotlight

07/20/22

Solve for increasing data complexity

Google Cloud

Priyanka Vergadia
Staff Developer Advocate,
Google Cloud

Closing the Data Value Gap

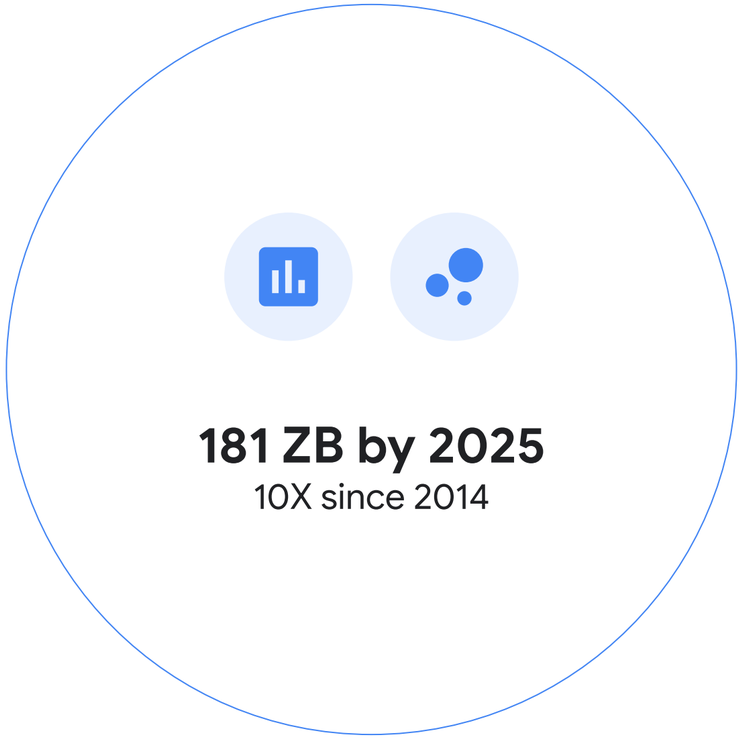
Data

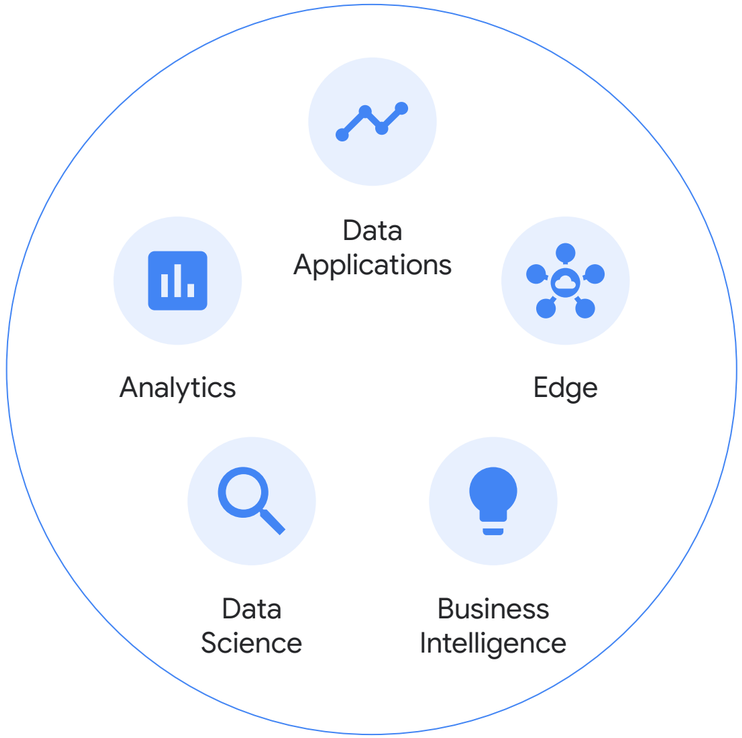
68%

of companies are **unable to realize tangible & measurable Value from Data.**

Value

Why is this happening?





Data volume, velocity & variety continues to grow at rapid pace

Data is increasingly distributed across storage systems, clouds, regions

Data is serving more users & use cases than ever before

- STATISTA, Feb 2022

What’s in store today at Data Engineer Spotlight

Latest how-to and best practices on our latest innovations; Q & A session

Curated sessions with demos and best practices to simplify data engineering from Google experts;

Q & A session following the curated sessions

Get hands-on and have some fun!

After the Q & A, we will host an **interactive Cloud Hero game**. Compete live with fellow aspiring and early career data engineers in a series of labs to learn the basics of BigQuery. Earn points as you complete the labs and if your name is on the leaderboard at the end, you might earn a prize!

Celebrate success and share resources!

Celebrate the winners of the **Learn to Earn Challenge**: go.qwiklabs.com/learn-to-earn

Google Cloud Community Forum: Join the conversation at googlecloudcommunity.com/gc/Data-Analytics

Announcing

Dataform PREVIEW

Q3 2022

Build and operationalize scalable SQL pipelines in BigQuery

End-to-end data transformation pipelines in BigQuery using SQL

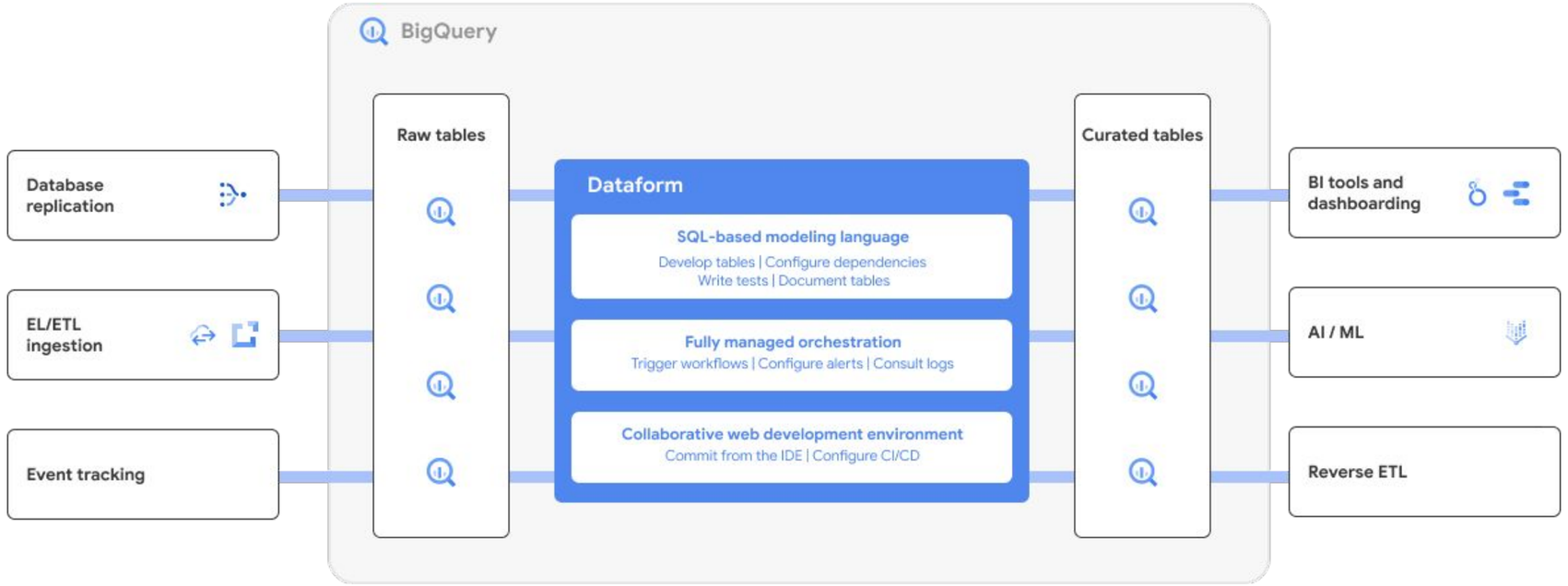
Develop and operationalize scalable data pipelines in BigQuery using SQL from a **single environment** and **without additional dependencies**.

Collaborate on SQL code using software development best practices

Follow software engineering best practices when managing SQL code, such as version control, environments, testing, and documentation. Connect your Dataform repository to **GitHub** or **GitLab** to leverage the best of CI/CD.

Empower data analysts to build production-grade SQL pipelines

Version control their code, manage dependencies, and orchestrate complex pipelines without additional skills other than knowing SQL.



Announcing

Latest Innovations for Dataflow

Smarter, easier, self-healing data pipeline optimization for truly unified streaming and batch data processing

> 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.
>
> — **Media customer**

55%
Developer productivity boost

50%
Infrastructure costs savings for streaming

<6 mo
Payback period

Ultimate collaboration without language compromises

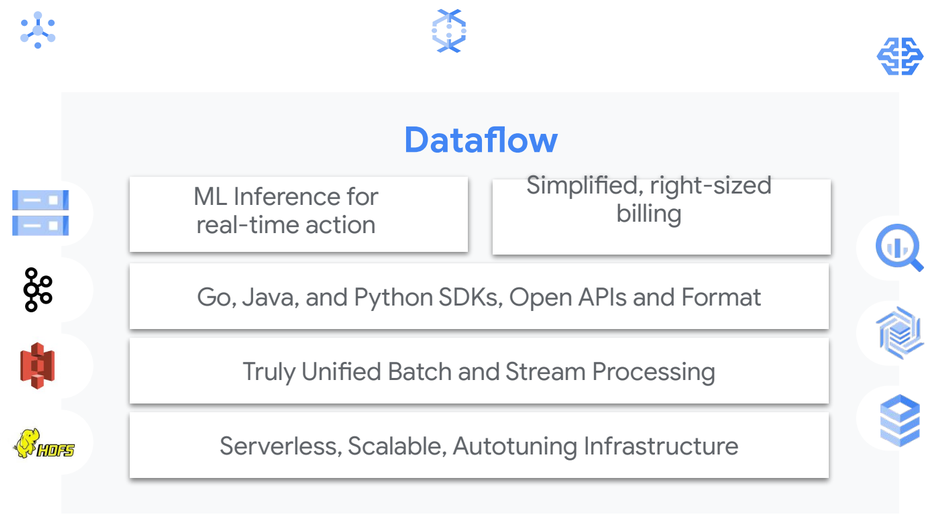
Dataflow Go enables performance-focused application engineers to write the core Beam pipelines in Go, data scientist colleagues to contribute to that pipeline with Python transforms, and data engineers to import their standard Java I/O connectors - and it all works perfectly well together in a single pipeline.

Apply ML models from all the popular frameworks, in real time

Dataflow ML makes it easy to apply Machine Learning models developed with PyTorch, TensorFlow, or scikit-learn to your application in real time, powering Continuous Intelligence for better business outcomes.

Ultimate simplicity for high performing data pipelines

Dataflow Prime removes the complexity of sizing and tuning while providing SLO-based smart diagnostics and simplified billing to enable developers to be more productive without worrying about machine types.



Announcing

Data Catalog is now a part of Dataplex.

Consistent and unified data governance across distributed data to accelerate time to business insights.

Unified metadata across distributed data

Automatic data discovery, & metadata harvesting, enriched with business context. Logically unify and organize your data without any data movement.

Centralized Security & Governance

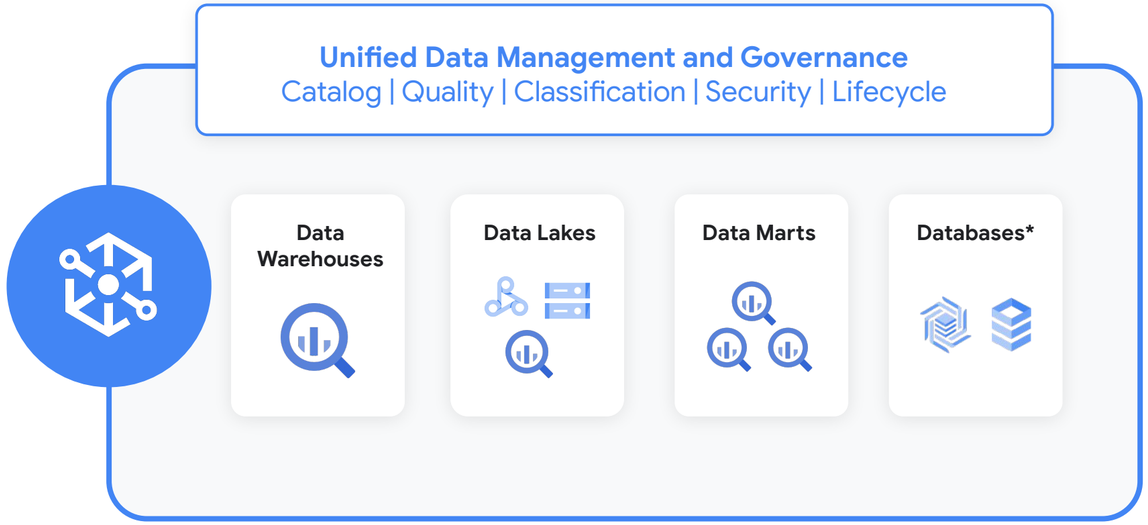
Central policy management, monitoring and auditing for data authorization, retention, and classification.

Intelligent Data Management

Built-in AI-driven intelligence with data classification, data quality, data lineage, and lifecycle management.

> We have **PBs of data** stored in Google Cloud, accessed by 1,000s of internal users daily. Dataplex enables us to deliver a **business domain-specific, self-service data platform across distributed data**, with decentralized data ownership but **centralized governance** and visibility. We are very excited to adopt Dataplex as a central component for building a **unified data mesh** across our analytics data.
>
> **— Saral Jain, Director of Engineering, Snap Inc**

Source: Build a data mesh on Google Cloud with Dataplex



**Databases support is a planned innovation*

BigQuery Migration Services (BQMS)

End-to-end migrations to BigQuery
Fast, predictable and free

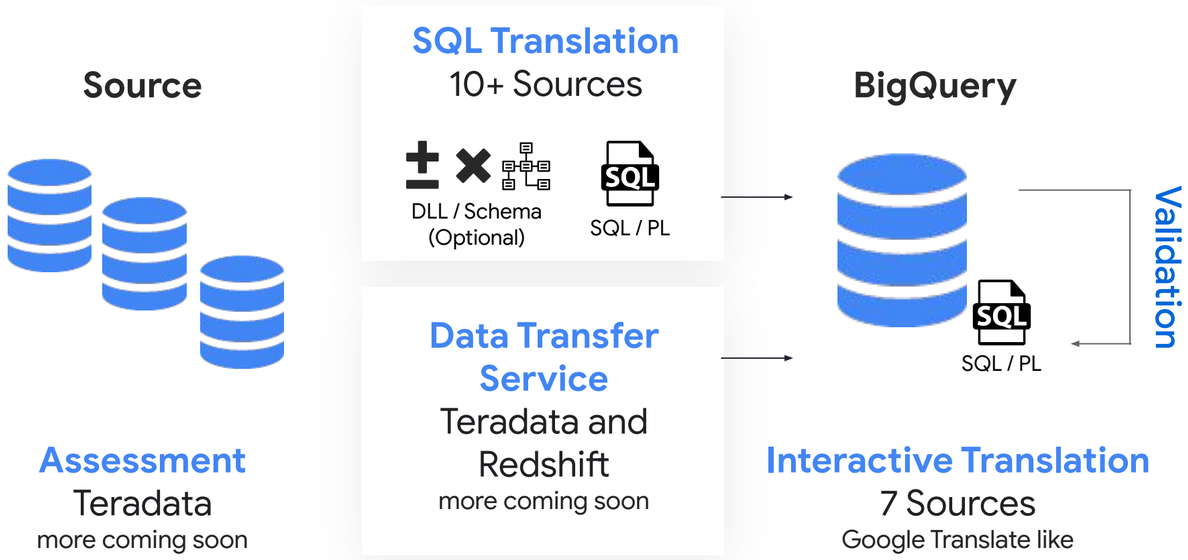
> With BQMS automated SQL translation, we were able to achieve over 90% translation accuracy in a matter of days. Our data platform teams couldn't be happier!
>
> — Large online retail customer

End-to-end data warehouse migrations
Suite of tools for migration planning, automated SQL/script conversion, data transfer and data validation

Reliably plan your migrations using Assessment
Get deep insights into how your existing usage will map to BigQuery, what order to migrate your workloads and how to optimize

Accelerate migrations with intelligent automated SQL translation
Receive semantically correct, intelligent, human readable translations of legacy SQL queries at scale, from most major data warehouses, with just a push of a button

Migrate data at scale and then easily verify data correctness
Customized multi-level validation functions at the table, column, and row level.



Google Cloud

Data Engineering Accelerators

Design Patterns

Combine technology with business value that can be used to delivery quick value or extended for specific use cases. Over 38 design patterns available.

Dataflow Templates

Simplify common data movement needs across Streaming, Batch, and Utility data workloads. Over 32 Google-provided and GCP-supported templates available in the UI.

Data Engineer Learning Path

Learn how to design and build data processing systems the Google way while getting certified!

Google Cloud Community

Seek support from peers and experts through the Google Cloud Community Forum

Thank you.

Google Cloud