Data Warehouse on Google Cloud

수많은 기업들이
BigQuery를 활용하는 이유는?!

Ver. 1.0
2022-07-14
update

Cloocus

AGENDA

Cloocus

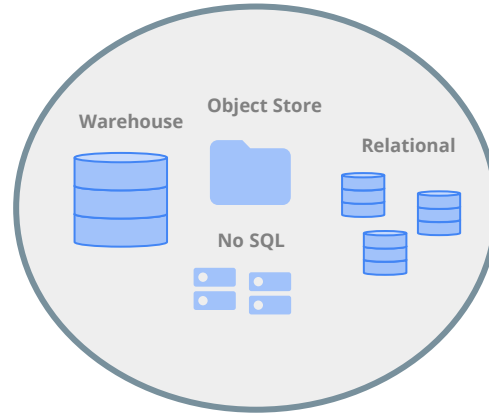# Closing the Data Value Gap
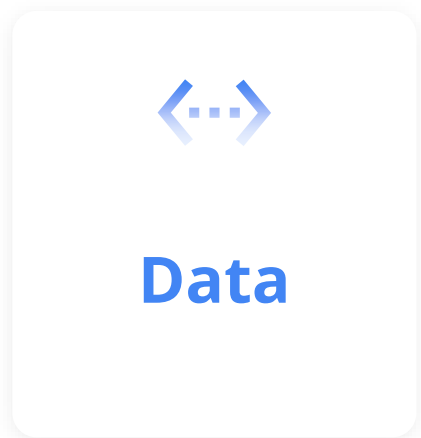
Data

68%

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

Value
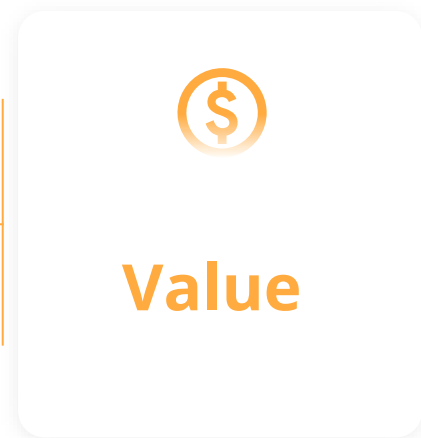
Accelerating trends are pushing the boundaries of existing data analytics stack

163 ZB by 2025

10X since 2014
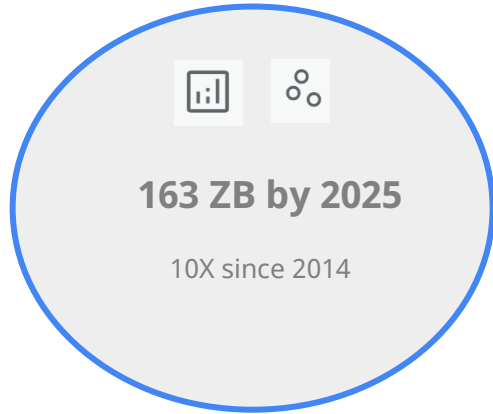
Data volume, velocity & variety continues to grow at rapid pace

Warehouse   Object Store   Relational   No SQL

Data is increasingly distributed across storage systems, clouds, regions

Analytics   Data Applications   Edge   Data Science   Business Intelligence
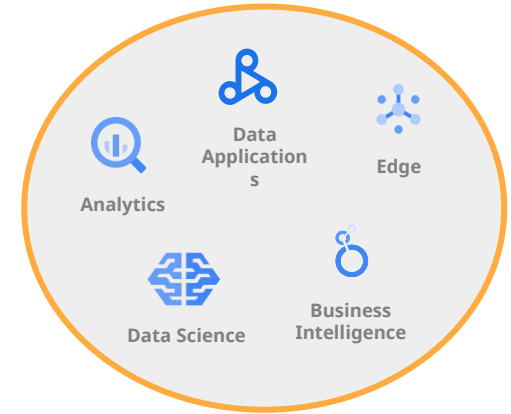
Data is serving more users & use cases than ever before

The evolution of business needs has created
a value gap

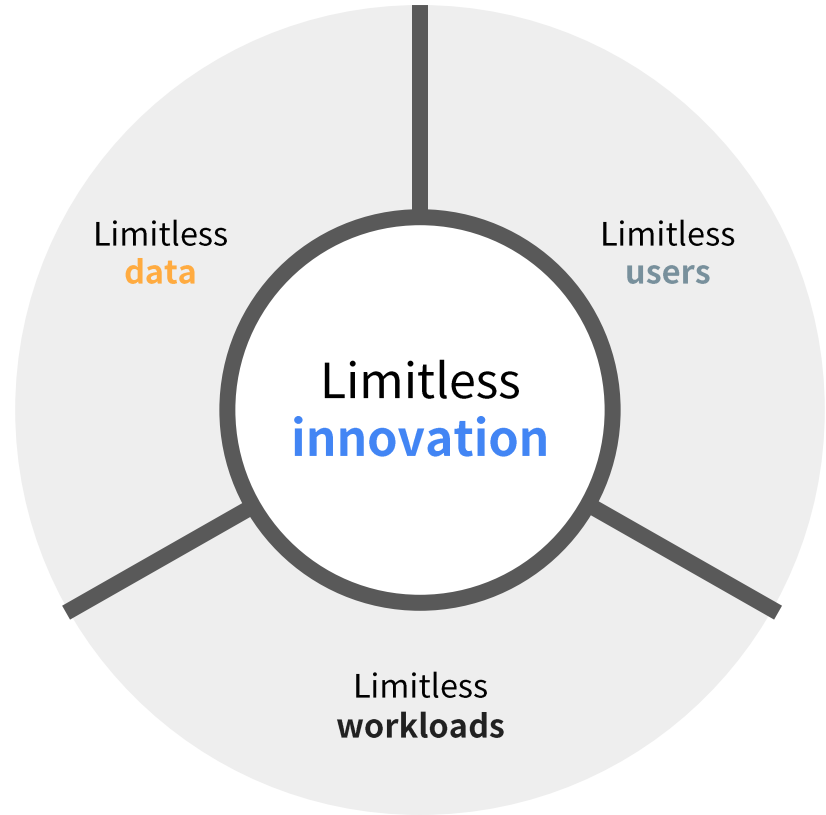**2022**: Exponential & distributed data, multiple use cases, complexity

2010s: More data and advanced analytics

2000s: Bring data together for analysis

Business needs over time

VALUE GAP **Scale, Cost**

VALUE GAP **Management overhead, Democratizing data**

VALUE GAP **Operational AI/ML, Streaming, Continuous Intelligence**

On-prem
Data Warehouses

Data Lakes
(HDFS)

Data Warehouses
in the Cloud and point
solutions

Google Cloud

Moving to the future requires a limitless data ecosystem to drive innovation

Limitless
data

Limitless
users

Limitless
innovation

Limitless
workloads

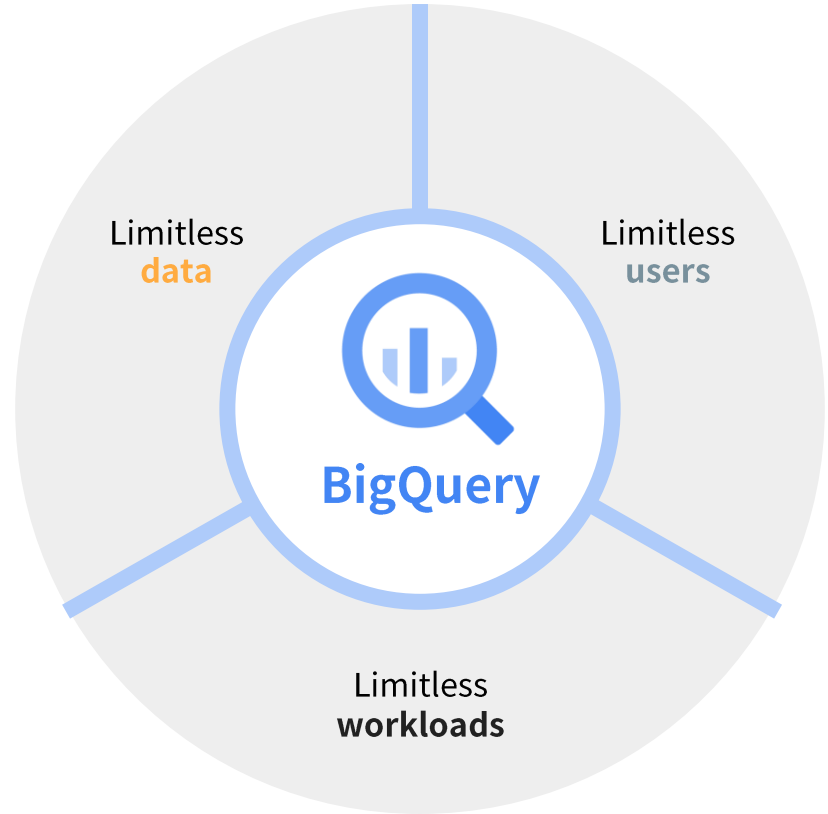# Google

Our mission is to organize the world's information and make it universally accessible and useful.

Customers unlock innovation with Google's Data Cloud

Limitless **data**

Limitless **users**

BigQuery

Limitless **workloads**

Google BigQuery

Data warehouse with customers ranging from TB to 100+ PB

- Cloud-scale enterprise data warehouse

- Standard SQL(ANSI 2011) with DML Support

- Encrypted, durable, highly available

Serverless platform — Unique

Real-time insights — Unique

Built-in **ML** — Unique

Insights for everyone — Unique

Continually improving performance

Graham Polley
@polleyg

Following

Holy shit balls! You're weren't kidding about those #BigQuery performance improvements @thetinot..🚀

```
 12    LANGUAGE
 13 ORDER BY
 14    total_views DESC;
```
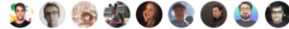
No cached results

▶ Run | ⬇ Save query | ⊞ Save view | 🕐 Sch

Query results   ⬇ SAVE RESULTS ▼

Query complete (16.7 sec elapsed, 4.1 TB processed)

11:40 PM - 7 Apr 2019

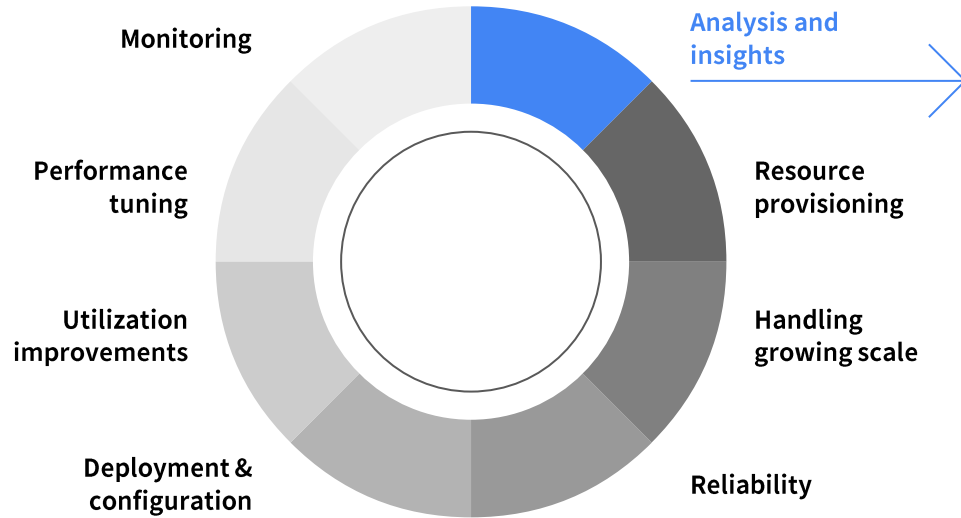15 Retweets 76 Likes

💬 7   ⟲ 15   ❤ 76   ✉

Tweet your reply

Graham Polley @polleyg · Apr 7
For comparison, this query used to take 50s-70s in the past.

Serverless big data analytics

Serverless big data analytics

Serverless data analytics

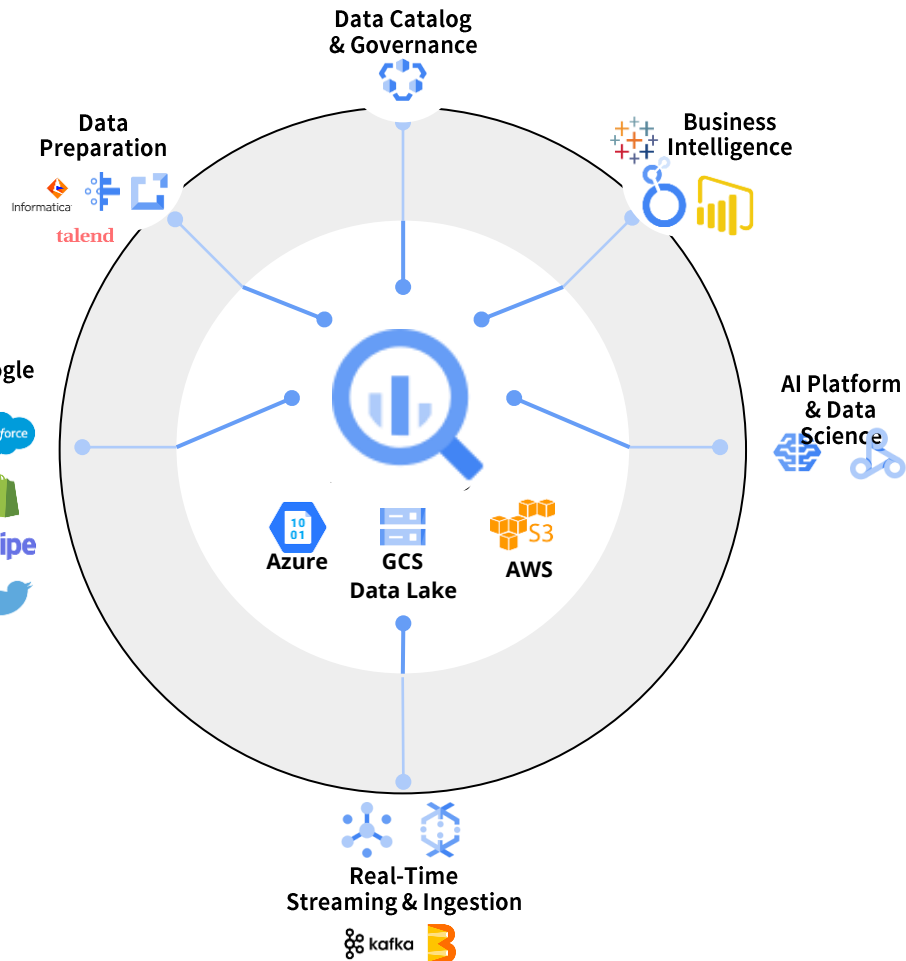# Limitless Data & Workloads

Intelligent: Self-optimizing architecture and ML "built-in" for broader adoption of advanced analytics

Simplified: Real-time and batch. Structured, unstructured, and spatial. In GCP, AWS and Azure

Reliable: Global scale at a predictable cost. Industry's best reliability with 99.99% SLA is 10x more reliable than other clouds

BigQuery Omni: Overview

BigQuery Control Plane
(UI / API / CLI) on Google Cloud

Google Cloud region

BigQuery Compute clusters (Dremel)

Distributed Memory
Shuffle Tier

Petabit Network

BigQuery Storage

Decoupled
compute & storage
for maximum
flexibility

AWS region (Omni)

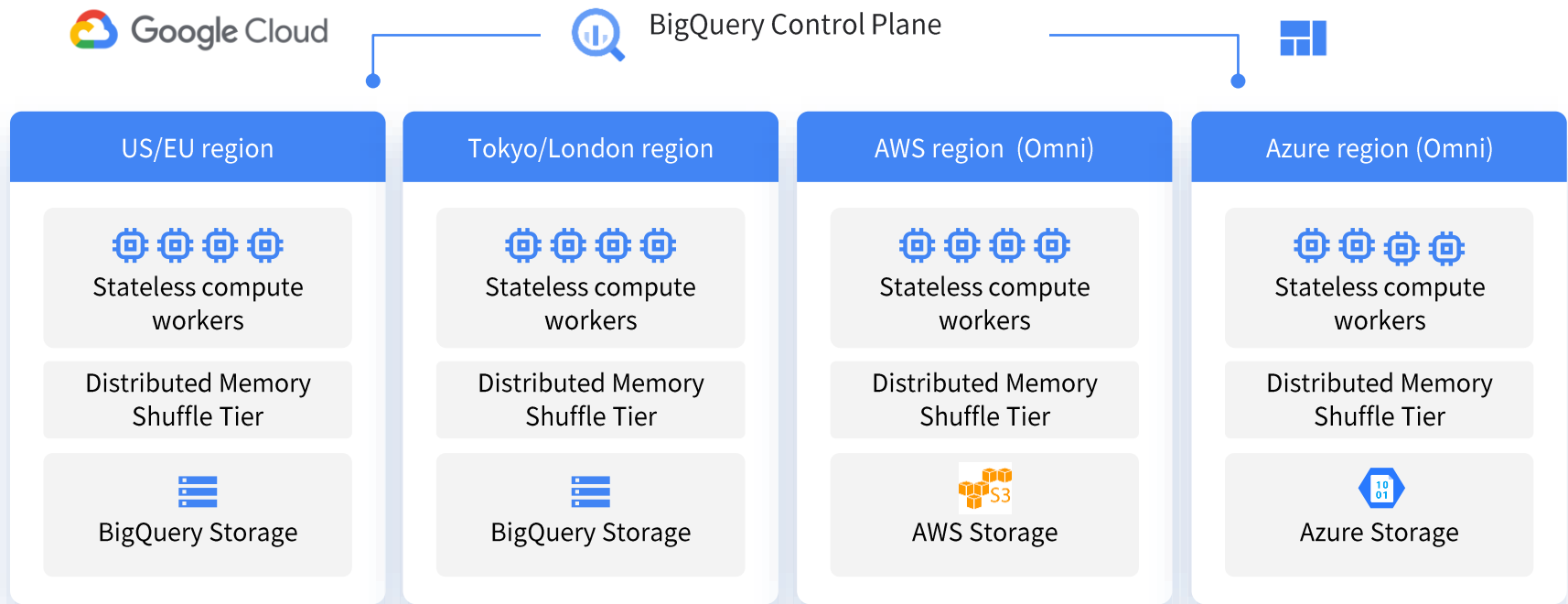**BigQuery Compute clusters (Dremel)**

Distributed Memory
Shuffle Tier

powered by Anthos
technology

AWS APIs

Customer S3 Storage

Embrace a flexible, multi cloud analytics solution with BigQuery Omni

Google Cloud — BigQuery Control Plane

US/EU region	Tokyo/London region	AWS region (Omni)	Azure region (Omni)
Stateless compute workers	Stateless compute workers	Stateless compute workers	Stateless compute workers
Distributed Memory Shuffle Tier	Distributed Memory Shuffle Tier	Distributed Memory Shuffle Tier	Distributed Memory Shuffle Tier
BigQuery Storage	BigQuery Storage	AWS Storage	Azure Storage

go/BigQueryOmni to learn more!

Cross-cloud data transfer

Google Cloud   BigQuery Control Plane

GCP Region

Stateless compute workers

Distributed Memory Shuffle Tier

BigQuery Storage

LOAD DATA FROM FILES
[Files on other cloud]

Use SQL to enable cross-cloud functionality

AWS / Azure Region

Stateless compute workers

Distributed Memory Shuffle Tier

AWS / Azure Storage

BigLake

Built on years of investment in BigQuery, BigLake is a **storage engine** that unifies data warehouses and lakes, by providing uniform fine-grained access control, performance acceleration across multi-cloud storage and **open** formats

Open API interface

Provides an Open API interface for query engines spanning across Google Cloud and open source runtimes to access your distributed data with consistent security & governance controls

BigLake

BigQuery	Object Store
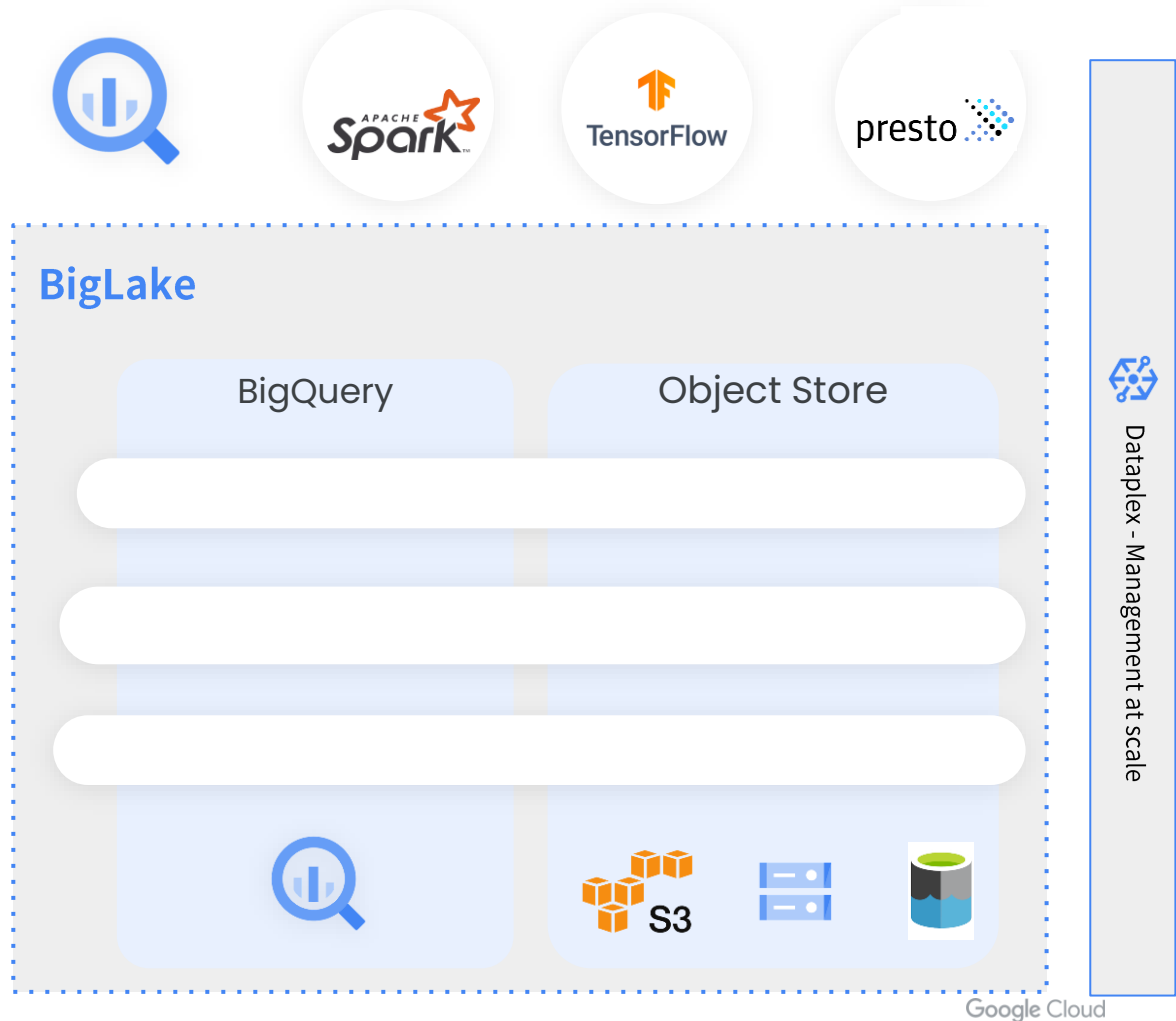
Dataplex - Management at scale

Google Cloud

BigQuery Security & Governance for Data Lakes

```sql
CREATE EXTERNAL TABLE bq_demo
WITH CONNECTION 'service account'
OPTIONS (uris=["gs://mybucket"])

CREATE ROW ACCESS POLICY row_access
ON bq_demo
GRANT TO group FILTER USING (c1='filter')
```

SQL

Authorized external tables extend the data management features of BigQuery tables to data lakes. Configured through a service account connection, authorized external tables allow you to manage end users without requiring you to grant permissions on files residing in data lakes.

Fine grain security for your data in object stores

Row and column level security for authorized external tables defined on Google Cloud Storage, Amazon S3 and Azure data lake storage.

Better interoperability with access through open APIs

Row, column level security **consistently enforced** when using BigQuery storage APIs from Spark and other Dataproc supported OSS engines.

Simplify user access management

Manage access similar to BQ tables **without needing to manage file level permissions**.

Google's open data platform around BigQuery delivers value from data

Intelligent: Self-optimizing architecture and ML built-in for broader adoption of advanced analytics.

Simplified: Real-time and batch. Structured, unstructured, and spatial. In GCP, AWS and Azure.

Reliable: Global scale at a predictable cost. Industry's best reliability with 99.99% SLA.
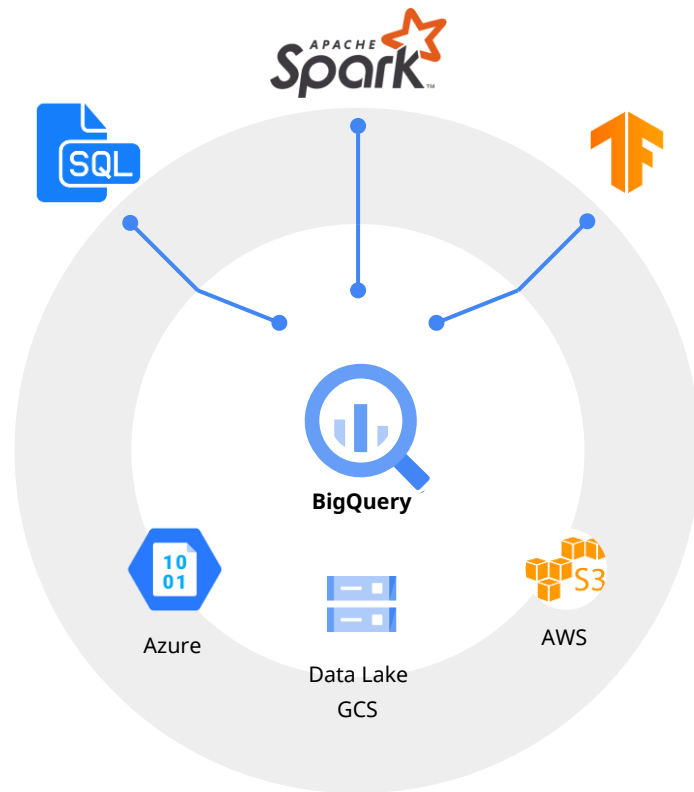
Announcing

Spark on Google Cloud

Industry's first Serverless Spark,
integrated with the best of Google Cloud.
Run and write spark where you need it across all
use-cases: ETL, data science and exploration.

Industry's First Serverless Spark for All Workloads

Auto-scale without any manual infrastructure provisioning or tuning for Spark. Empowers customers to shift from managing clusters to workloads.

Pervasive Spark Experience

Connect, analyze and execute Spark jobs from BigQuery, Vertex AI or Dataplex in 2 clicks, without any custom integrations, using the best of Google-native and Open Source tools.

Flexibility of Consumption

One size does not fit all. Choose between Serverless, Google Kubernetes Engine (GKE), and compute clusters for your Spark applications

APACHE
Spark™

BigQuery   Vertex AI   Dataplex   Composer
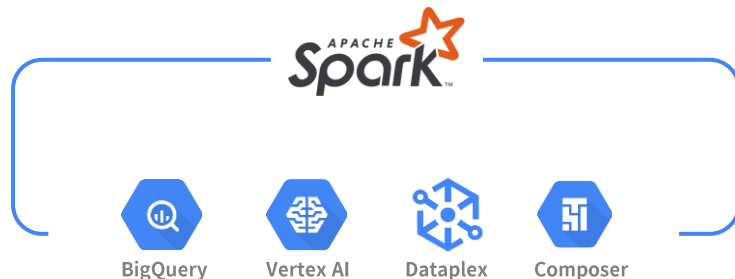
Serverless Spark: no clusters to create or manage

Google Cloud
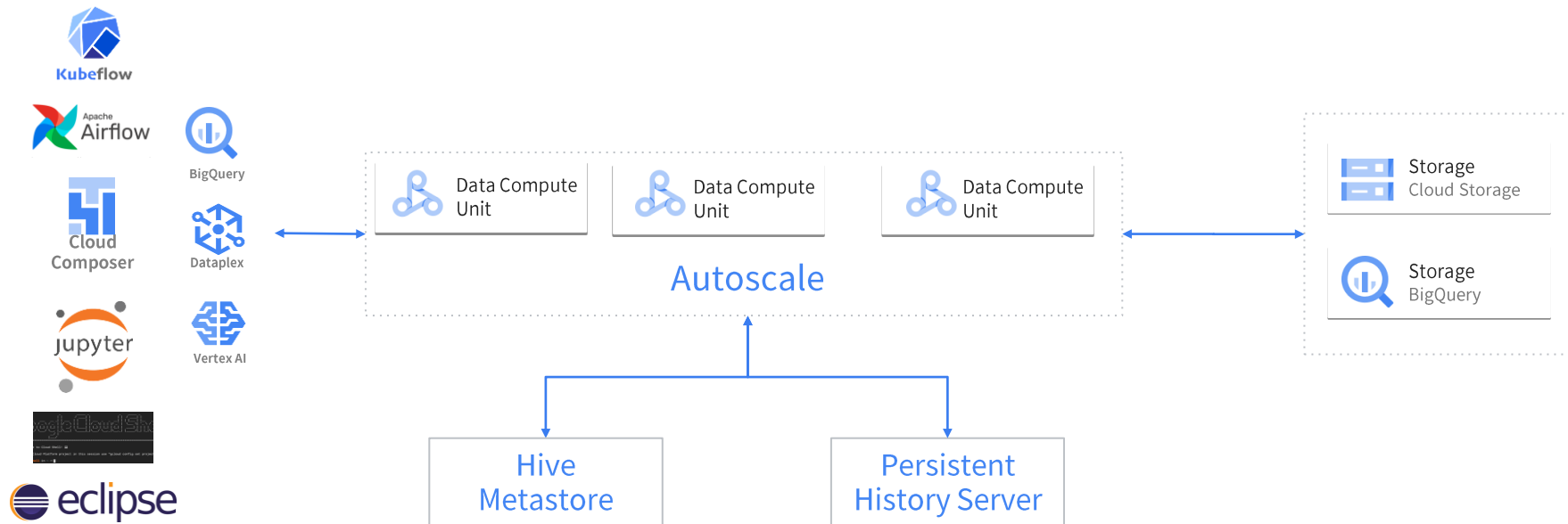
```
gcloud beta dataproc batches submit pyspark \
  --project=${PROJECT} \
  --region=${REGION} \
gs://dataproc-spark-preview/spark_terasort.py \
-- --gbs 100 --partitions 1000 \
  --base_dir gs://${BUCKET}/spark_terasort_100/${RANDOM} \
  --phases gen sort
```
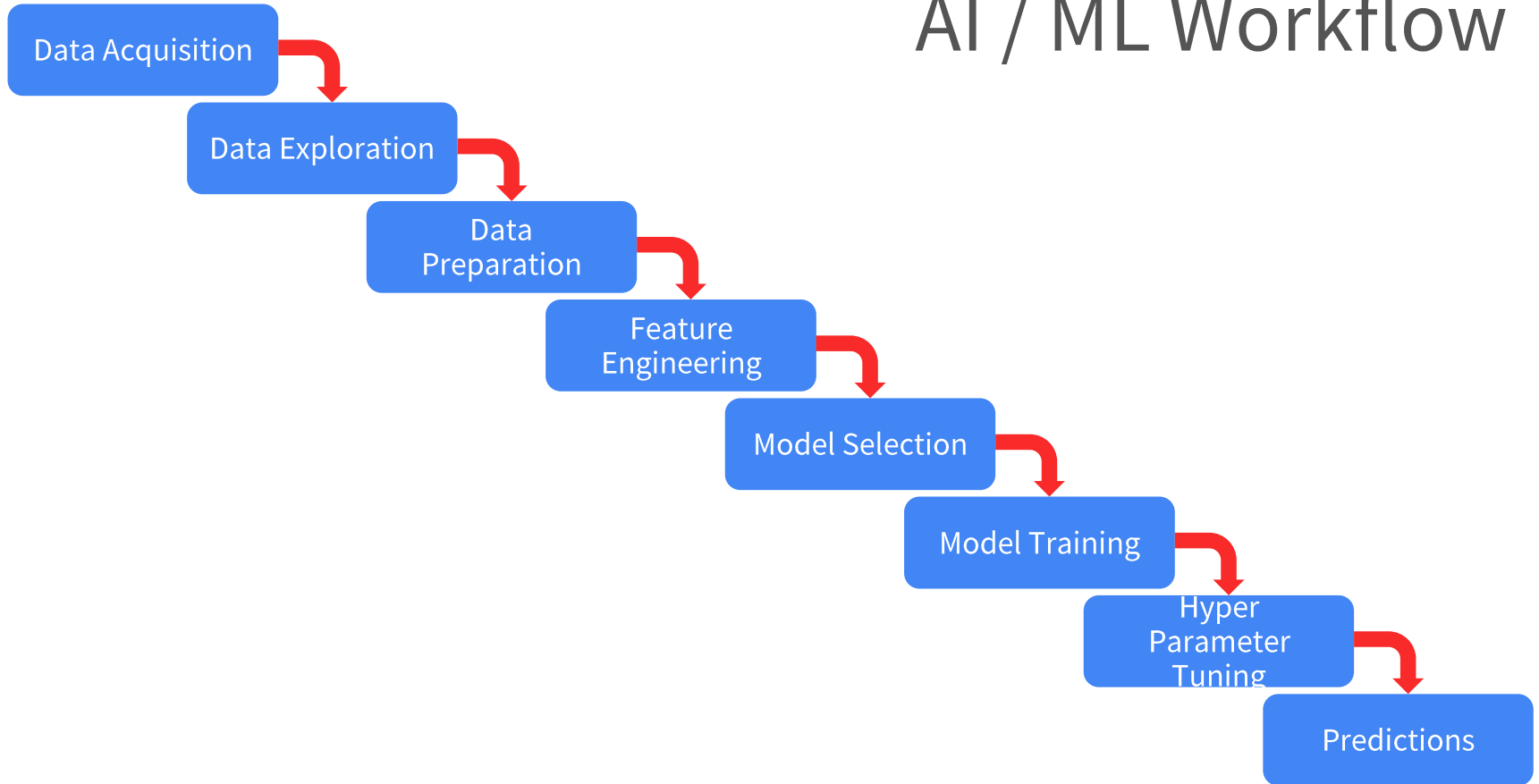
BQML

In-Database ML engine

AI / ML Workflow

Data Acquisition

Data Exploration

Data Preparation

Feature Engineering

Model Selection

Model Training

Hyper Parameter Tuning

Predictions

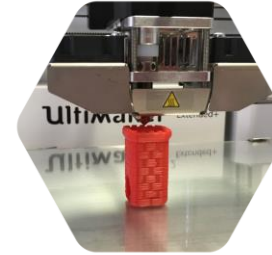# Pipeline for ML workflow

| Define objectives | Collect data | Understand and prepare the data | Create the model | Refine the model | Serve the model |

Inputs → Pre processing → Asset Creation → Train model → Model + Assets

Typical ML Workflow

Streaming/batch data

BigQuery (data warehouse) → Data processing → Export data → Train ML model (e.g. Python/R) → Operationalize ML model

Issues:

Data Governance?

Manage resources?

Where do I host?
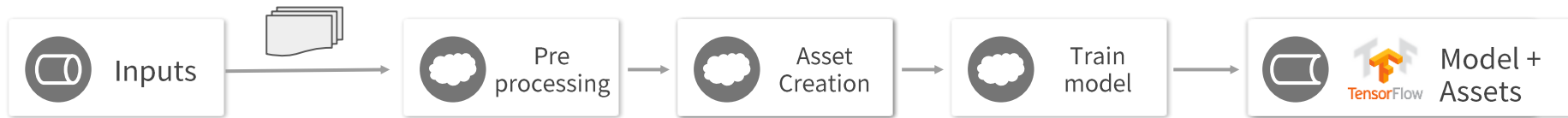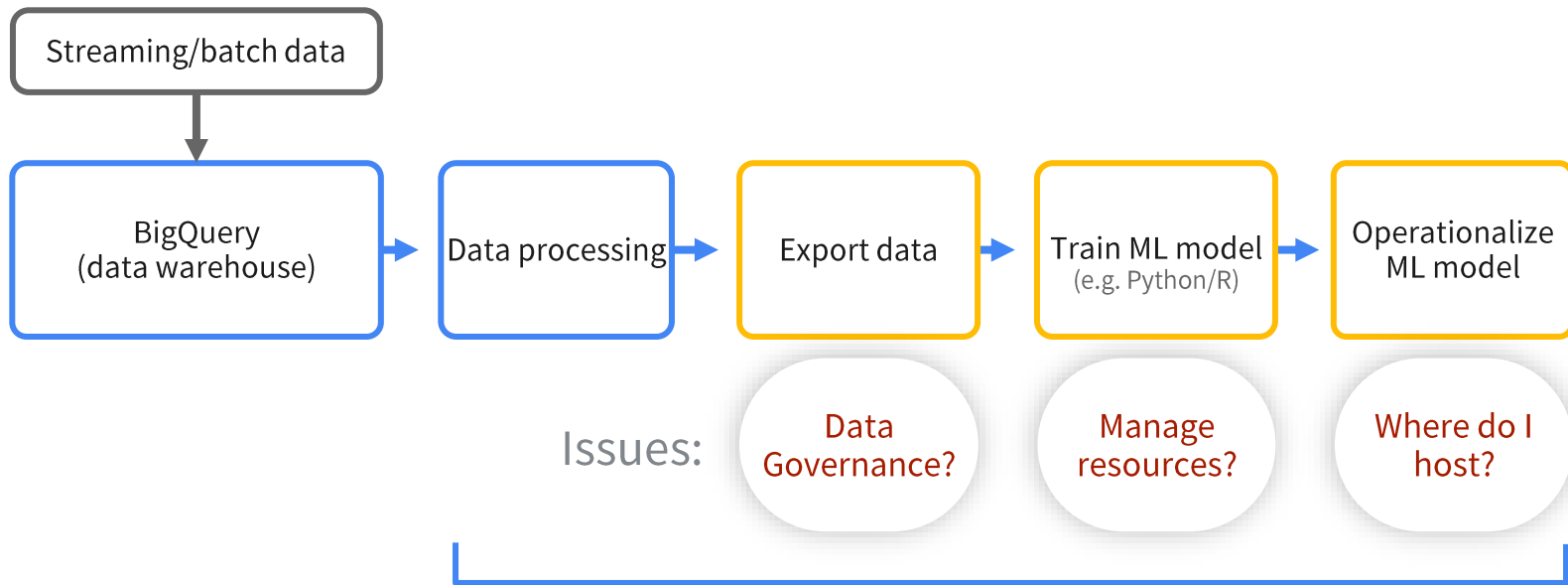
Multiple products & roles can lead to unnecessary complexity & costs

Typical ML Workflow

```
┌─────────────────────────┐
│   Streaming/batch data  │
└─────────────────────────┘
             │
             ▼
```

| BigQuery (data warehouse) | → | Data processing | → | Export data | → | Train ML model (e.g. Python/R) | → | Operationalize ML model |

Issues: Data Governance?   Manage resources?   Where do I host?

Simplify with BigQuery ML
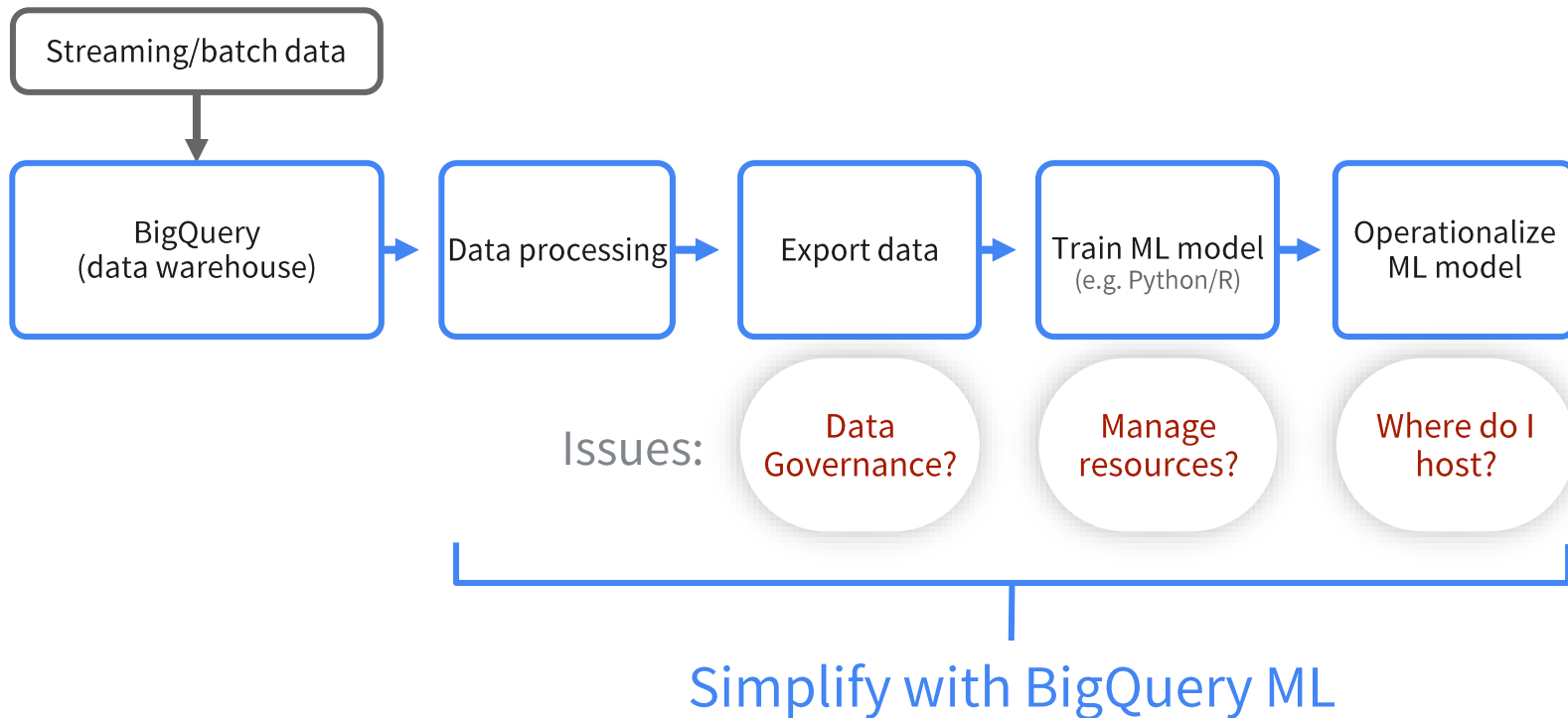
Key challenges affecting ML

Complex and time consuming to set up data pipelines and multiple tools

Data governance required across tools

Needs **advanced data science** and **programming skill** set
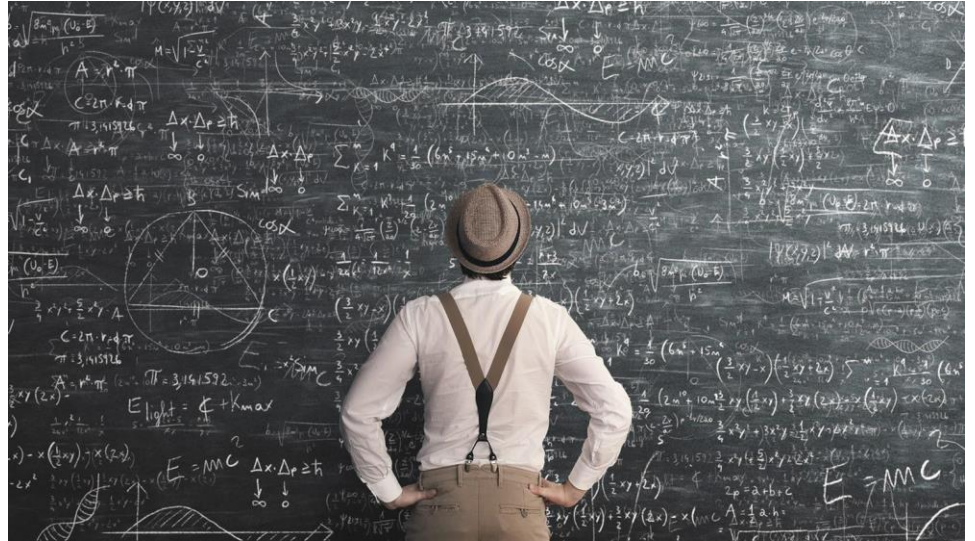
BigQuery ML makes machine learning super easy

The easiest way to unlock powerful AI/ML capabilities in BigQuery

- **Train and deploy** ML models in SQL

- **Execute** ML workflows without moving data from BigQuery

- **Automate** common ML tasks

- **Simplify** infrastructure management, security & compliance

Behind the scenes - BigQuery ML

Through SQL and within BigQuery

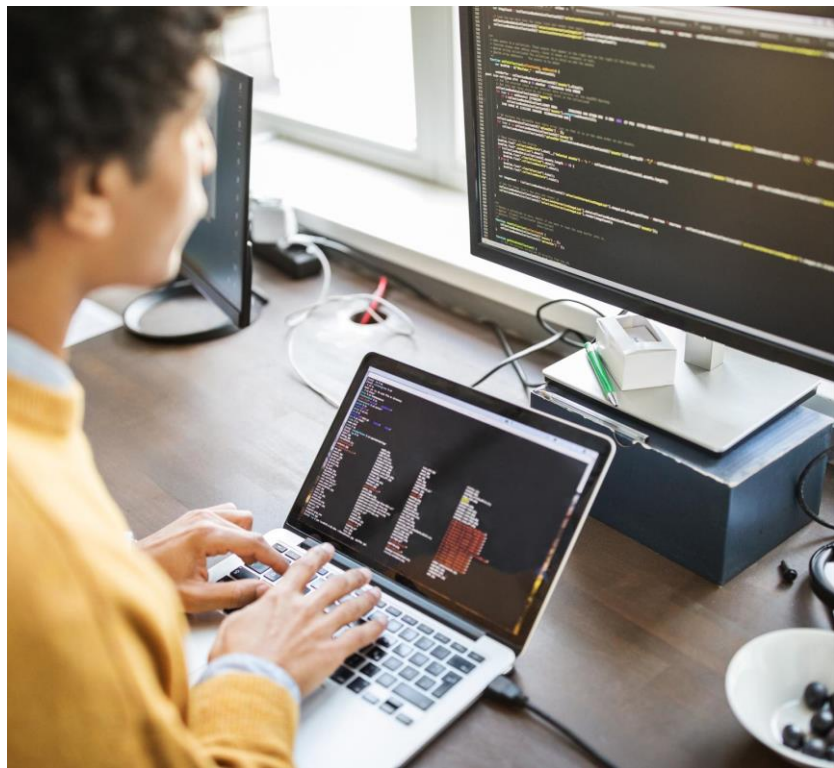Leverage BigQuery's processing power to build a model

Auto-tuned learning rate

Auto-split of data into training and test

Null imputation

Standardization of numeric features

One-hot encoding of strings

Class imbalance handling

Build and train with Create Model

Linear Regressor: given the weather, how many NYC Citibikes are in circulation?

🚴

```
CREATE OR REPLACE MODEL `sara-
bqml.bikes.num_trips`
OPTIONS
  (model_type='linear_reg',
    labels=['num_trips']) AS (
SELECT
    temp,
    visib,
    prcp,
    fog,
    rain_drizzle,
    snow_ice_pellets,
    num_trips
FROM
  `sara-bqml.bikes.trips_weather_nyc18`)
```

Use the model with ML.PREDICT

```sql
SELECT
  predicted_num_trips,
  num_trips
FROM
  ML.PREDICT(MODEL `sara-bqml.bikes.num_trips_nyc`,
    (
    SELECT
        temp,
        visib,
        prcp,
        fog,
        rain_drizzle,
        snow_ice_pellets
    FROM
    `sara-bqml.bikes.2017_nycbikes_test` ))
```
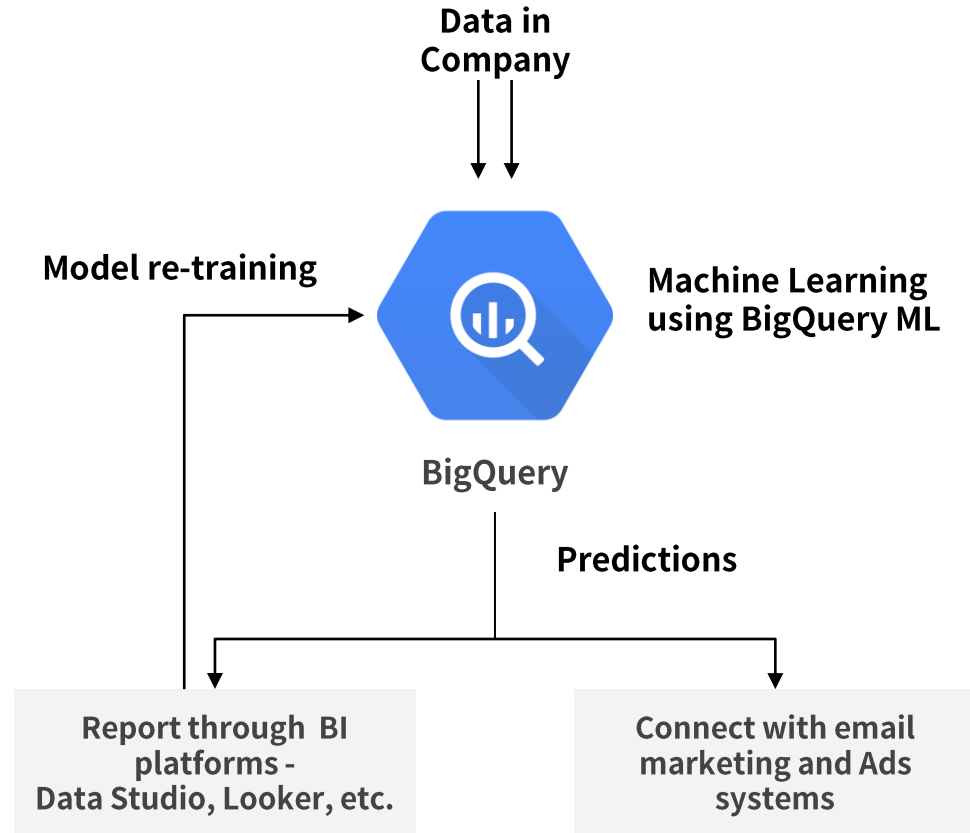
Evaluate the model with ML.EVALUATE
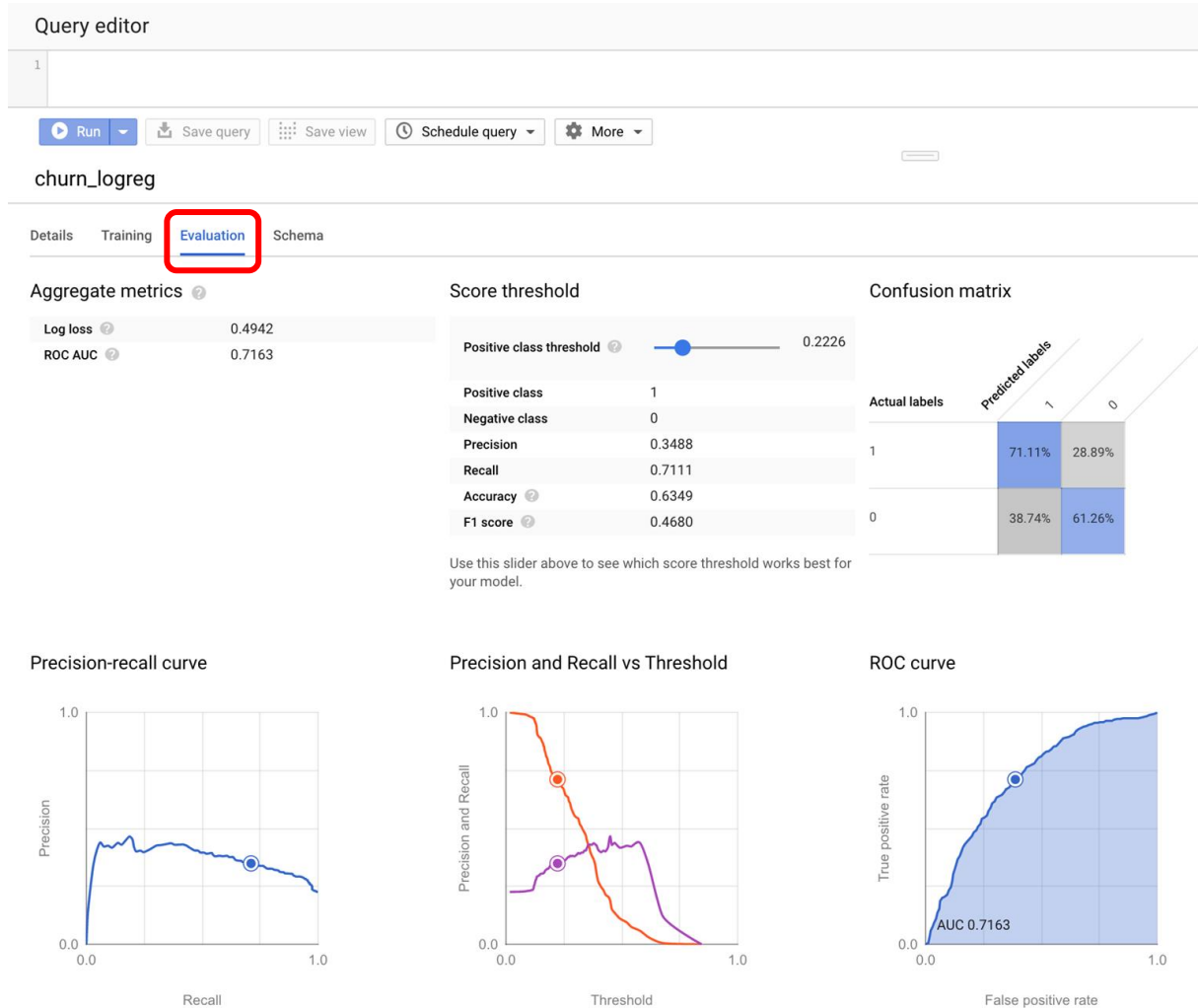
Evaluate the model with ML.EVALUATE

```
SELECT
  *
FROM
  ML.EVALUATE(
    MODEL [DATASET].churn_logreg
  )
```

Row	precision	recall	accuracy	f1_score	log_loss	roc_auc
1	0.4367816091954023	0.10555555555555556	0.7672521957340025	0.17002237136465326	0.49420724043767983	0.7162997002997002

Google Cloud

ML without data transfer

Data in Company

Model re-training

Machine Learning using BigQuery ML

BigQuery

Predictions

**Report through BI platforms -
Data Studio, Looker, etc.**

Connect with email marketing and Ads systems

BigQuery ML supported models and features

Classification
- Logistic regression
- DNN classifier (TensorFlow)
- XGBoost
- AutoML Tables
- Wide and Deep NNs

Regression
- Linear regression
- DNN regressor (TensorFlow)
- XGBoost
- AutoML Tables
- Wide and Deep NNs

Other Capabilities
- k-means clustering
- Time series forecasting (ARIMA+)
- Recommendation: Matrix factorization
- Anomaly Detection
- Principal Component Analysis
- Hyperparameter tuning using Cloud AI Vizier
- Explainable AI

ML Ops
- Importing TensorFlow models for batch prediction
- Exporting models from BigQuery ML for online prediction

MLOps tooling for BigQuery ML

Traditional data science

Data engineering

Sampling data for training

Programming (Python, Scala)

Deployment & productionisation

Extract value

Resource provisioning

Handling growing scale

Expensive, no-code platforms

BigQuery ML

Extract value
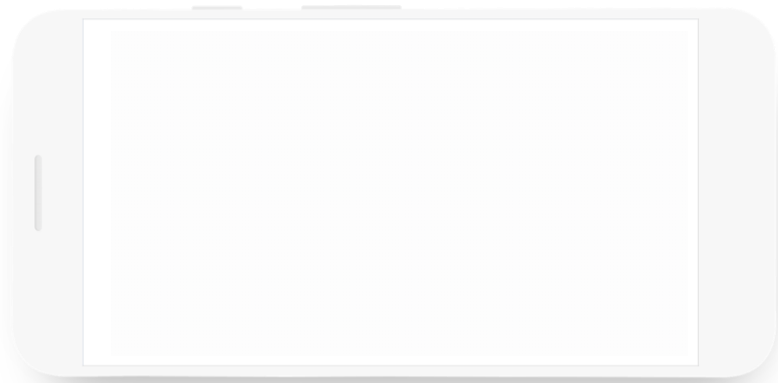
Manage production models using **Vertex AI Model Registry**

Use models in **BigQuery** or deploy behind a **REST endpoint**

Orchestrate data prep, training, and eval with **Vertex AI Pipelines**
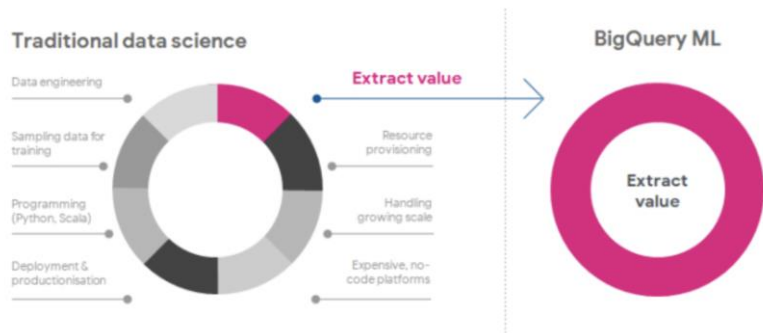
Manage production models using **Vertex AI Model Registry**

- **Register, organize, track, and version** your trained and deployed ML models supporting AutoML, BQML, and custom workflows

- **Store model metadata** and runtime dependencies for deployability.

- **Integrate with the model evaluation and deployment** capabilities and track online and offline evaluation metrics for the models. [on roadmap]

- **Govern the model launch process**: review, approve, release, and roll back. [on roadmap]

Use models in **BigQuery** or deploy behind a **REST endpoint**

Batch Predictions inside BigQuery
Model prediction inside of BigQuery queries

Good for data analysis, especially over large batches of data

Example
```
#standardSQL
SELECT
  *
FROM
  ML.PREDICT(MODEL `bqml_tutorial.penguins_model`,
  (
  SELECT * FROM
    `bigquery-public-data.ml_datasets.penguins`
  WHERE
    body_mass_g IS NOT NULL
    AND island = "Biscoe"))
```

Vertex Online prediction
Get model predictions for single examples over a REST API

Good for integration with applications, especially requiring low-latency

Leverage Vertex Model Monitoring
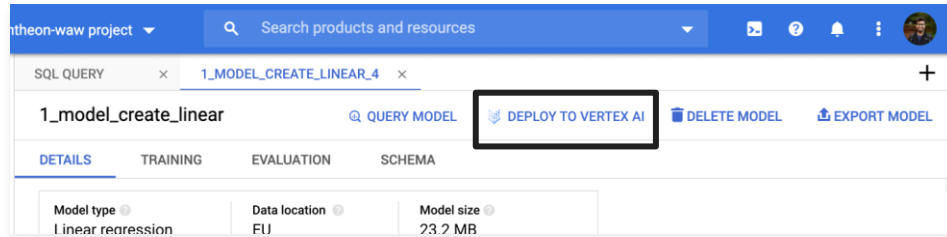
Orchestrate data prep, training, and eval with **Vertex AI Pipelines**

Easy to use Python SDKs
Build pipelines using Data Scientist friendly SDKs like TensorFlow Extended and Kubeflow Pipelines.

Automated, Scalable and Cost Effective
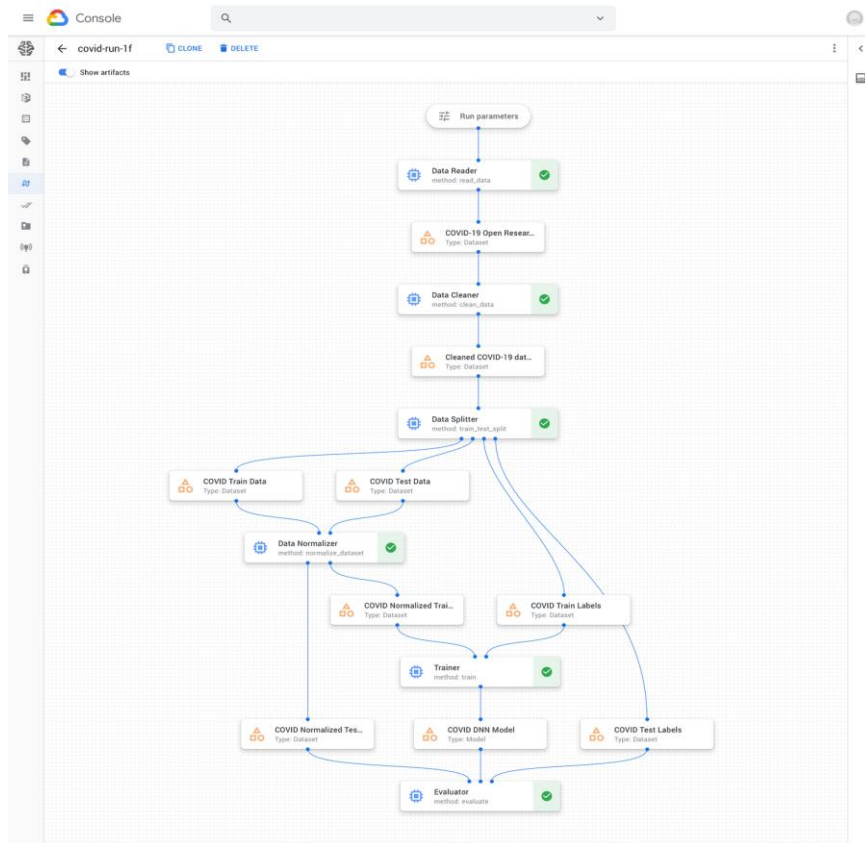Leverage GCPs managed services to build scalable pipelines. Pay for only the resource you use.

Streamlined MLOps
Automatic metadata tracking

BigQuery Model operators

- BigqueryQueryJobOp
- BigqueryCreateModelJobOp
- BigqueryExportModelJobOp
- BigqueryPredictModelJobOp
- BigqueryEvaluateModelJobOp

Limitless Users

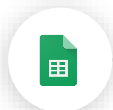By delivering data experiences for all data people

Deliver the right tools
to the right people.

Google's varied data
experiences make it
easy for any team to
drive analysis from its
leading DBMS platform.

Self-Service BI
Data Studio

EDW-Powered Spreadsheets
Google Sheets

Governed BI with
Semantic Model
Looker

Natural Language
Data QnA

Partner Ecosystem

MicroStrategy   Informatica   +ableau

Power BI   DOMO   sisense

Qlik Q   IBM Cognos Analytics

Introducing Connected Sheets

Combining the best of Big Query and the familiarity of Sheets to empower workforces and assist with:

- Unlocking big data insights
- Accelerating data workflows
- Improving cost-efficiency
- Strengthening data security

Sheets

Easy to use & shareable

Familiar interface

Light-weight analysis

Introducing
Connected Sheets

Analyze billions of rows of data in Sheets, without any need for specialized knowledge.

Big Query

Analyze petabytes of data

Complex queries

Increase time to insight

(1)

Single source of truth

Access billion rows of BigQuery data directly in Sheets without compromising security and performance.
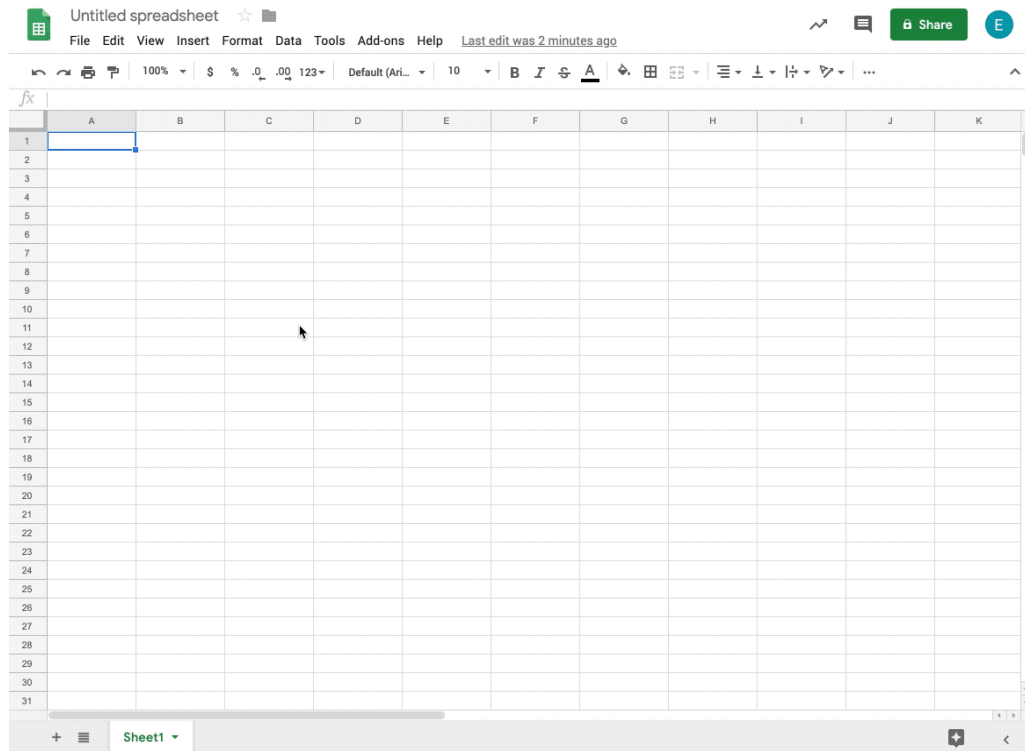
Do more with familiar tools

Unlock insights from your data using features you're already familiar with in Sheets, such as pivot tables, charts, and formulas.
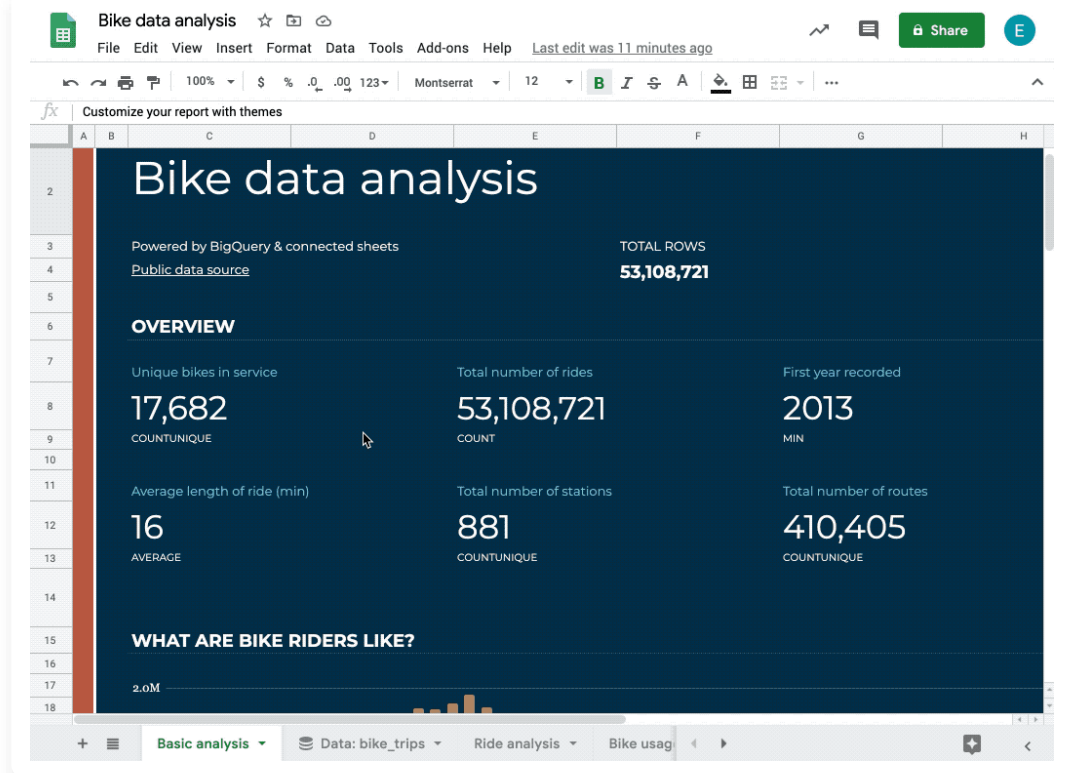
Fresh insights

Help ensure insights are based on the most up-to-date BigQuery data by setting up automatic data refreshes in Sheets.

Help secure collaboration

Share insights and collaborate securely with colleagues, analysts, or other stakeholders in a familiar spreadsheet interface.
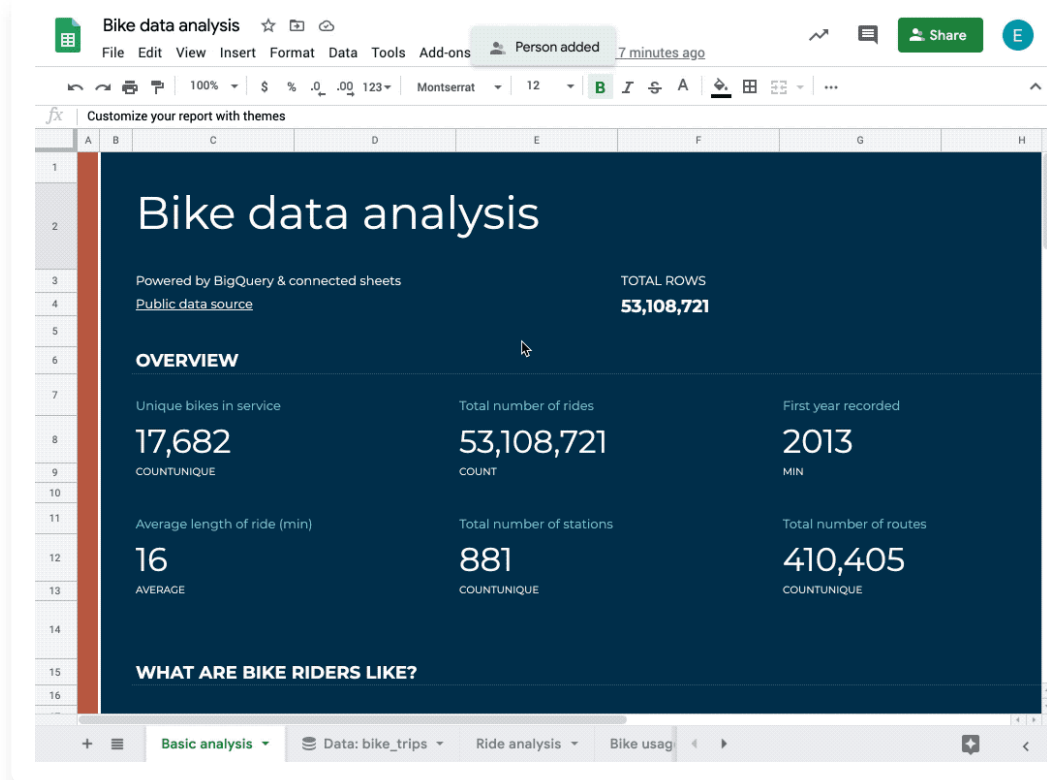
01   Intro

02 **Session 1**

Common Data Analytics Pipeline on GCP

1. Data Pipeline on GCP
2. Data Insight on GCP
3. Reference

1. Data Pipeline

BigQuery

2. Data Insight

3. Reference

BigQuery를 활용하는 이유
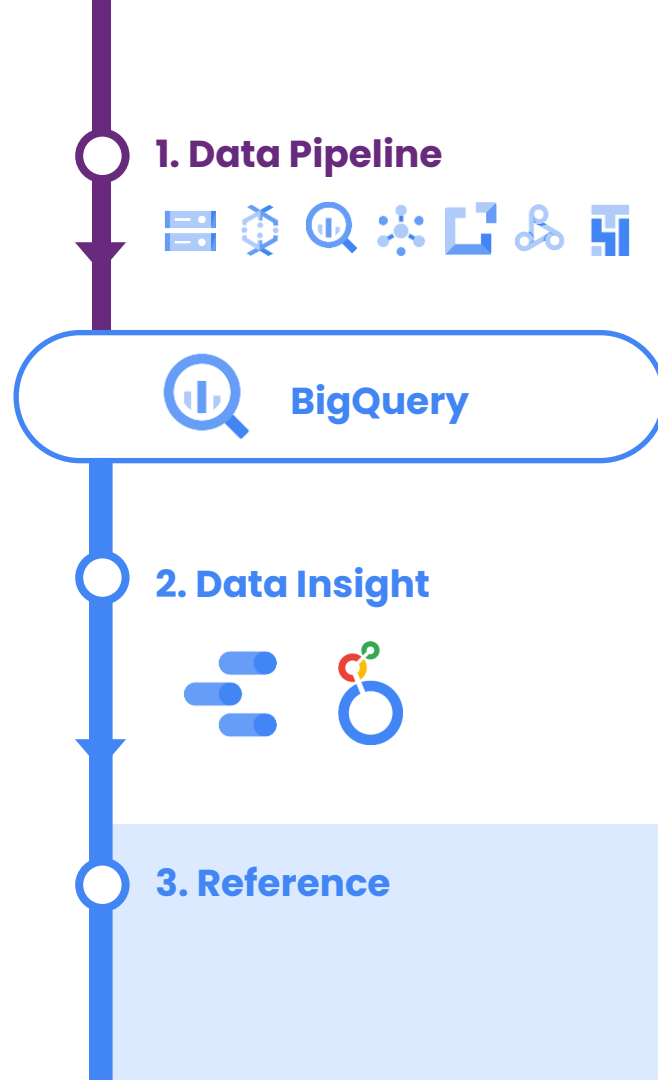
Enterprise Data Warehouse인 BigQuery의 장점을 소개합니다.

**PB 단위 까지 지원 가능한
Enterprise Data Warehouse**

**저렴한 비용
- Query 처리 용량 당 과금**

완전 관리형 & 서버리스

**ACL(Access Control List)
- Dataset / Table**

BigQuery

표준 SQL 지원

BQML(BigQuery ML)
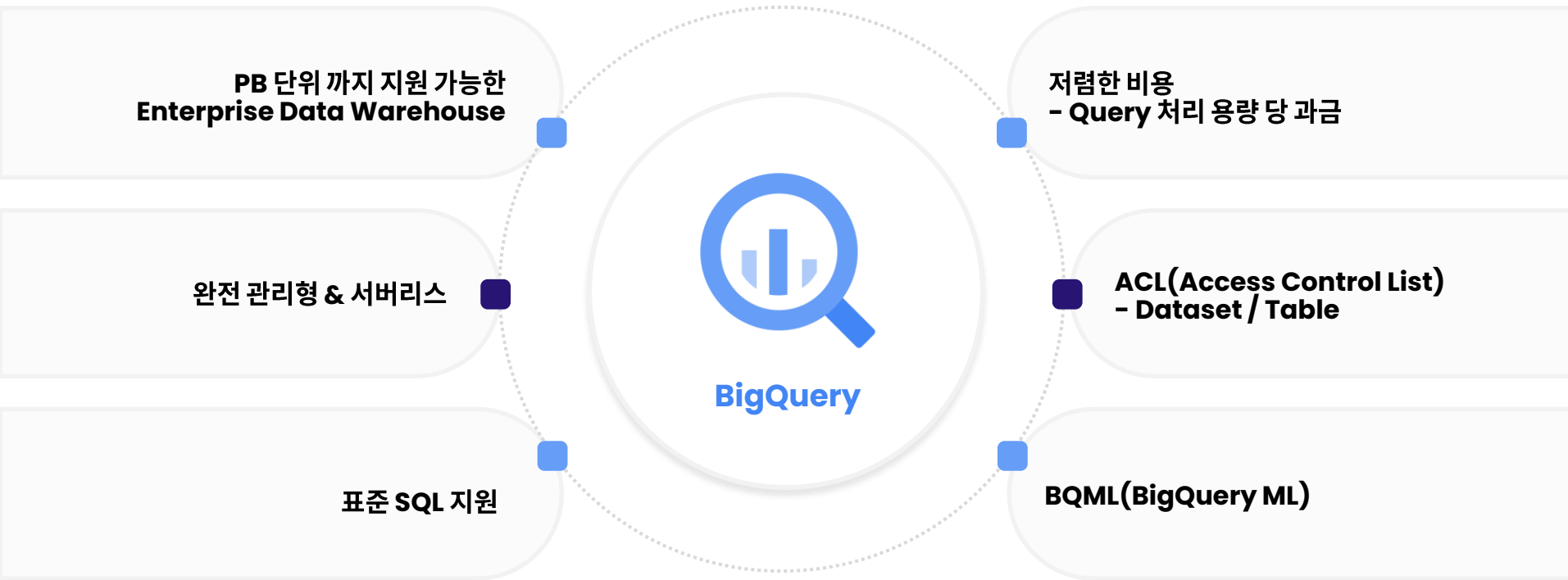
2-1 **Data Pipeline on GCP**

BigQuery

2-2 Data Insight on GCP

Google Cloud 기반의 일반적인 Data Pipeline 예시입니다.

Architecture: Streaming & Batch Data Analytics Pipeline

Data Pipeline on Google Cloud

Google Cloud 기반의 일반적인 Data Pipeline 예시입니다.

Architecture: Streaming & Batch Data Analytics Pipeline

Data Pipeline on Google Cloud

Google Cloud 기반의 일반적인 Data Pipeline 예시입니다.

Architecture: Streaming & Batch Data Analytics Pipeline

54

Data Processing Service in Google Cloud
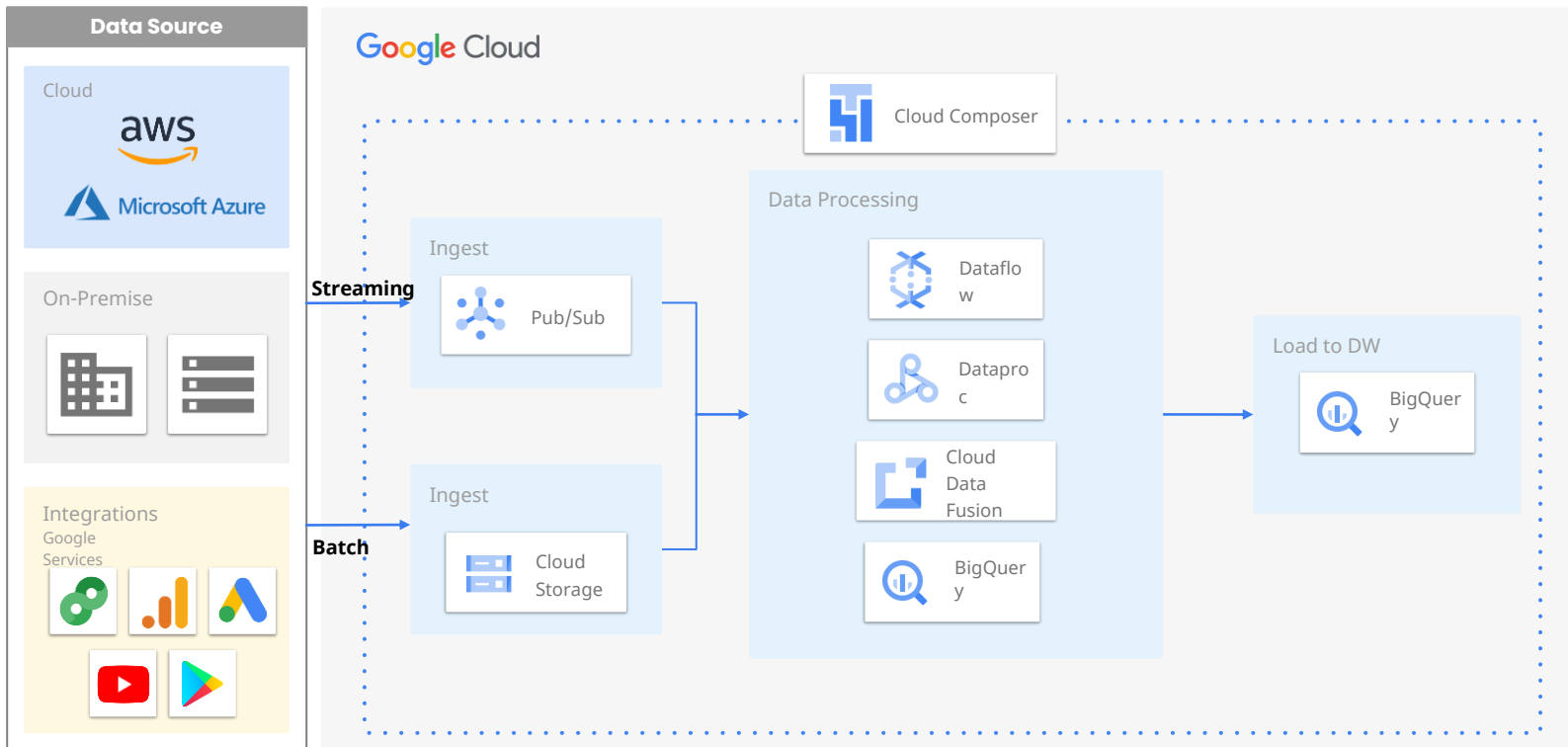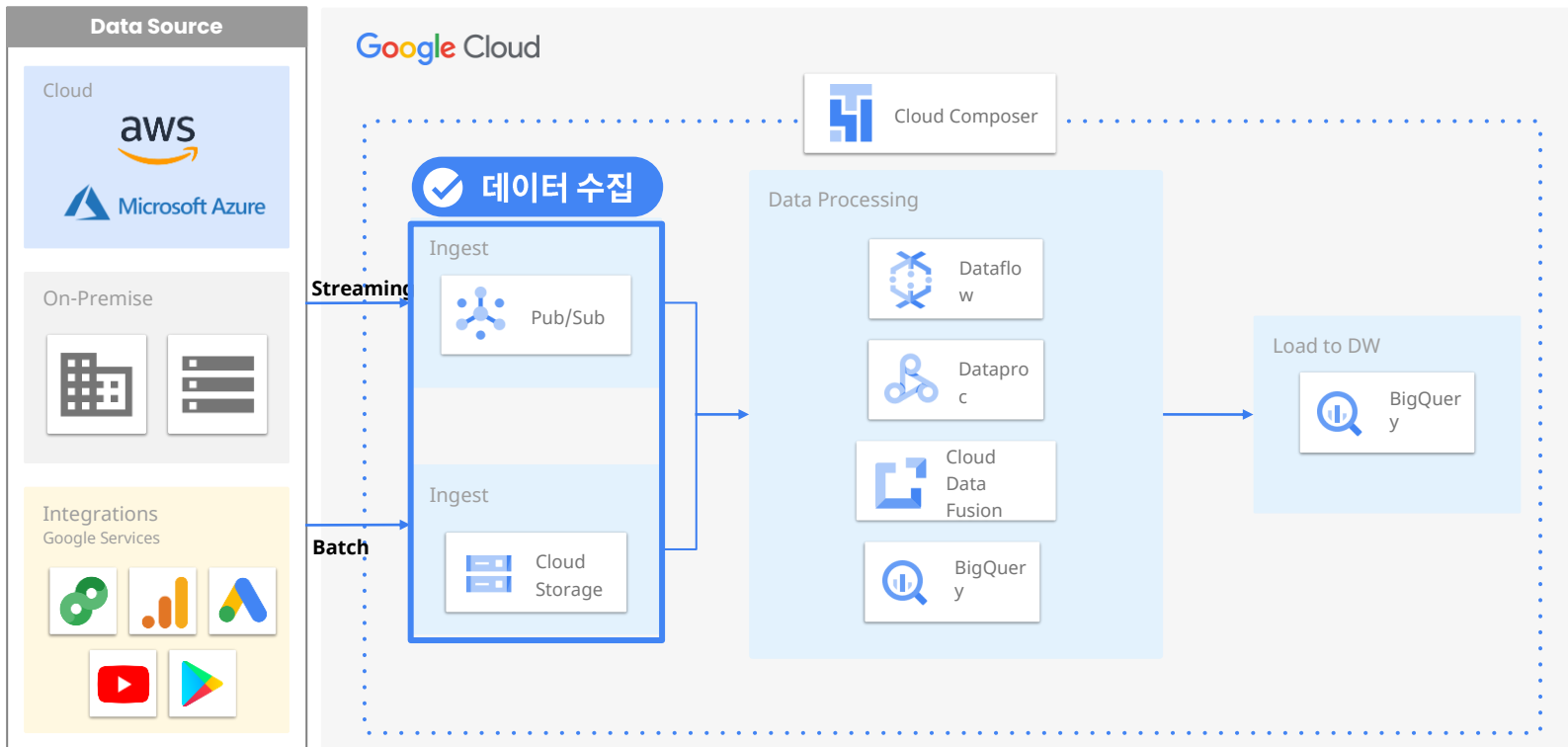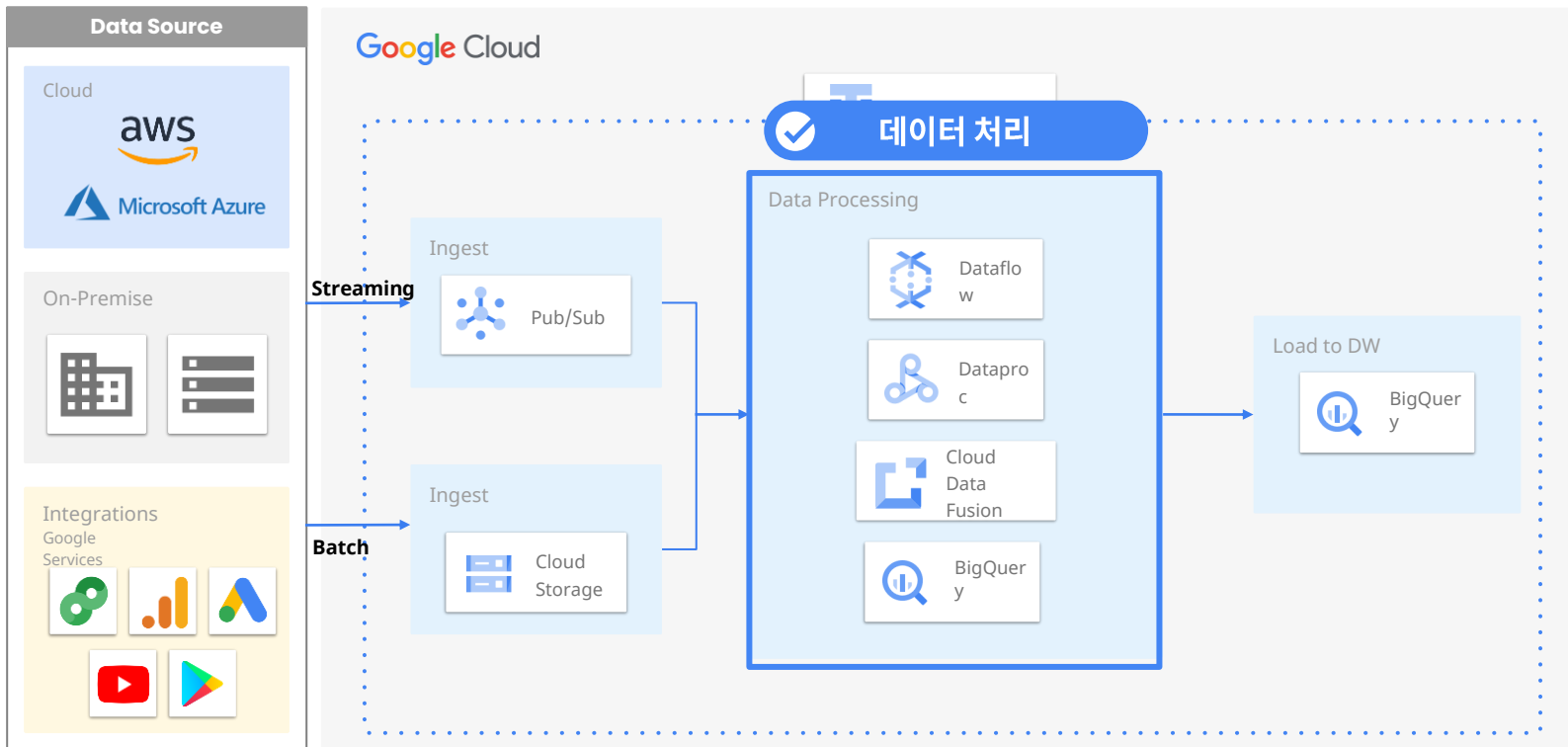
Google Cloud에서 데이터를 처리할 때 사용하는 서비스를 소개합니다.
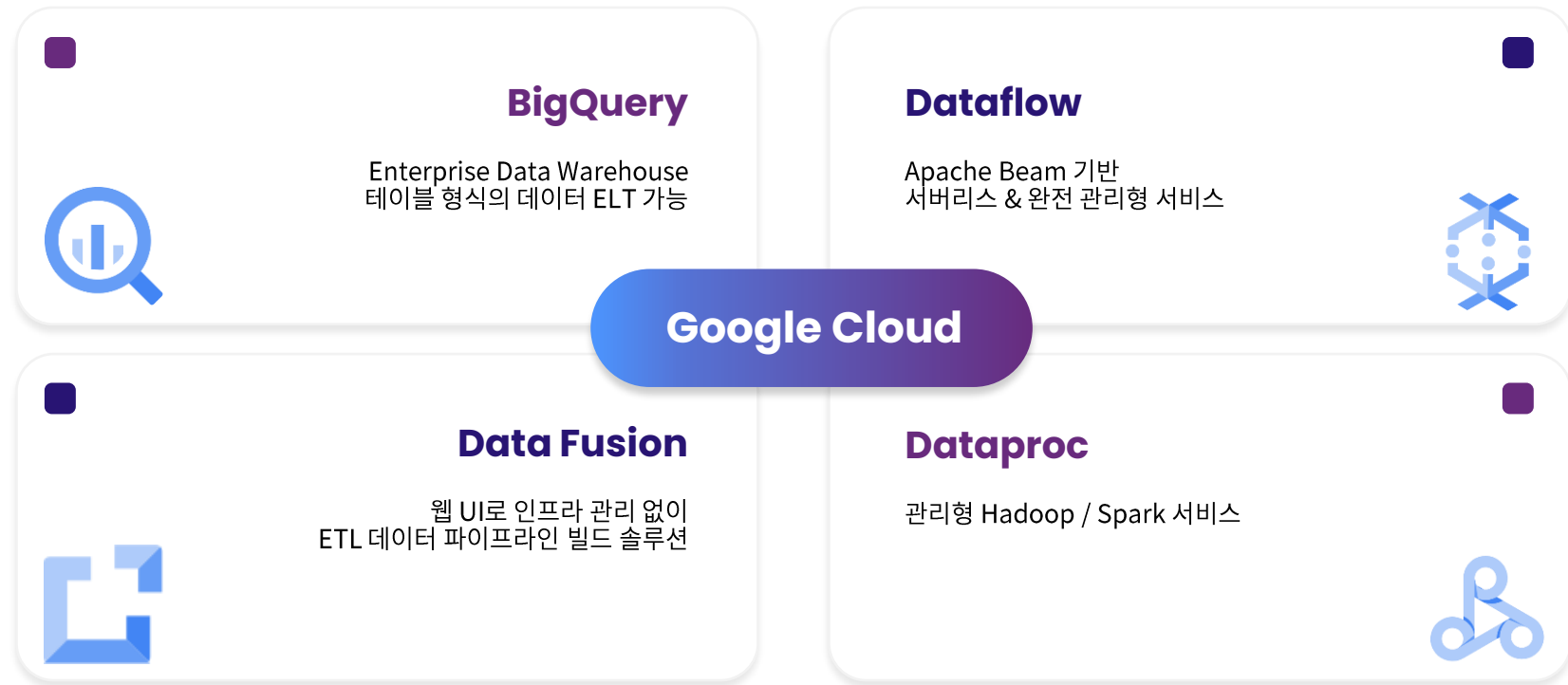
BigQuery

Enterprise Data Warehouse
테이블 형식의 데이터 ELT 가능

Dataflow

Apache Beam 기반
서버리스 & 완전 관리형 서비스

Google Cloud

Data Fusion

웹 UI로 인프라 관리 없이
ETL 데이터 파이프라인 빌드 솔루션

Dataproc

관리형 Hadoop / Spark 서비스

Google Cloud 기반의 일반적인 Data Pipeline 예시입니다.

Architecture: Streaming & Batch Data Analytics Pipeline
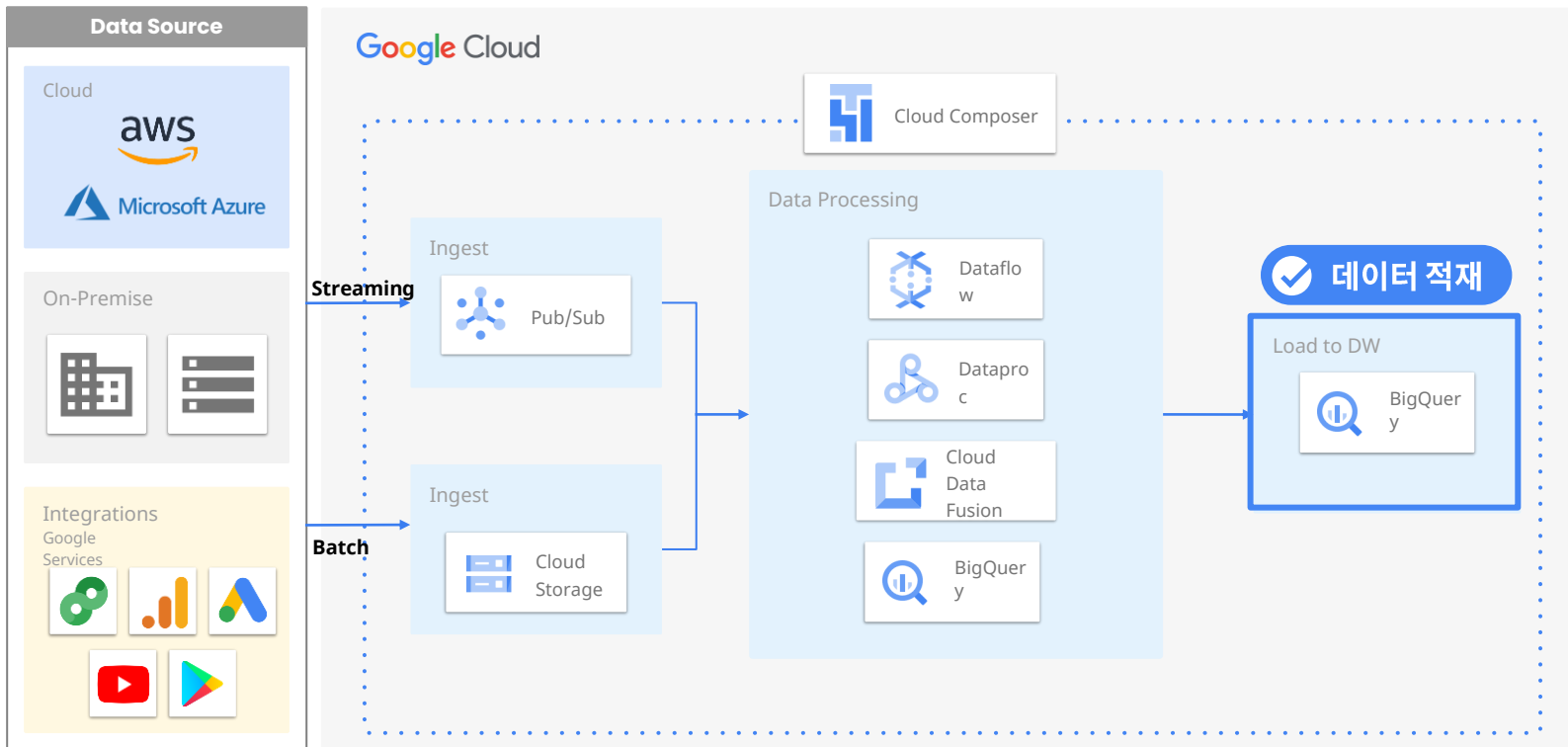
BigQuery

2-2 Data Insight with GCP

2-3 Reference

BigQuery를 Data Mart로 활용

BigQuery ACL(Access Control List)과 View 기능을 활용해 Data Mart로서 BigQuery를 활용합니다.

Using BigQuery as a Data Mart

BigQuery를 Data Mart로 활용
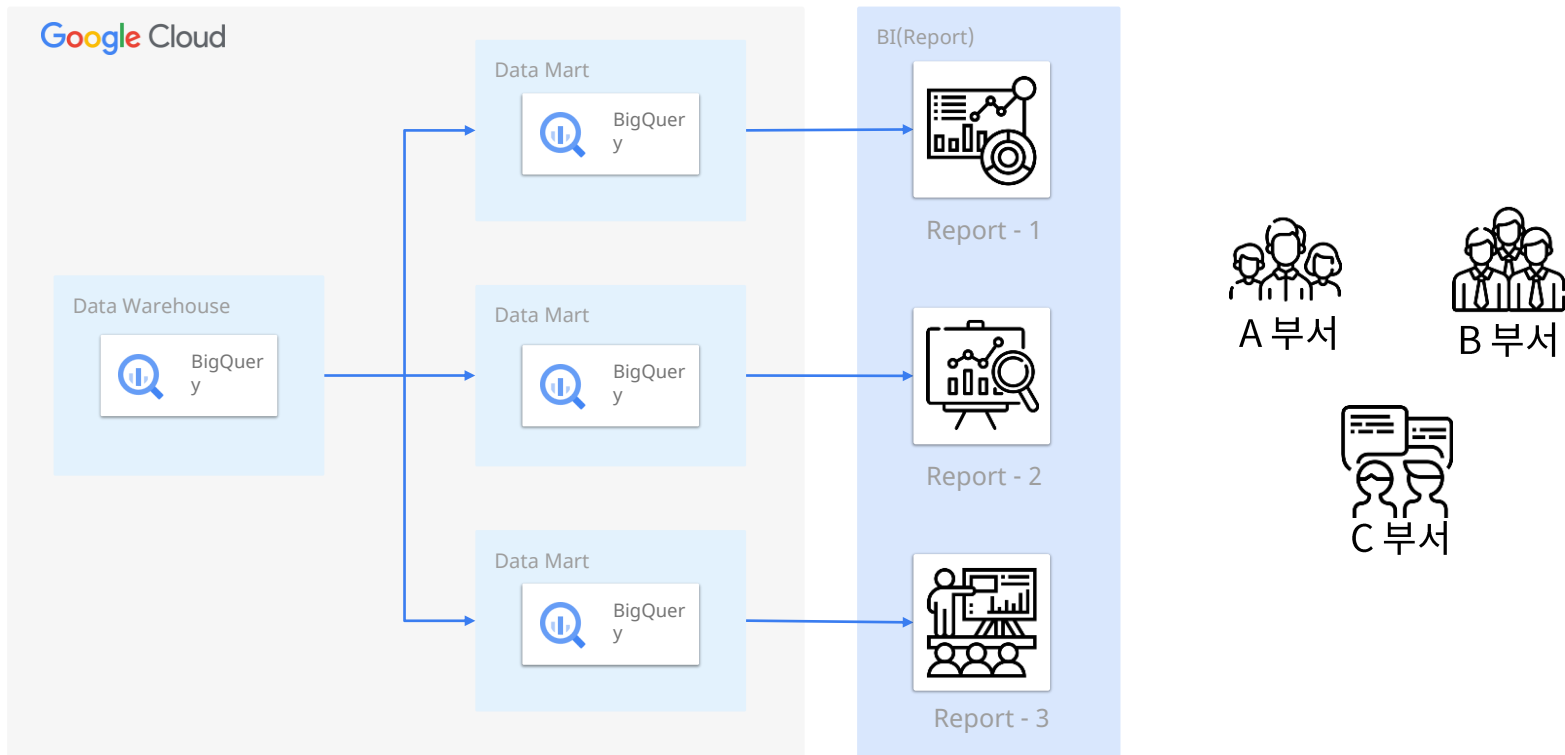
BigQuery ACL(Access Control List)과 View 기능을 활용해 Data Mart로서 BigQuery를 활용합니다.

Using BigQuery as a Data Mart
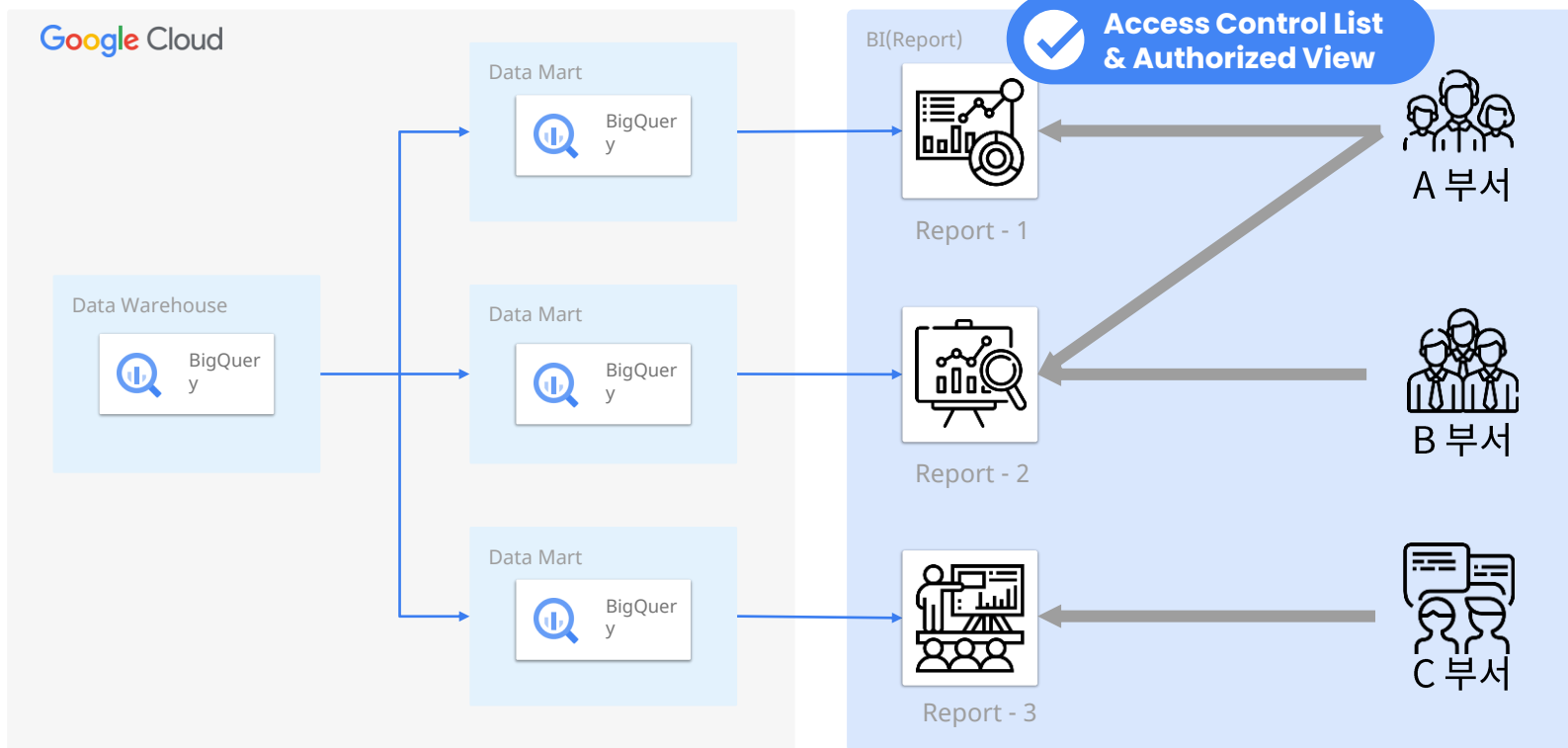
Google Cloud에서 데이터를 시각화할 때 주로 사용되는 두 가지 서비스를 소개합니다.
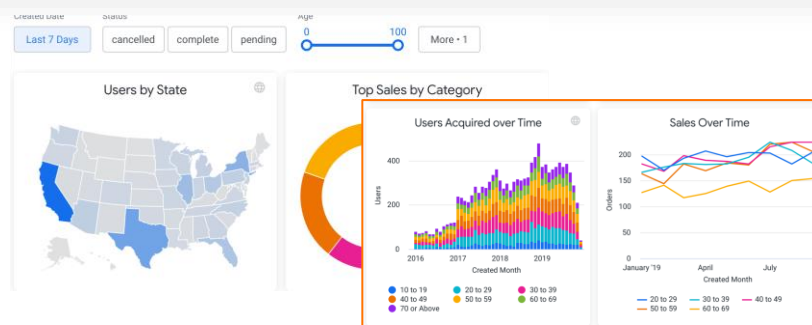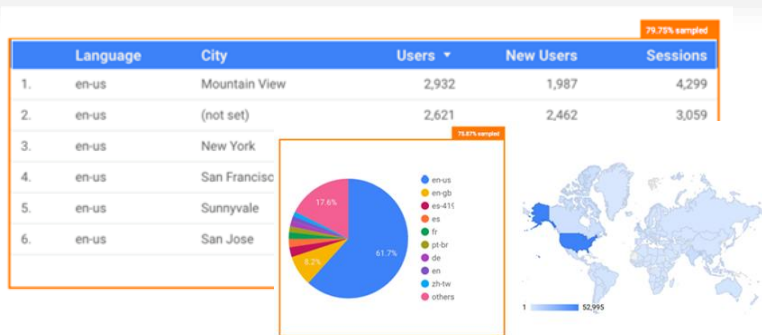
Data Studio

- **무료**
- 다양한 데이터 소스 지원
- **Code / Query 없이 대시보드 생성**

Looker

- 유료
- 접근 제어 - IP Allow List & ACL
- **LookML로 Custom Logic 설정**
- **개발/배포 환경 분리**
- **연결한 DataBase에 직접 Query**

2-3 Reference

유명 패션업계 제조업체 K사

Reference) As-Is Architecture

본 사례의 기존 분석계 시스템 아키텍쳐 입니다.

AS-IS Architecture

내부망

On-premises

On-premises

AWS
Amazon RDS

ODS

EDW(Sybase IQ)

WAS

BI(Report)

외부망

Google Cloud

GA360

Compute Engine

BigQuery

BigQuery

Reference) As-Is Architecture
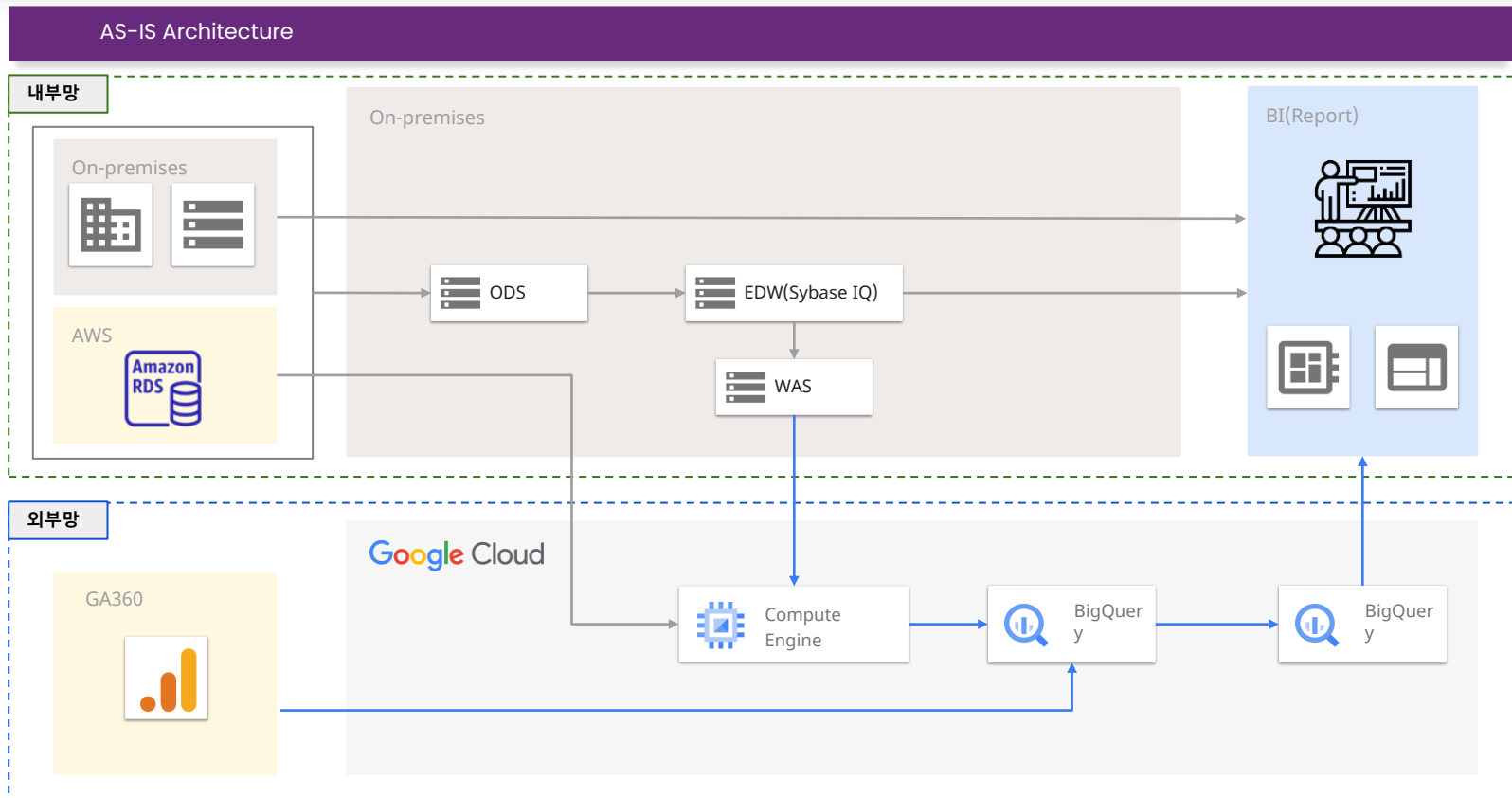
본 사례의 기존 분석계 시스템 아키텍쳐 입니다.

본 사례의 기존 분석계 시스템 아키텍쳐 입니다.

Reference) As-Is Architecture

본 사례의 기존 분석계 시스템 아키텍쳐 입니다.

AS-IS Architecture

내부망

On-premises

On-premises

AWS
Amazon RDS

ODS

EDW(Sybase IQ)

WAS

BI(Report)

외부망

Google Cloud

GA360

Compute Engine

BigQuery

BigQuery

Reference) As-Is Architecture

본 사례의 기존 분석계 시스템 아키텍쳐 입니다.

Reference) To-Be Architecture

Google Cloud를 기반으로 통합된 분석계 환경을 구축한 사례입니다.

TO – BE Architecture

AWS
Amazon RDS

On-premises

GA360

Google Cloud

Cloud Composer

Cloud VPN

Dataproc

BigQuery

Data Mart

BigQuery

BigQuery

67

Reference) To-Be Architecture

Google Cloud를 기반으로 통합된 분석계 환경을 구축한 사례입니다.

Reference) To-Be Architecture

Google Cloud를 기반으로 통합된 분석계 환경을 구축한 사례입니다.

TO – BE Architecture

Reference) To-Be Architecture

Google Cloud를 기반으로 통합된 분석계 환경을 구축한 사례입니다.

TO - BE Architecture

AWS

On-premises

GA360

Google Cloud

Cloud Composer

Cloud VPN

Dataproc

BigQuery

Data Mart

BigQuery

BigQuery

70

Google Cloud를 기반으로 통합된 분석계 환경을 구축한 사례입니다.

TO - BE Architecture

AWS
Amazon RDS

On-premises

GA360

Google Cloud

Cloud Composer

Cloud VPN

Dataproc

BigQuery

Data Mart

BigQuery

BigQuery

71

Reference) To-Be Architecture

Google Cloud를 기반으로 통합된 분석계 환경을 구축한 사례입니다.

TO - BE Architecture

AWS
Amazon RDS

On-premises

GA360

Google Cloud

Cloud Composer

Cloud VPN

Dataproc

BigQuery

Data Mart

BigQuery

BigQuery

72

Reference) 도입 효과

본 사례의 도입 효과입니다.

GCP 기반 통합 분석 환경
(BigQuery)

운영 및 분석
비용 절감

도입효과

Private 환경 구축

파이프라인
간소화

Reference) 도입 효과

본 사례의 도입 효과입니다.

GCP 기반 통합 분석 환경
(BigQuery)

운영 및 분석
비용 절감

도입효과

Private 환경 구축

파이프라인
간소화

Reference) 도입 효과

본 사례의 도입 효과입니다.

GCP 기반 통합 분석 환경
(BigQuery)

운영 및 분석
비용 절감

도입효과

Private 환경 구축

파이프라인
간소화

Reference) 도입 효과

본 사례의 도입 효과입니다.

GCP 기반 통합 분석 환경
(BigQuery)

운영 및 분석
비용 절감

도입효과

Private 환경 구축

**파이프라인
간소화**

Reference) 도입 효과

본 사례의 도입 효과입니다.

GCP 기반 통합 분석 환경
(BigQuery)

운영 및 분석
비용 절감

도입효과

Private 환경 구축

파이프라인
간소화

Reference) Next Step

본 프로젝트 진행 이후 단계에서의 활용 및 기대 효과입니다.

Google Cloud   **BigQuery**

➕
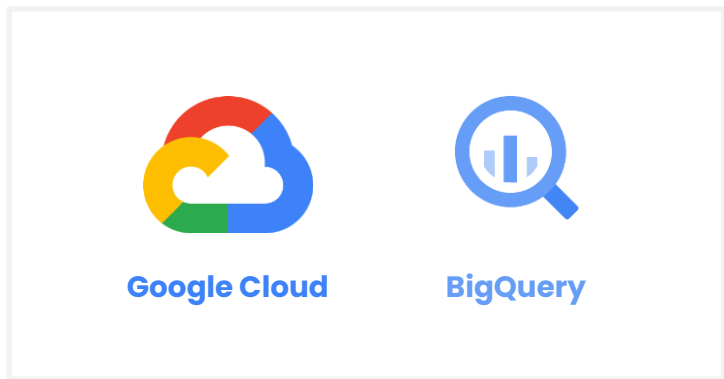
■ Next Step with BigQuery ML

BQML을 활용한 이점
- BigQuery에 저장되어 있는 데이터를 별도 이동없이 그대로 활용해 ML 모델 빌드
- SQL이 익숙한 분석 실무자들이 직접 ML 모델 빌드 및 실행

1 매출 증가	2 재고 관리	3 우수 고객 확보
• 추천 시스템 도입 -> 매출 증대	• 수요 예측 -> 재고 관리 효율 증가	• 고객 Segmentation -> 타겟 마케팅, 우수 고객 확보

03 **Session 2**

Integration BigQuery with GWS & GA
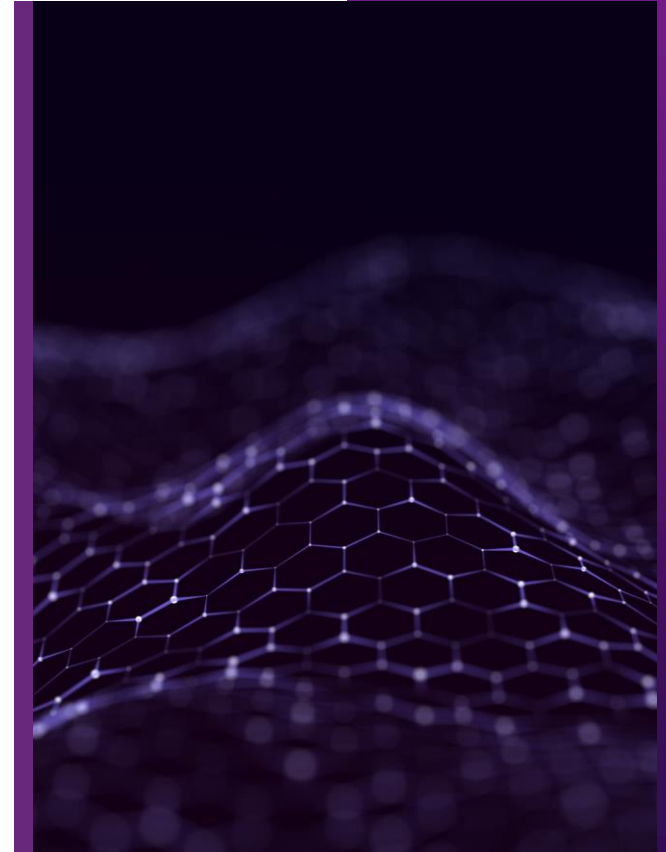
03 Session 2

Integration BigQuery with GWS & GA

3-1. BigQuery with Various Data
3-2. Reference
3-3. Demo

1. Google Services

BigQuery

2. Reference

3. Demo
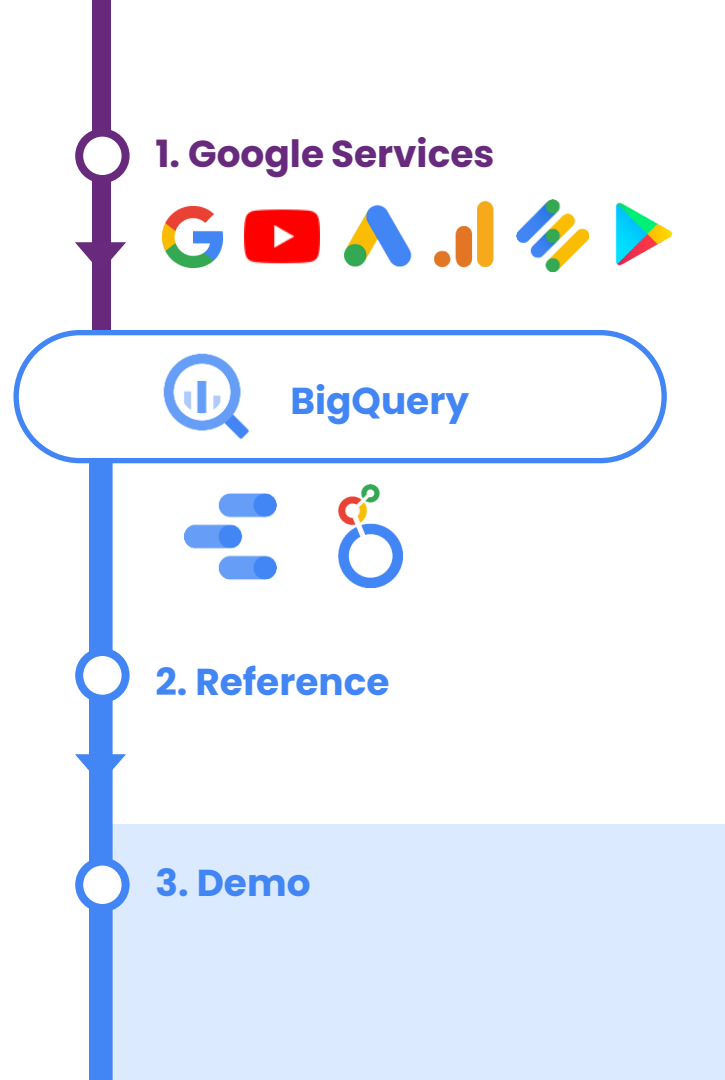
3-1 BigQuery with Various Data

BigQuery

3-2 Reference

Google Cloud 외부 리소스와 연결 가능한 BigQuery

BigQuery는 구글 내 서비스부터 3rd Party 제품까지 다양한 데이터 소스를 지원합니다.

Google Workspace

- GWS 리소스 활동 감사
- GWS 서비스 사용량 수집
- 조직 자산의 보안관리 지원

Google Analytics

- 사용자의 접속정보 및 활동기록 수집
- Web & Application 모두 지원

BigQuery

Additional Resources

- Youtube 채널 사용자 활동 로그,
  Play Store 예상매출 및 순익 등
  사용자의 컨텐츠 활동기록

Google Spreadsheet

- 데이터 분석을 위한 단일소스 저장소
- 익숙한 인터페이스를 통한 데이터 관리
- 파일 링크만으로 BigQuery에서 쿼리 수행 가능

Google Workspace and BigQuery

BigQuery에 조직 내 리소스 사용에 대한 감사로그를 저장하고, GWS 로그를 분석 할 수 있습니다.

GWS 활동데이터 수집

자산 활동 분석

맞춤형 리소스 보고서

Google Workspace

- Gmail, Chat, Drive , Chrome등 GWS 자산 활동 집계
- GWS 서비스 사용로그 집계

BigQuery

- 조직 내 트렌드 및 개괄적인 사용 정보 분석
- 잠재적 보안위협을 지닌 의심스러운 활동 탐지

Data Studio

- Data Studio 등의 분석도구 사용한 맞춤형 보고서
- GCP 리소스 활용한 log 기반 Alert 구성
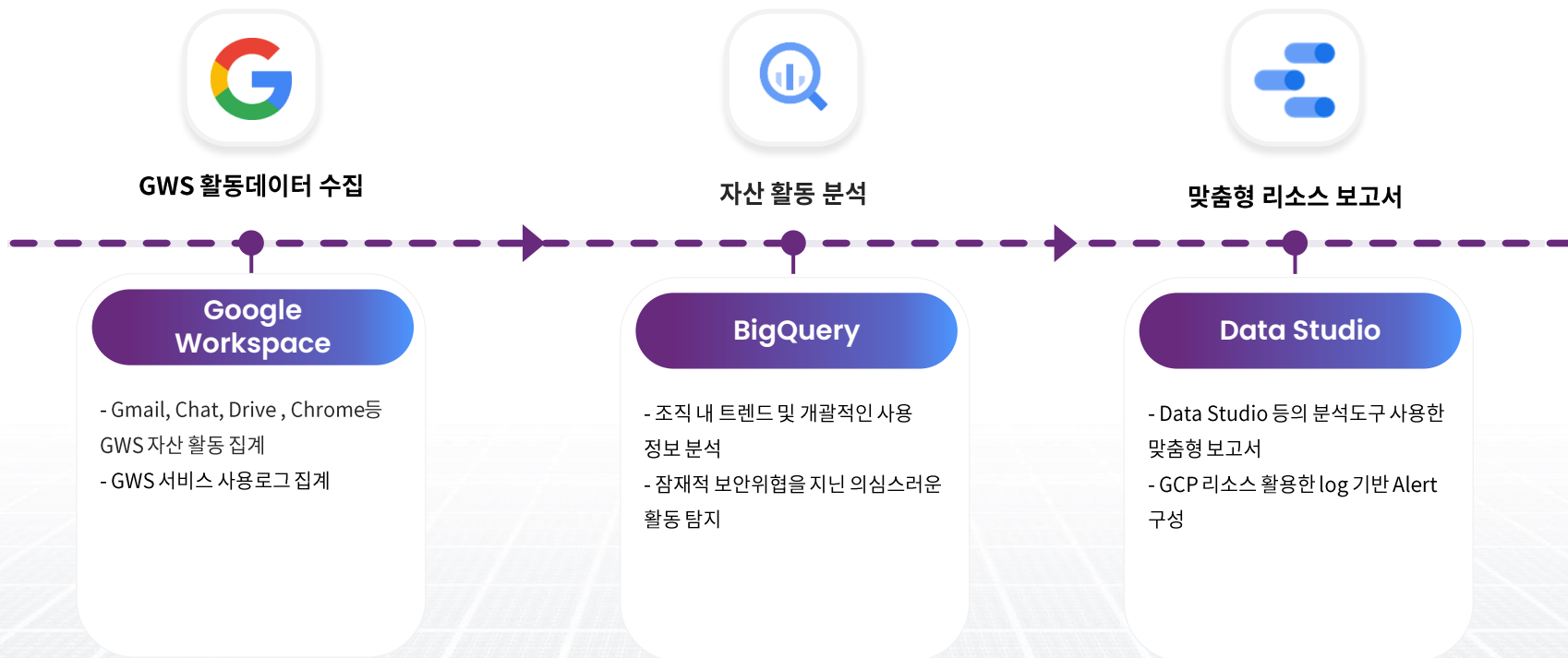
지원버전: Enterprise. Education Standard, Education Plus

Google Analytics를 통해 수집한 고객의 행동데이터를 BigQuery와 연결하여 고객에 대한 이해도를 높일 수 있습니다.

사용자 행동데이터 수집

패턴 분석

미래 예측

전략 수립

Google Analytics	**BigQuery**	**BigQuery ML**	**Marketing**

- 사용자의 유입경로 및 트래픽 수집
- 정형, 비정형 데이터 측정
	- 고객 터치포인트 분석
	- 웹사이트 성능 분석
	- 특정 사용자그룹의
	  중복 행동 패턴 분석
		- 미래 동향 예측
		- 고객전환경로 및
		  기여분석 모델 비교 평가
			- 마케팅 ROI를 최적화하여 선제적
			마케팅 전략 수립
			- 개별 고객 Lifetime Value
			  최적화
			- 퍼포먼스 마케팅

GA Standard의 일일 BigQuery Export 한도: 이벤트 100만 개

Google Services and BigQuery

BigQuery Data Transfer Service를 통해 다양한 Google 제품과 3rd party 제품을 분석할 수 있습니다.

로그 취합

패턴 분석

맞춤형 리소스 보고서

컨텐츠 전략 수립

Google Services	BigQuery	Data Studio	Marketing
- Youtube, Play Store, Ads 등 다양한 Google Service 지원 - BigQuery Data Transfer 서비스 통한 3rd party 지원	- 매입, 매출 분석 - 사용자 활동 기반 컨텐츠 퍼포먼스 분석 - 사용량 추이 분석	- 재무보고서 - 사용자 통계 보고서 - 실적보고서 - 잠재고객 분석 보고서	- 컨텐츠 사용자 동향 파악 - 컨텐츠 영향도 파악

Spreadsheet와 BigQuery

정형데이터 포맷인 Spreadsheet는 다양한 방법으로 BigQuery와 함께 사용할 수 있습니다.

Connected Sheet

- 익숙한 인터페이스에서 다른 이해관계자와 공동작업 수행
- 별도 내보내기 작업 없이 데이터 출처 확인 가능
- 보고 및 대시보드 Workflow 간소화
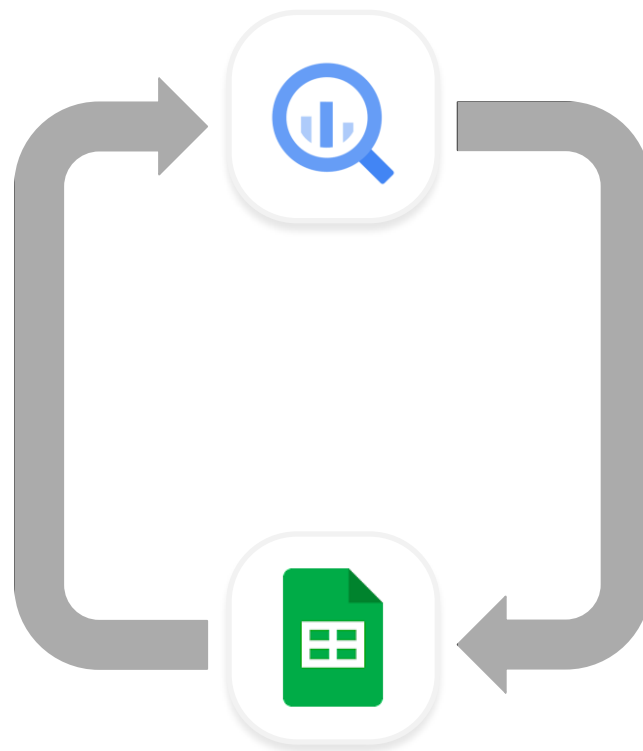
External Table

- 원본 데이터 유지
- 쿼리 대상 범위 지정히여 사용 기능
- 쿼리 대상 데이터가 BigQuery 외부에 있어 속도 저하
  ⇒ 일회성 프로세스에 유용

87

3-2 Reference

국내 환경 제조업체 E사

3-3 Demo

Reference

Spreadsheet와 BigQuery를 결합하여 데이터를 수집, ETL, 저장 및 분석하는 환경을 구축한 사례입니다.

Architecture
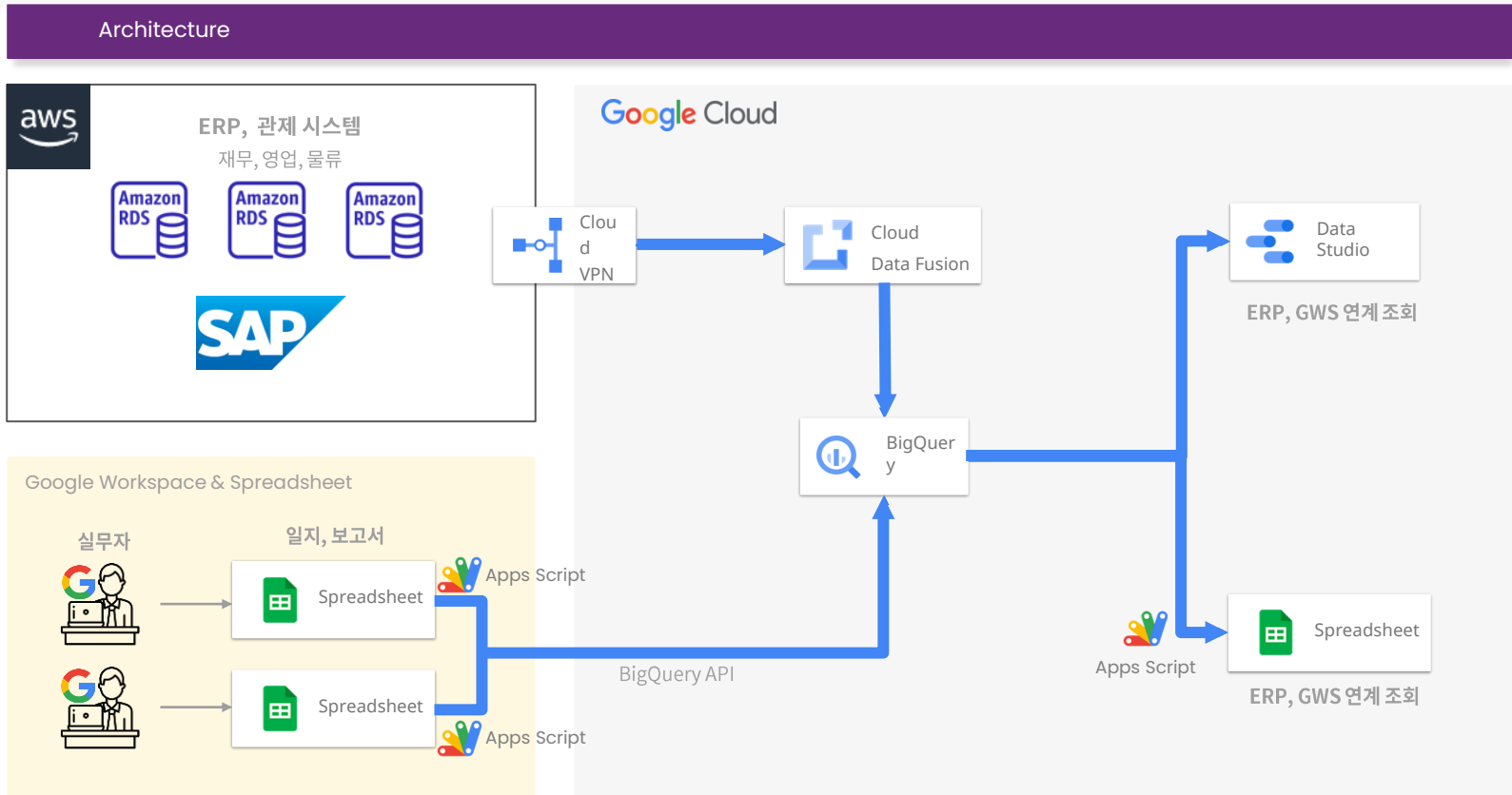
Reference) 도입효과

본 사례의 도입 효과입니다.

GWS로
사용자를 인증하는
Private 환경

운영 및 분석비용
절감

도입효과

BigQuery에 통합된
실시간 분석 환경 구축

시각화 된 대시보드로
주요 KPI 분석 가능

3-3 Demo

04 Summary

Summary

BigQuery를 사용하는 이유

Thank you!

Cloocus