



Entire Data Lifecycle @Google Cloud

Google Cloud / Partner Engineer 김현정

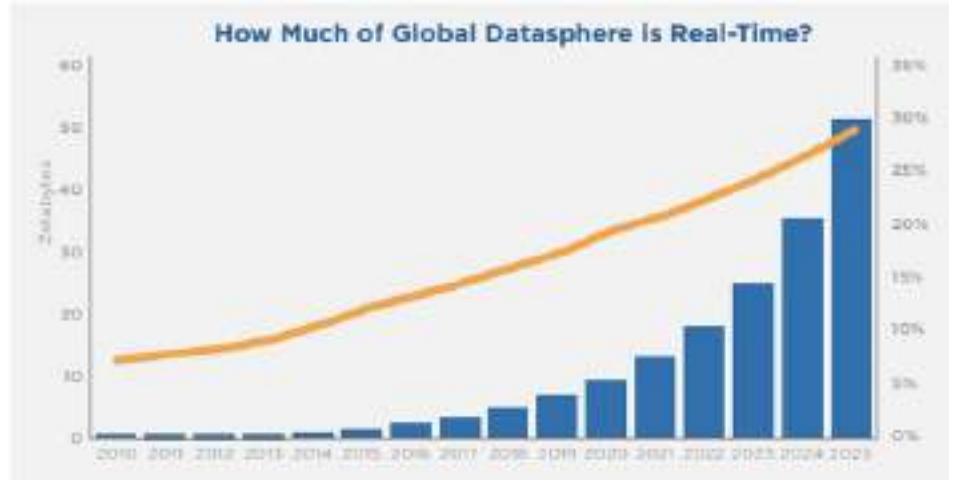




By 2025, more than a quarter of data created in the global datasphere will be real time in nature

[IDC](#)

5년 후 글로벌 데이터량 175 ZB,
Real-Time 데이터의 급속한 성장



Data analytics remains **untapped**

데이터 분석 사용률 30%

69%

69% of companies report that they have not created a data-driven organization

71%

And 71% report that they have yet to forge a data culture

Challenges with Big Data Projects

빅데이터는 *Big Problems*?!

1 복잡하고 어려운 시스템
유지/관리

2 다양한 데이터 소스 수집/저장

3 급속 성장하는 데이터량

4 보유 데이터에 대한
가치발굴

5 데이터 수집부터
의사결정까지의 소요시간

6 빅데이터의 통합/혁신 방안
마련

7 관련 조직간의 협업

8 데이터 보안 유지

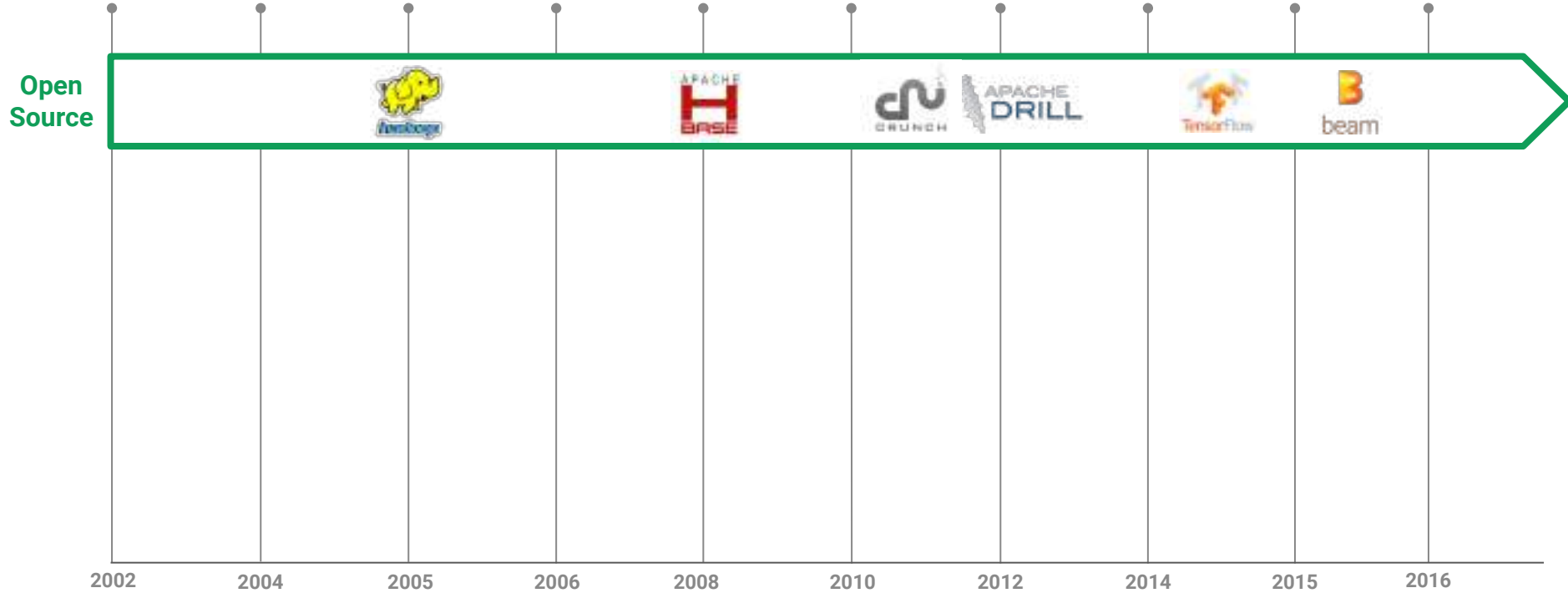
9 시스템 안정성 유지

Google faced many of those challenges

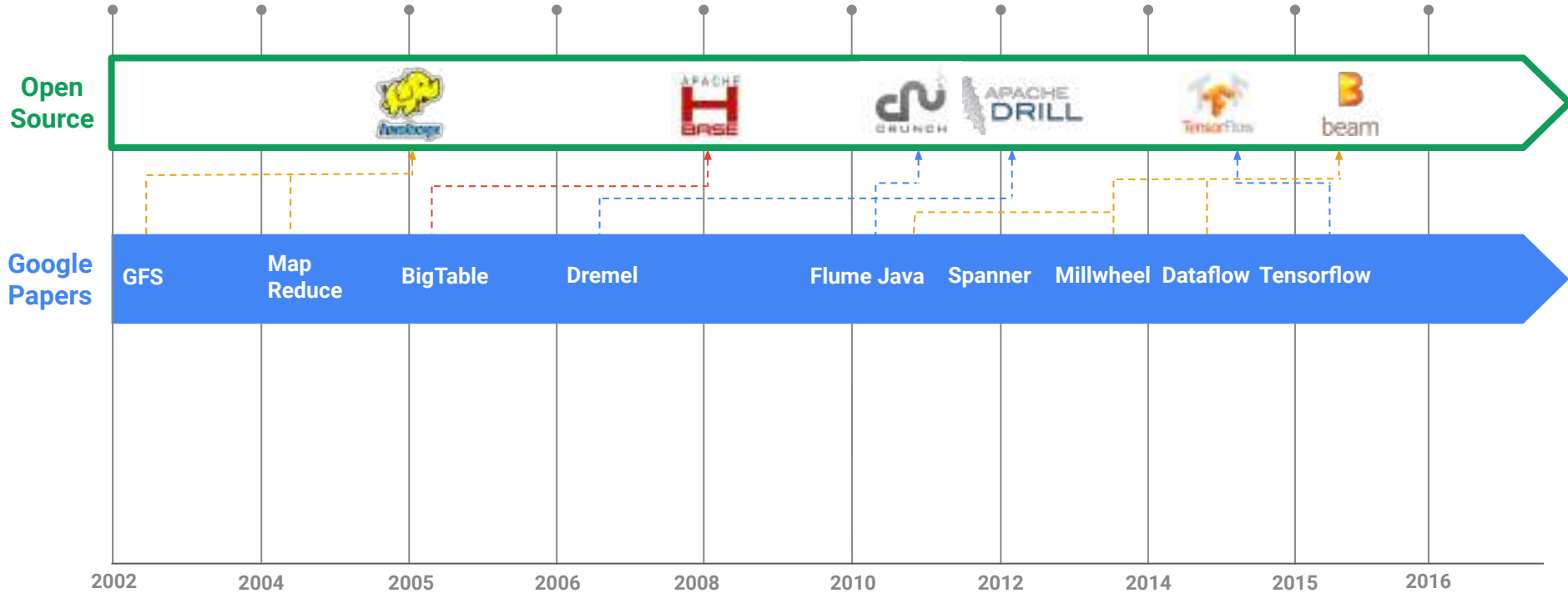
구글은 이러한 고민을 어떻게 해결했을까요?



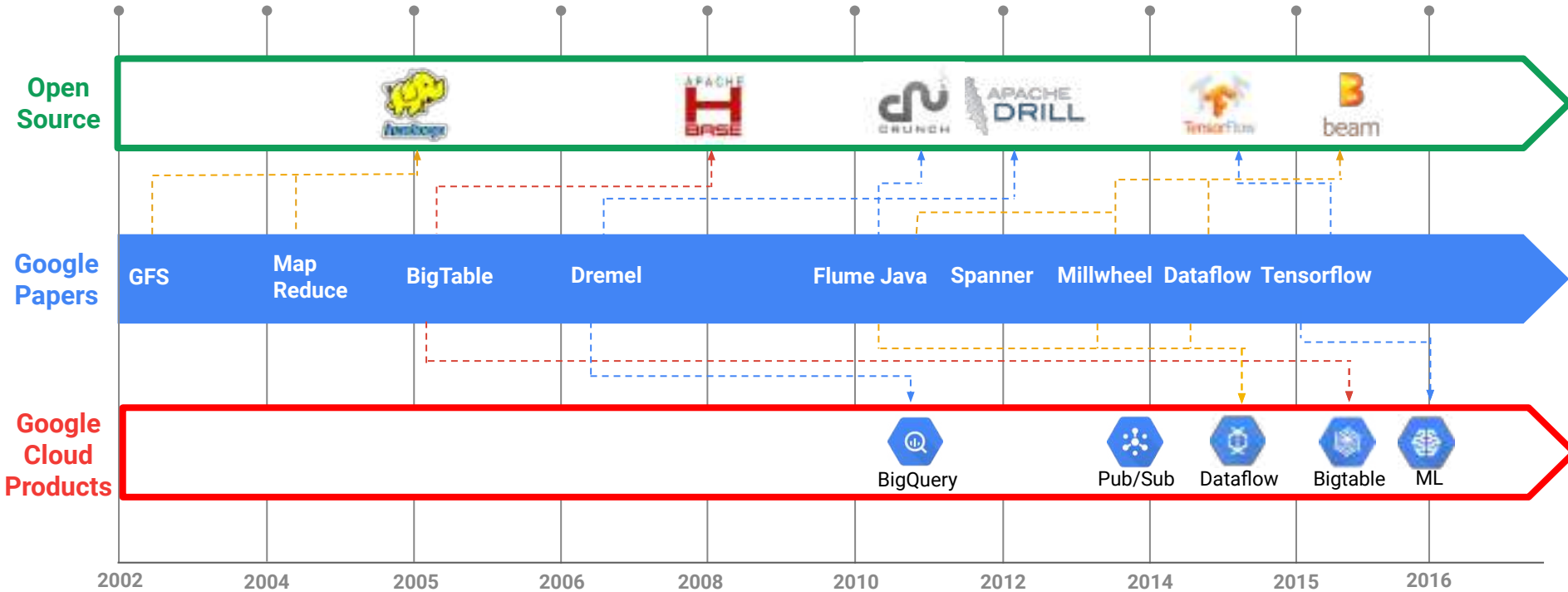
Technologies Inside of Google



Technologies Inside of Google



Technologies Inside of Google



Google is a **world leader** in Data Analytics & AI/ML to real-world situations, inside and outside of Google.

구글 클라우드 서비스를 통해 구글 인사이트 기술을 여러분께
제공드립니다.



Search
Search Ranking
Speech Recognition



Android
Keyboard and
Speech Input



Play
App Recommendations
Game Developer Experience



Gmail
Smart Reply
Spam Classification



Drive
Intelligence in Apps



Chrome
Search by Image



Photos
Photos Search



YouTube
Video Recommendations
Better Thumbnails



Maps
Street View Image
Parsing Local Search



Translate
Text, Graphic and Speech
Translations



Cardboard
Smart Stitching



Ads
Richer Text Ads
Automated Bidding



Self Driving Car
1.5MM miles driven



Data Center
Reduced cooling energy usage

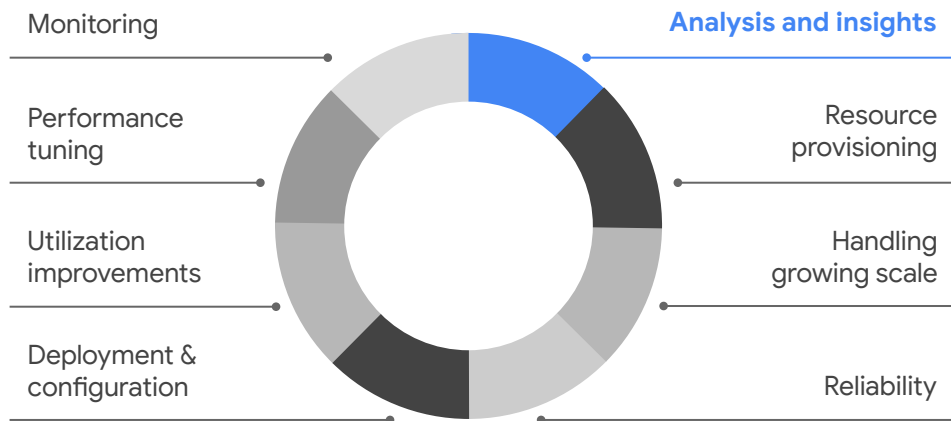


Alpha Go
First AI to beat a world Go
champion (2016)

Serverless data analytics

서버리스 데이터 분석을 통해 모든 리소스를 분석에 집중하세요

The traditional data analytics platform



The serverless data analytics model



Key Solutions Powered by Google Cloud

실시간 데이터를 활용하여 미래의 비즈니스를 미리 예측하세요



Cloud Data Warehouse

데이터웨어하우스
현대화

Modern Data
Warehousing which
builds foundation for AI



Streaming Data Analytics

실시간 데이터
분석 기반
인사이트

Process Streaming
Data along with batch
data to generate
real-time insights

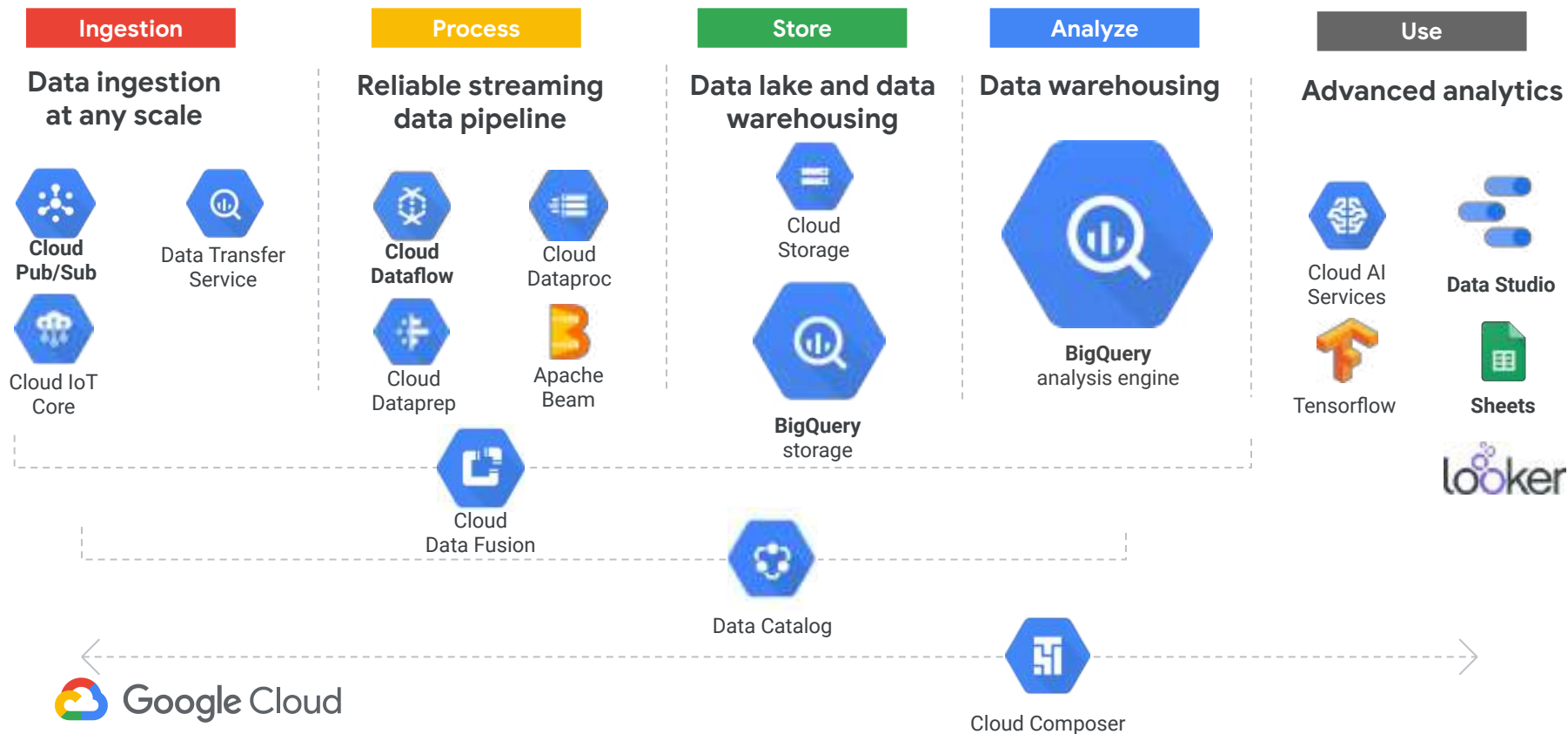


Predictive Analytics / ML

고객요구 사전예측
/ ML 자동화

Anticipate customer
needs and automate
delivery with Machine
Intelligence

Modern Data Warehouse on Google Cloud



Google BigQuery

BigQuery만의 차별점

Fully managed and **Serverless**
for maximum agility and scale

High-speed, in-memory
BI Engine for faster reporting and analysis

Built-in **ML** for out-of-the-box
predictive insights



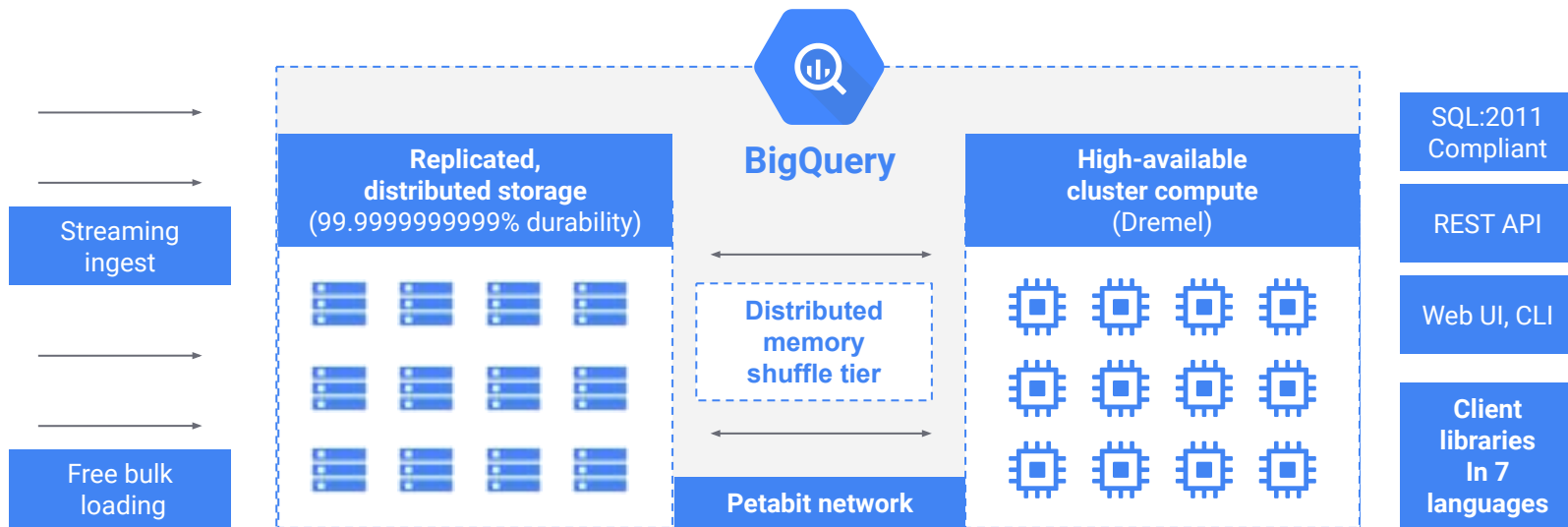
Petabyte-scale storage and
queries

Encrypted, durable and
highly available

Real-time insights from streaming data

BigQuery: Architecture

컴퓨팅과 스토리지 노드의 분리 구조 - 유연한 동적 할당 가능



BigQuery Pricing

작업	비용	참고사항
저장	GB당 \$0.02/ 월	매일 최초 10GB는 무료입니다. 자세한 내용은 저장소 가격 책정 을 참조하세요.
		1TB 당 월 20\$
장기 저장	GB당 \$0.01/ 월	장기 저장소 가격 책정 을 참조하세요.
		장기 저장(90일 이상) 1TB 당 월 10\$
스트리밍 삽입	GB당 \$0.05	저장소 가격 책정 을 참조하세요.
쿼리	TB당 \$5	매일 최초 1TB는 무료입니다. 자세한 내용은 주문형 가격 책정 을 참조하세요. 대량 사용 고객은 정액제 를 이용할 수도 있습니다.
		1TB 당 5\$
데이터 로드	무료	BigQuery에 데이터 로드 를 참조하세요.
데이터 복사	무료	데이터 복사 를 참조하세요.
데이터 내보내기	무료	BigQuery에서 데이터 내보내기 를 참조하세요.
메타데이터 작업	무료	list, get, patch, update, delete 호출입니다.

BigQuery ML for predictive analytics

ML 모델 생성 및 사전예측 분석
지원

1

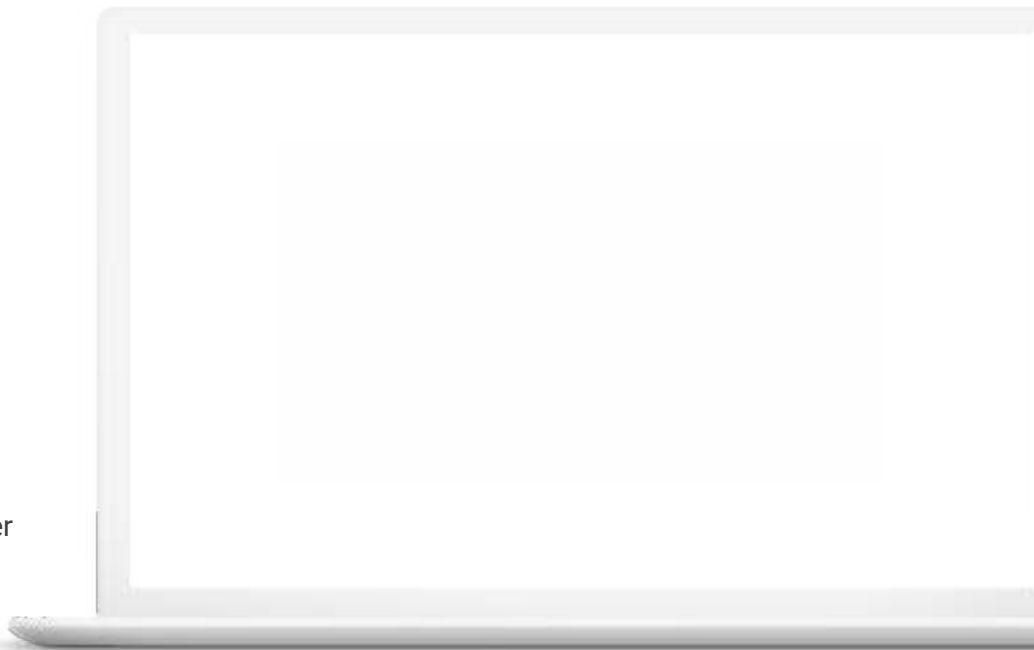
Execute ML initiatives without moving data from BigQuery

2

Iterate on models in SQL in BigQuery to increase development speed

3

Automate common ML tasks, and hyperparameter tuning

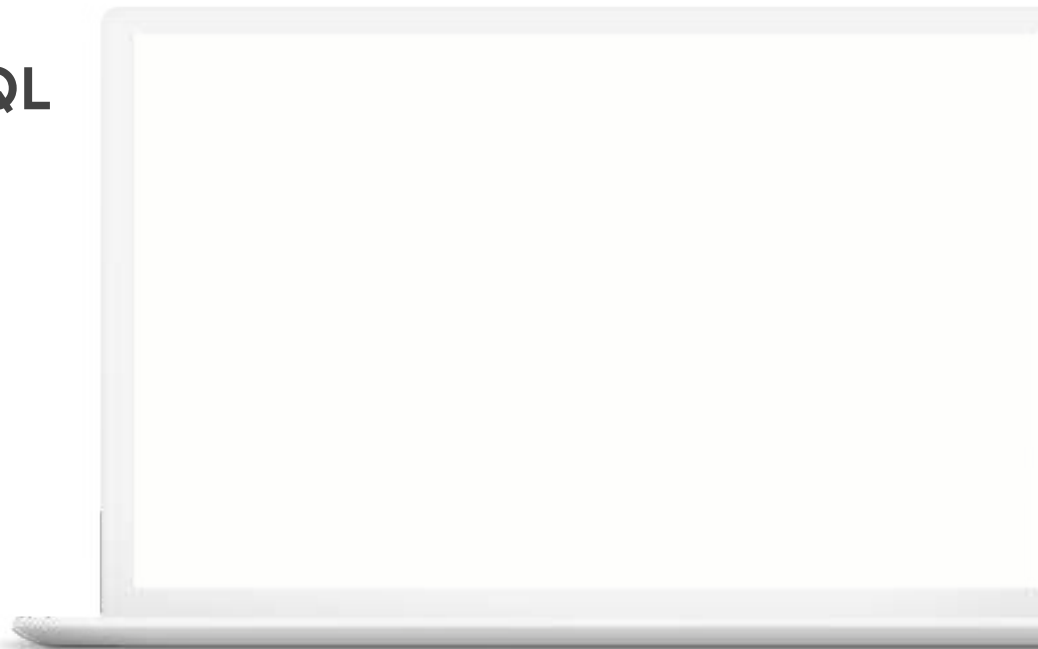


Analyze GIS data in BigQuery with familiar SQL

GIS 데이터 분석 지원

Accurate spatial analyses with
Geography data type over **GeoJSON**
and **WKT** formats

Support for core **GIS functions** –
measurements, transforms,
constructors, etc... – **using familiar SQL**



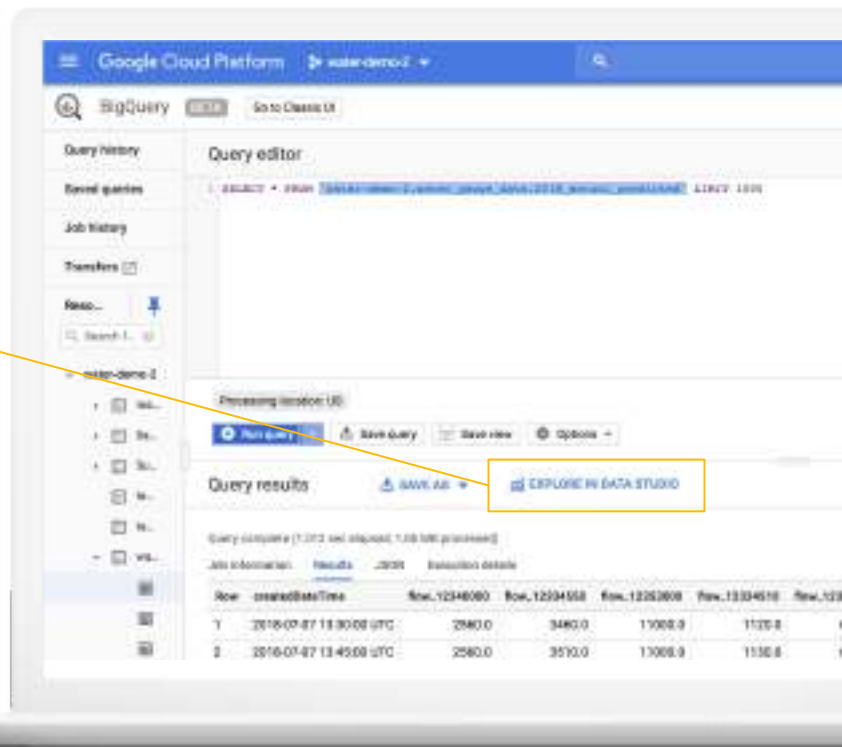
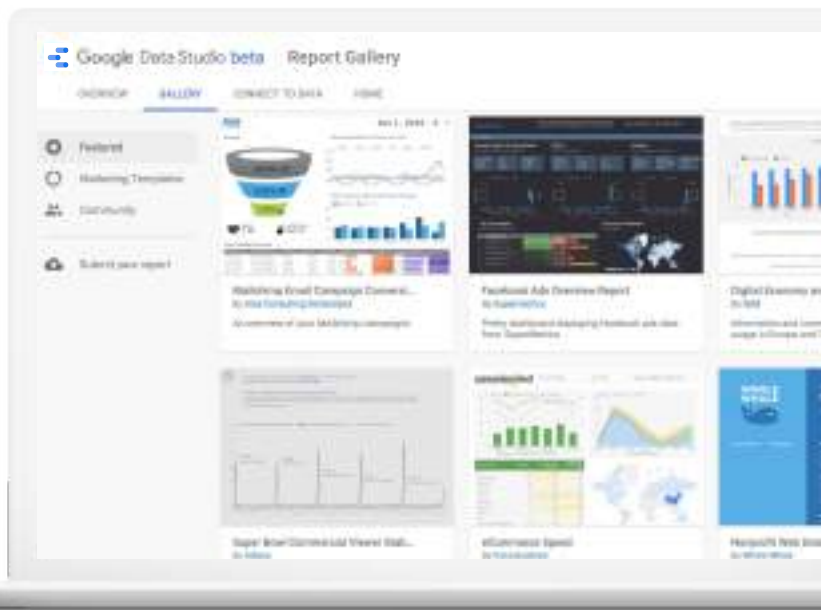
Visualize to identify patterns using Data Studio

Data Studio 로 분석 결과 시각화



See your BigQuery data in one click with Data Studio Explorer

원클릭 BigQuery 분석 결과 리포팅



Cloud Pub/Sub Scalable Event Ingestion and Delivery

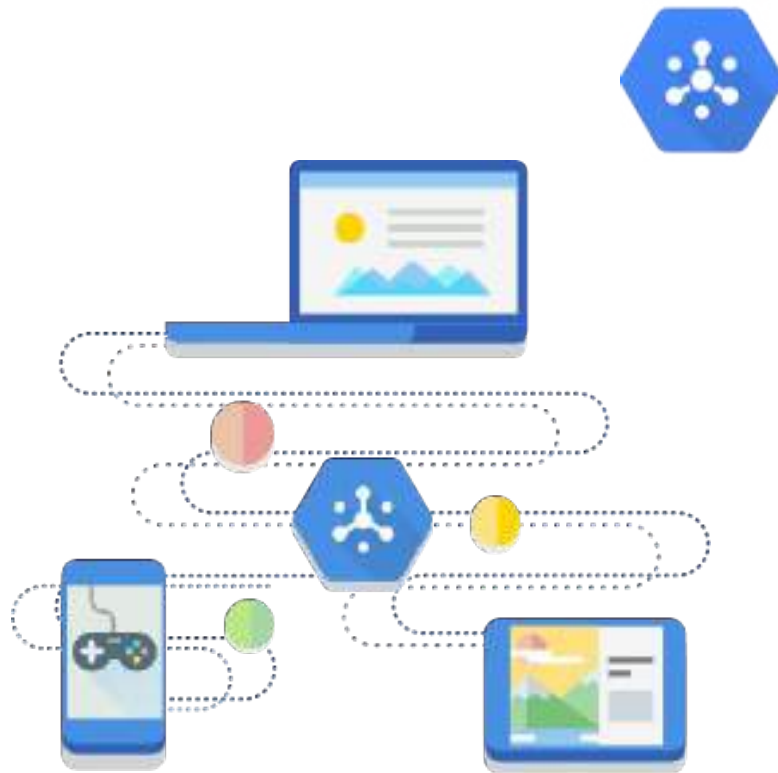
확장 가능한 글로벌 이벤트 처리 서비스

Global, fully-managed event delivery

Ingest millions of events per second

Integrated with **Cloud Dataflow for stream processing**

Streaming Inserts to BigQuery



Cloud Dataflow managed data processing service



스트리밍 & 배치 데이터 ETL 파이프라인 프로세싱

Managed service for **parallel processing** (e.g. ETL) of large data sets in Google Cloud

Supports both **Streaming and Batch** processing (no need for separate pipeline code development)

Autoscales up and down based on processing load

A screenshot of the Google Cloud Platform console showing a Dataflow job. The main area displays a pipeline diagram with five stages: 'PTableLoadTransform0' (59 sec), 'PMapTransform' (8 sec), 'PMapTransform2' (3 sec), 'PMapTransform4' (32 sec), and 'PMapTransform5' (8 sec). The 'Job Logs' panel at the bottom shows execution logs for 'Executing operator PMapTransform0/Combine Statistics/Combine/ParRay/GroupBy/Join/Close'. The right sidebar contains a 'Summary' section with job details: Job Name 'cloud-dataprep-dcm-wrangle2-\$16-by-kartik-trasi', Job ID '20170901_28_33_18-792349617056583397', Job Status 'Running (Stop job)', Job Type 'Batch', and a 'Worker History' graph at the bottom right showing a single worker's activity over time.

Dataprep bringing data closer to analysts

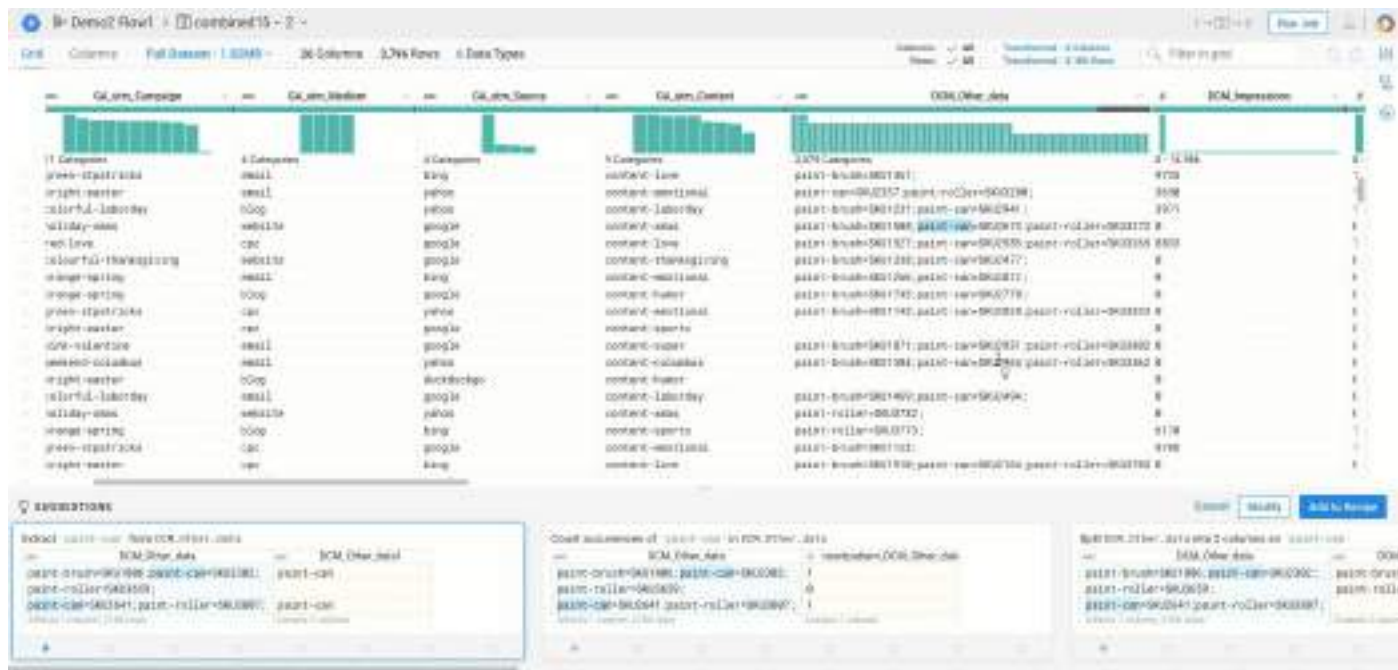


사전 데이터 분석 시각화 (by Trifacta)

Data cleansing and other transformations driven via UI

Connect <-> Data Lake with Data Warehouse

Runs Dataflow pipeline "under the hood"



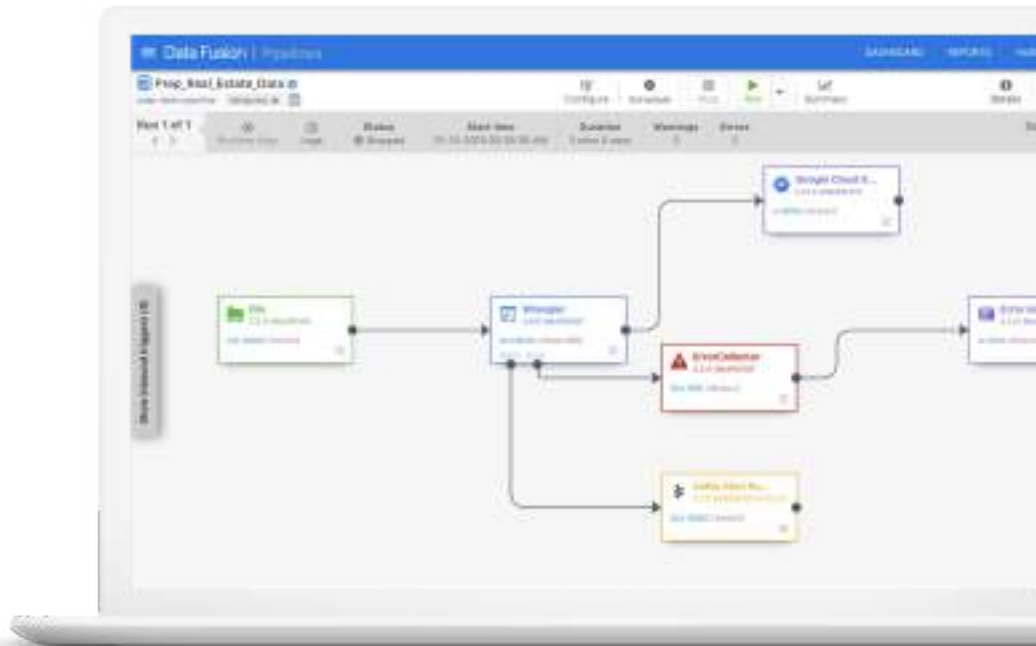
Cloud Data Fusion

코드 작성없이 손쉽게 ETL 파이프라인 관리



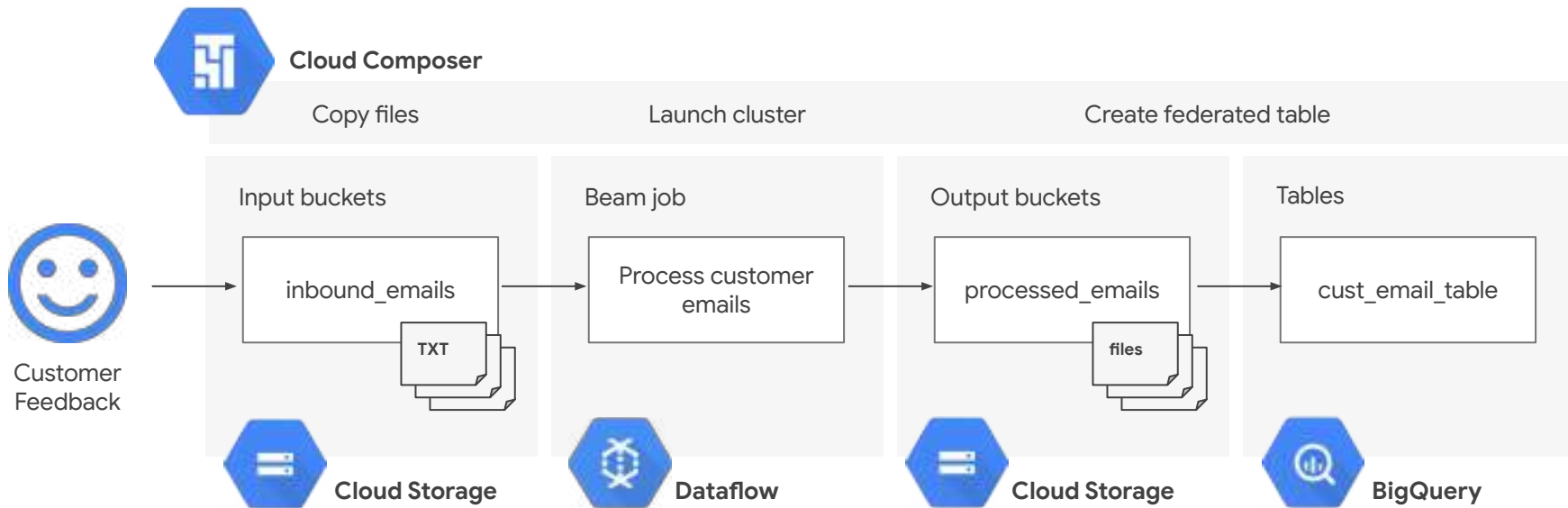
Fully-managed, **code-free** data integration service (ETL/ELT) for hybrid and multi-cloud environments

Based on CDAP **open-source** software



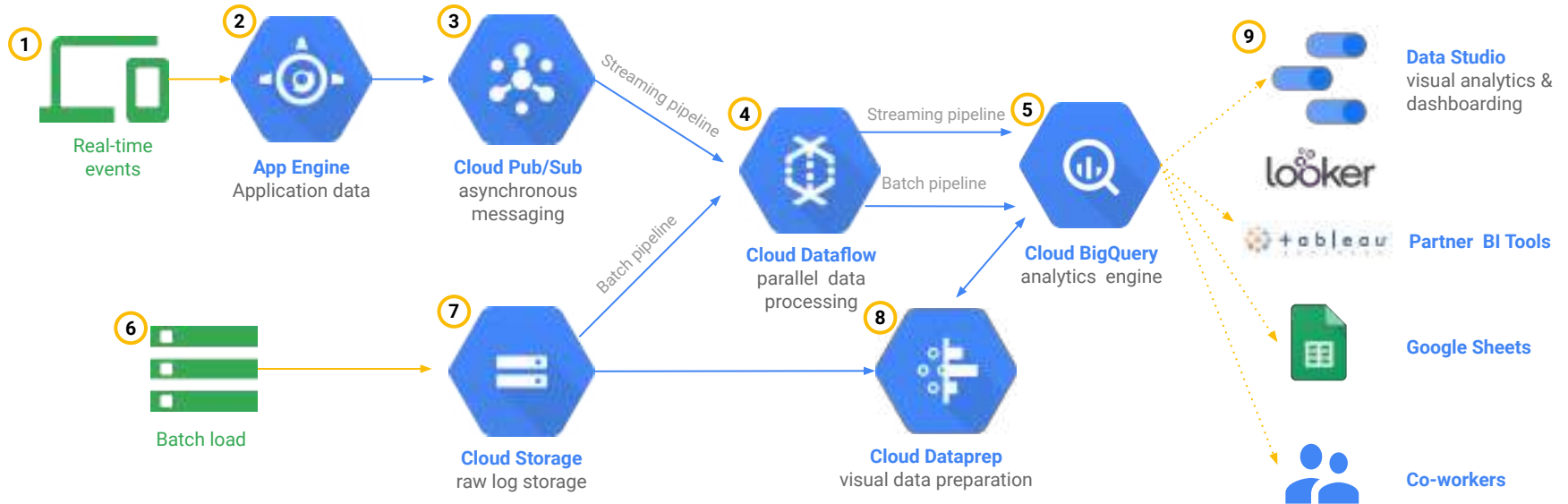
Cloud Composer example

전체 워크플로우 오케스트레이션



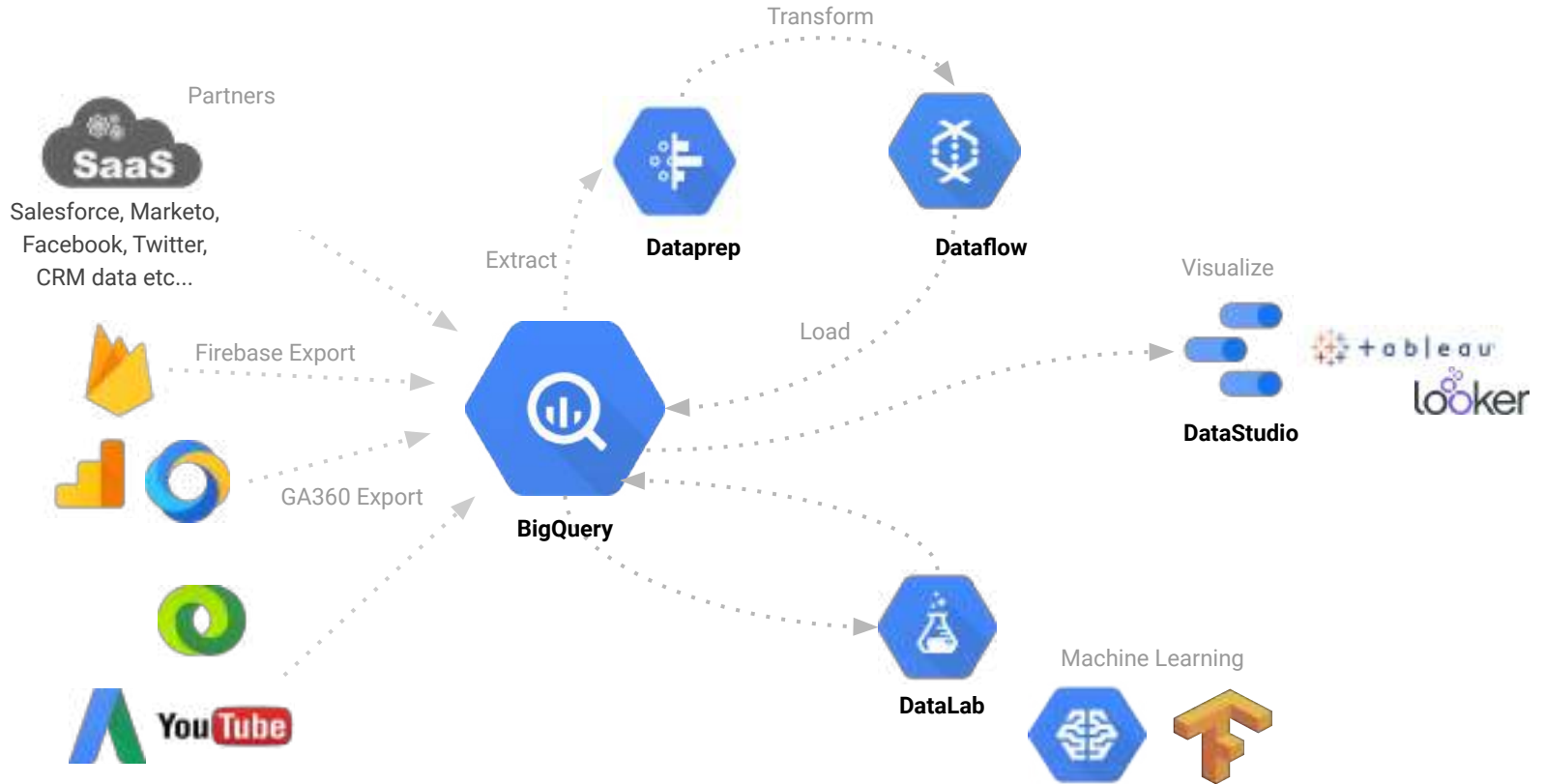
Modern data warehouse on Google Cloud Platform

구글 클라우드 플랫폼 통한 데이터 라이프사이클 구현

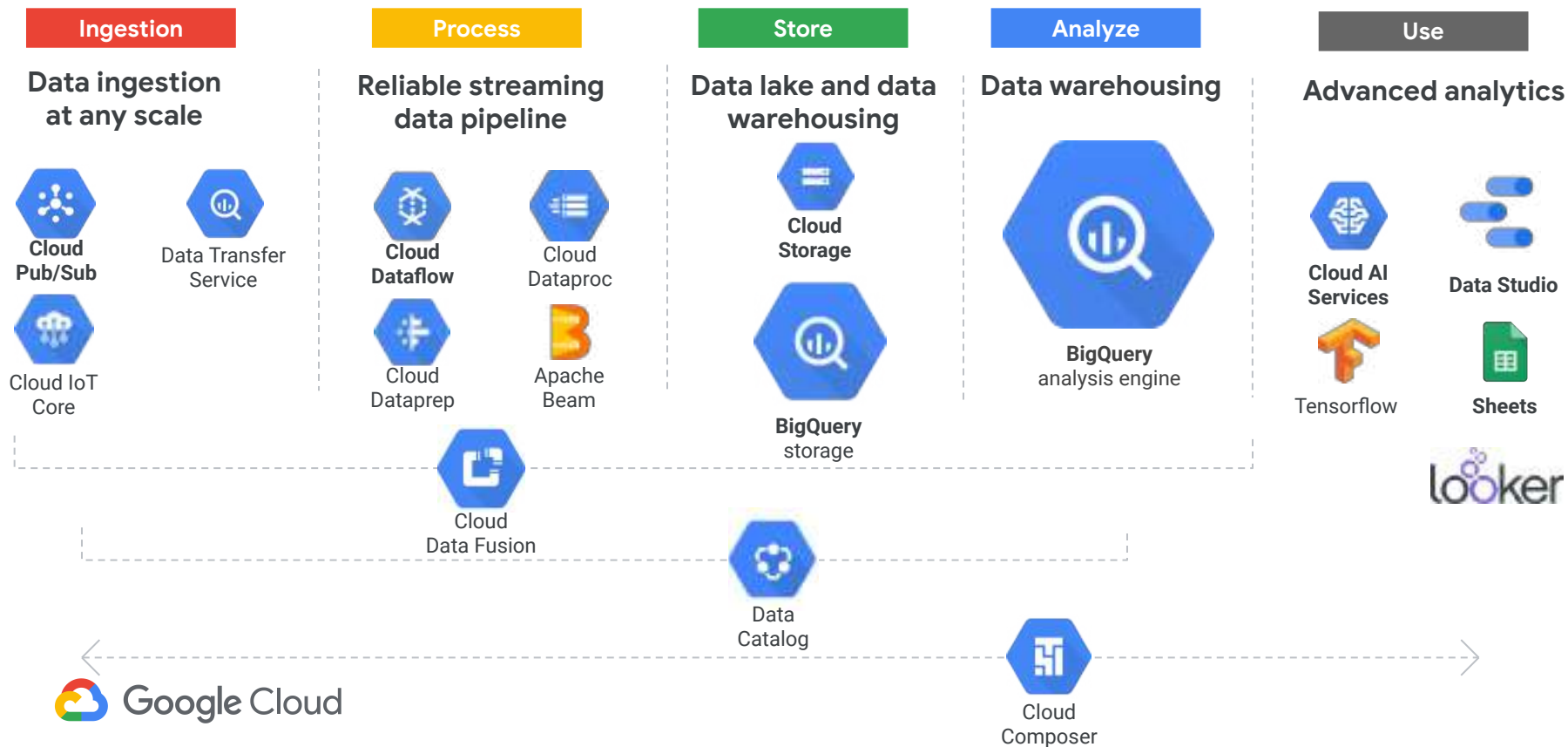


Marketing Analytics with Automated Data Delivery

향상된 마케팅 분석 플랫폼



Modern Data Warehouse on Google Cloud





“데이터 분석, Google Cloud 와 함께
시작해보세요”

Google Cloud