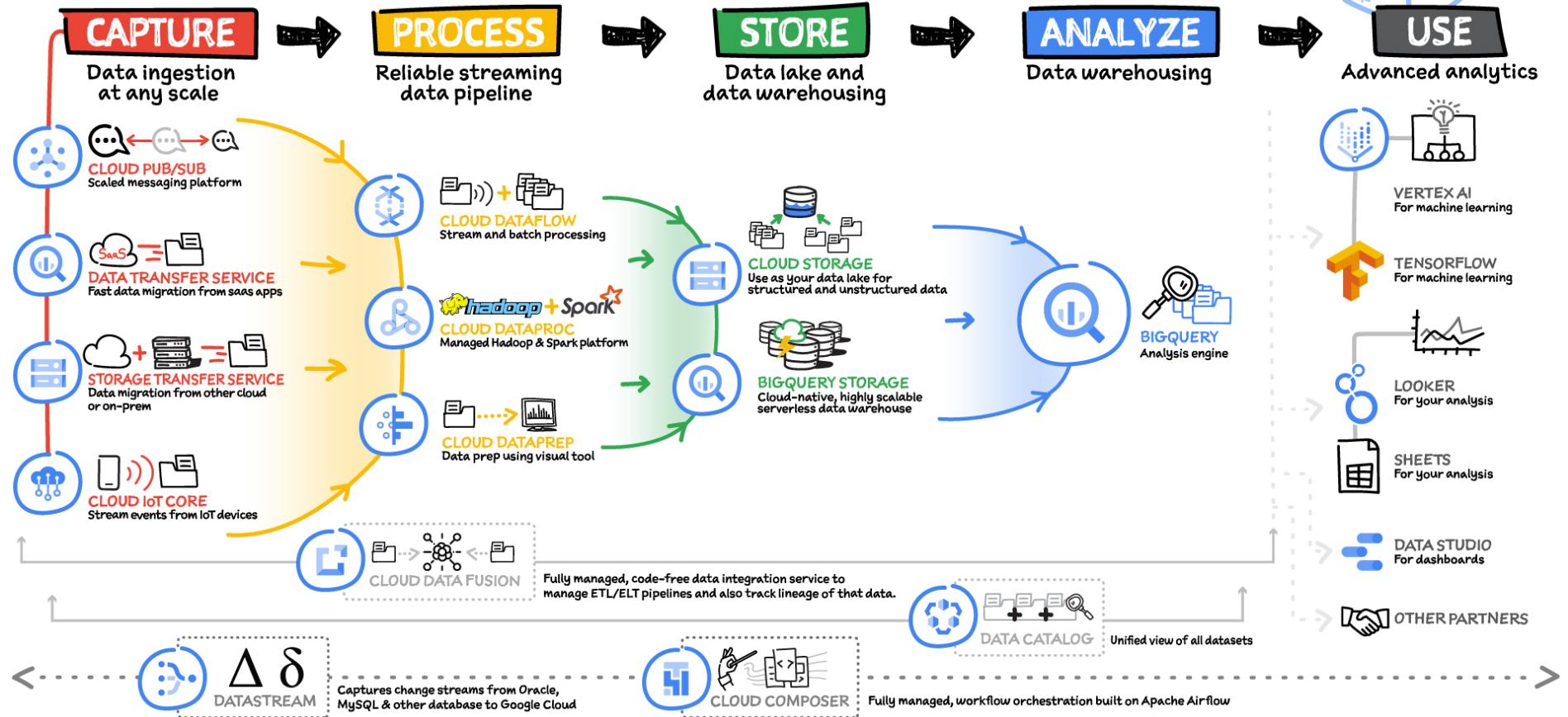




@PVERGADIA
THECLOUDGIRLDEV
04.05.2021

How to build a scalable DATA ANALYTICS PIPELINE





Cloud Storage

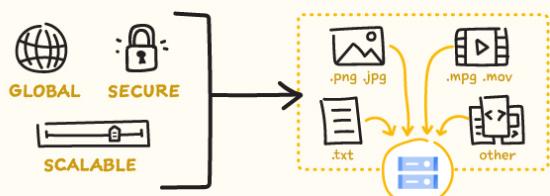
#GCPSketchnote

@PVERGADIA THECLOUDGIRL.DEV 4.23.2021

What is (电影) Cloud Storage?

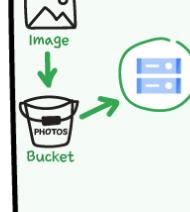
A GLOBAL, SECURE, AND SCALABLE OBJECT STORE

GLOBAL **SECURE** **SCALABLE**



How does it WORK?

WORKFLOW



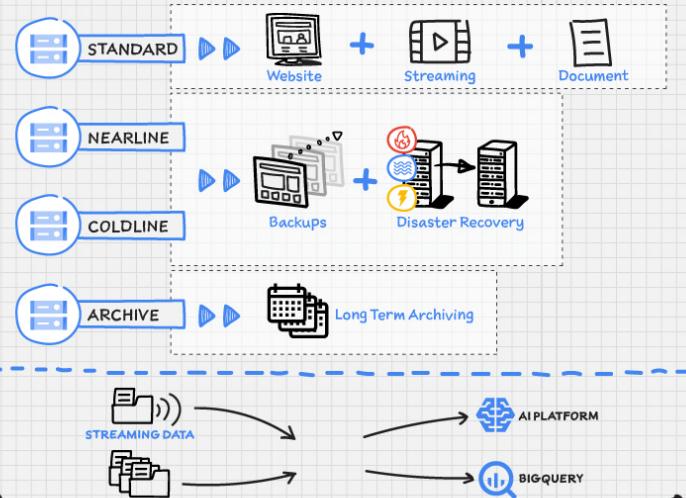
4 STORAGE CLASSES
Based on Budget, Availability, and Access Frequency

STANDARD	NEARLINE	COLDLINE	ARCHIVE
Frequent access High Availability	Once a month	Once a quarter	Once a year
>30 Days	>90 Days	>180 Days	>365 Days

OBJECT LIFECYCLE MANAGEMENT

AUTO VERSIONING

Cloud Storage Use case example



SECURITY for Cloud Storage

- Default encryption at rest Google-managed keys
- Can bring your own encryption key CMEK - Customer Managed CSEK - Customer Supplied



Cloud Storage PRICING

STANDARD	NEARLINE	COLDLINE	ARCHIVE
\$\$\$\$	\$\$\$	\$\$	\$

How to USE Cloud Storage

ONLINE TRANSFER gsutil, API, UI <gsutil cp image.png gs://my-bucket>

TRANSFER SERVICE Transfer data from other clouds & on-premise

TRANSFER APPLIANCE Hardware for >10TB data transfer





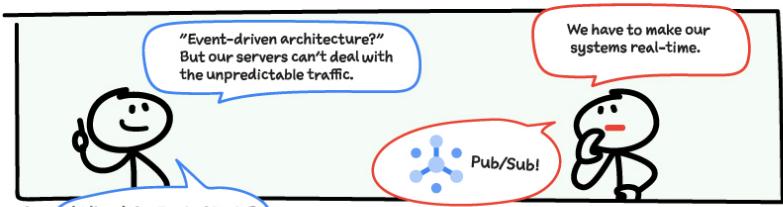
Cloud Pub/Sub

#GCPsketchnote

@PVERGADIA

THECLOUDGIRL.DEV

01.05.2021



What is Pub/Sub?

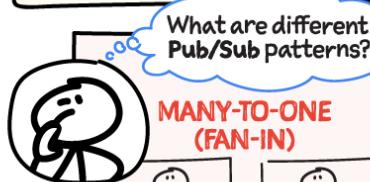
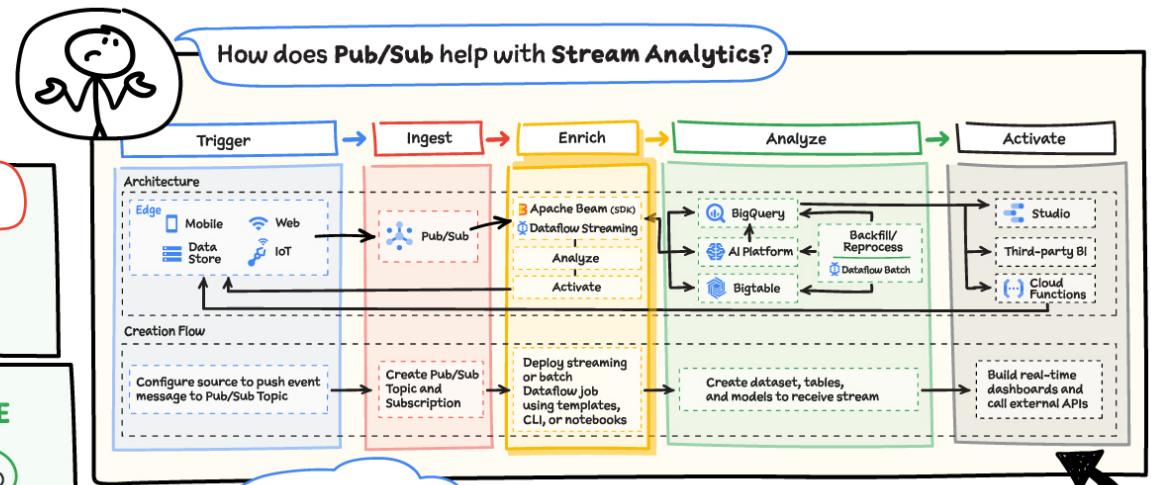
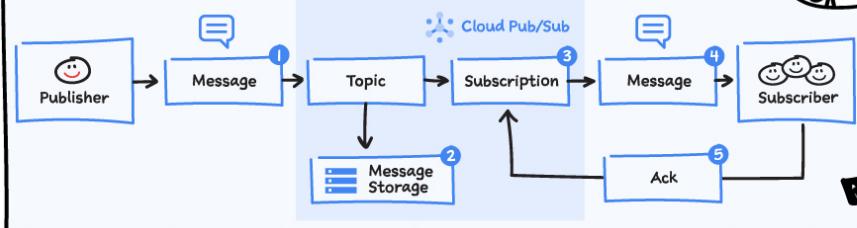
EVENT DRIVEN ASYNCHRONOUS MESSAGING SERVICE

- Service (Produce event) → → Service (Process event)
- In-order delivery at scale
 - Exactly once processing
 - Dead letter topics
 - At-least-once delivery in conjunction with Dataflow
 - Filtering
 - Seek & Replay
 - Pub/Sub Lite Cost optimized delivery
 - Offline message inspection

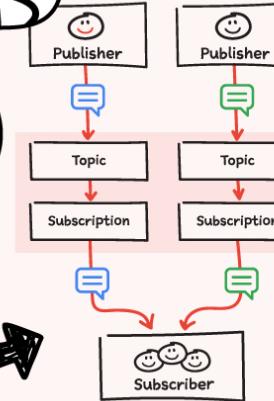


How does Pub/Sub work?

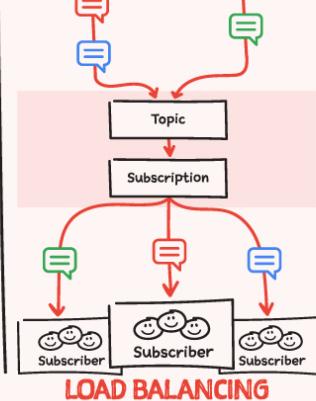
PUB/SUB MESSAGE FLOW



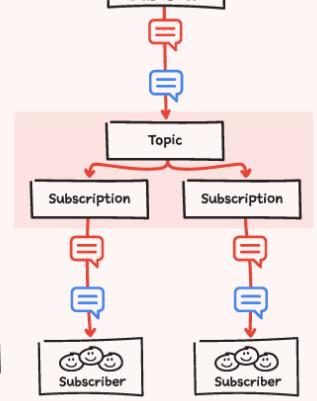
MANY-TO-ONE (FAN-IN)



MANY-TO-MANY



ONE-TO-MANY (FAN-OUT)

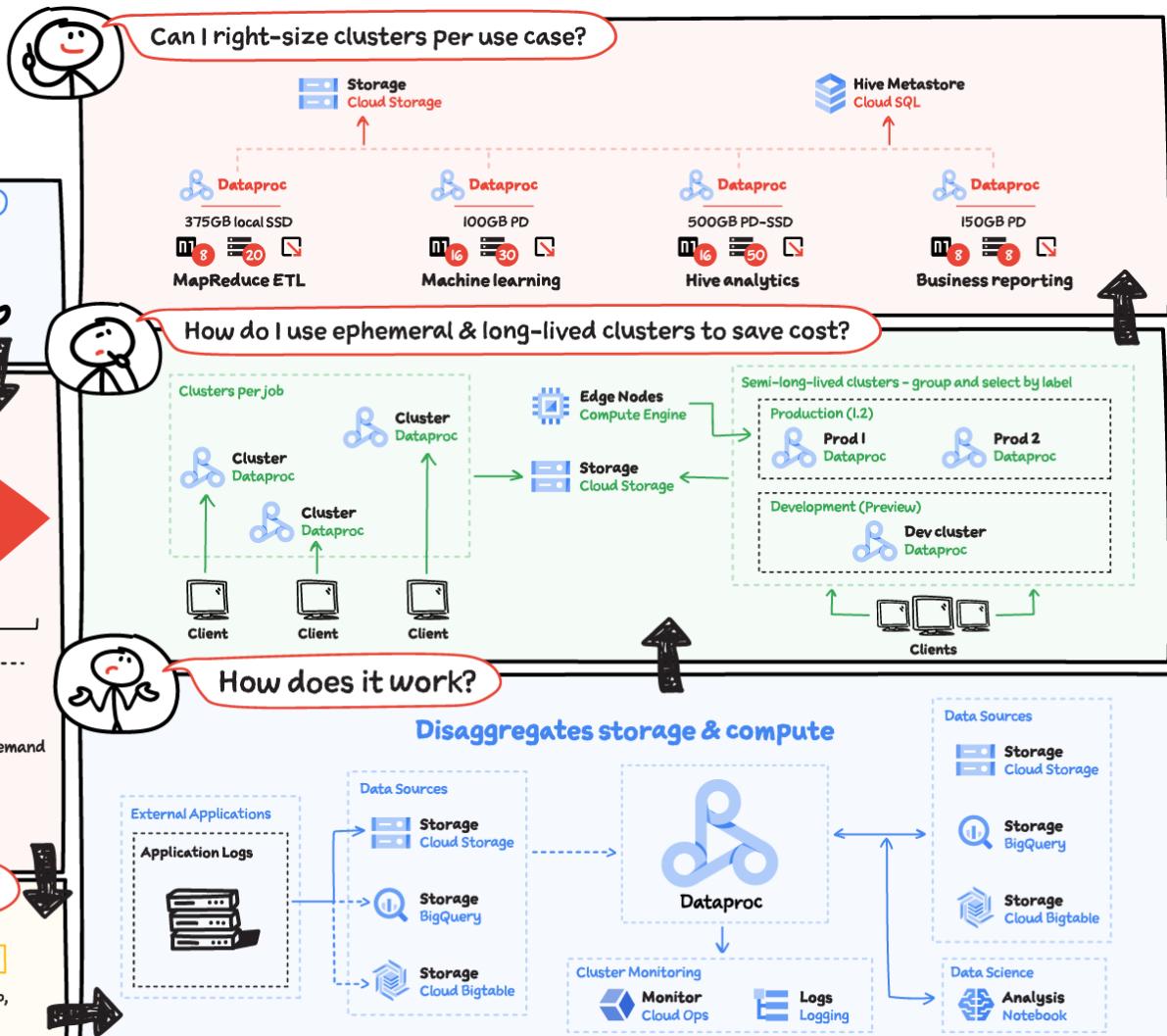
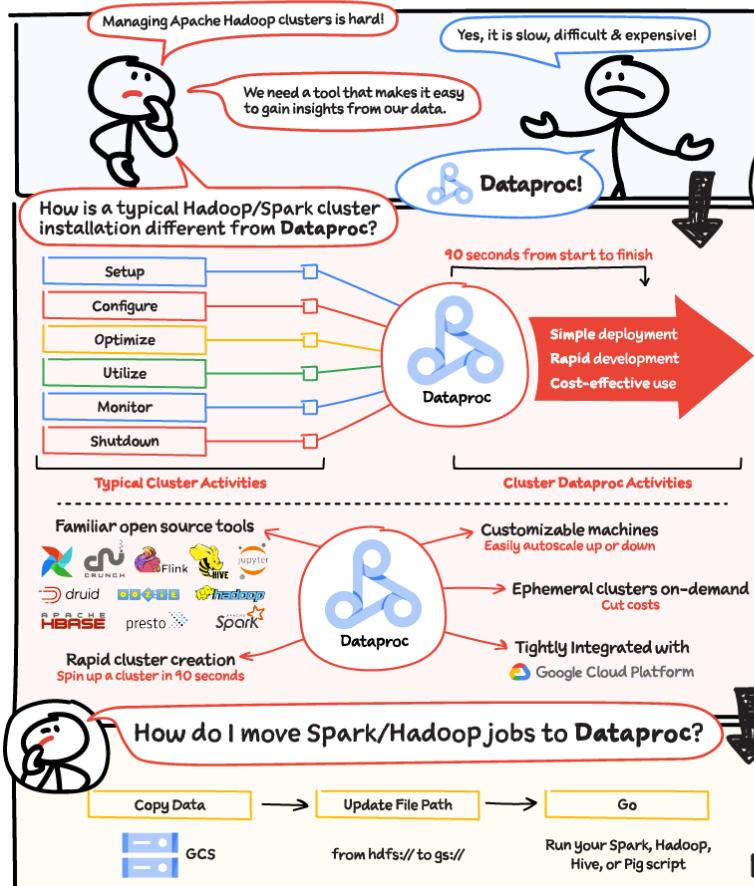


LOAD BALANCING



Dataproc #GCPSketchnote

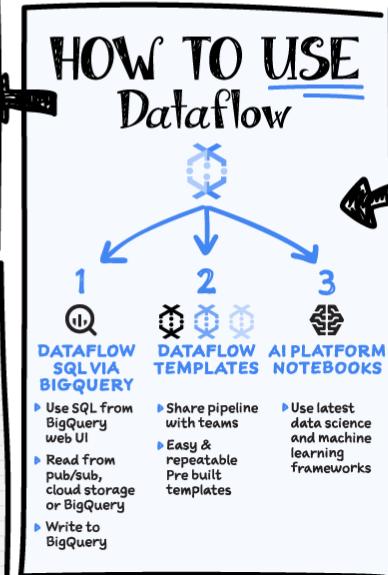
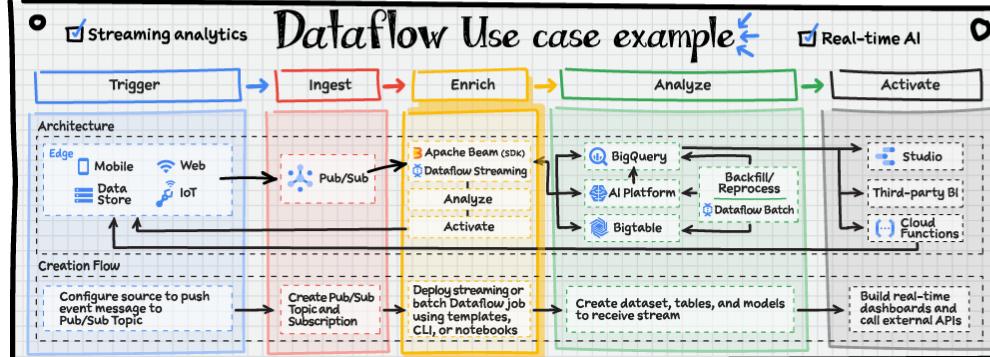
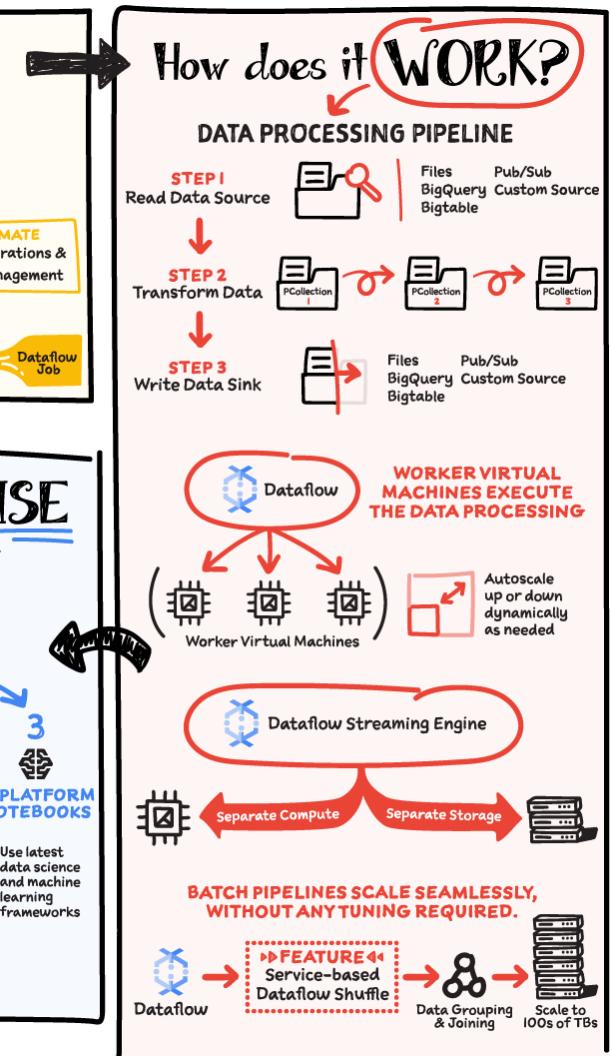
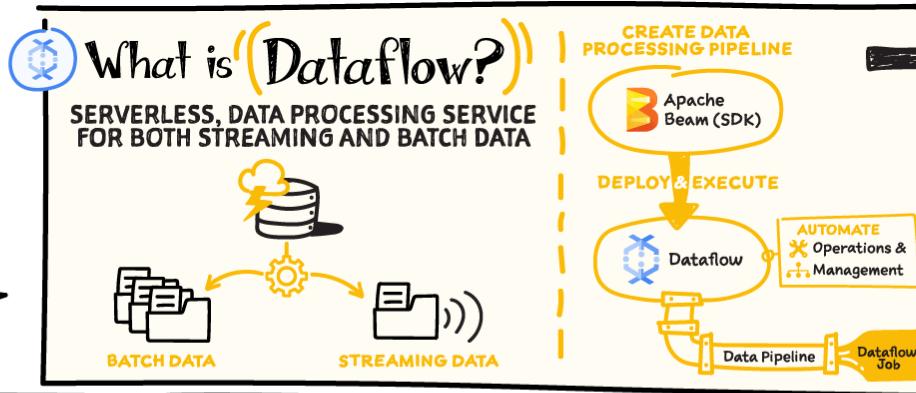
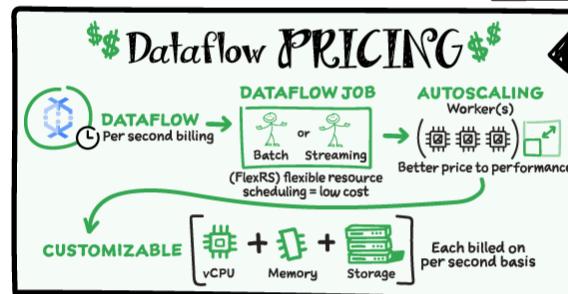
@PVERGADIA THECLOUDGIRL.DEV 09.05.2021



Dataflow

#GCPsketchnote

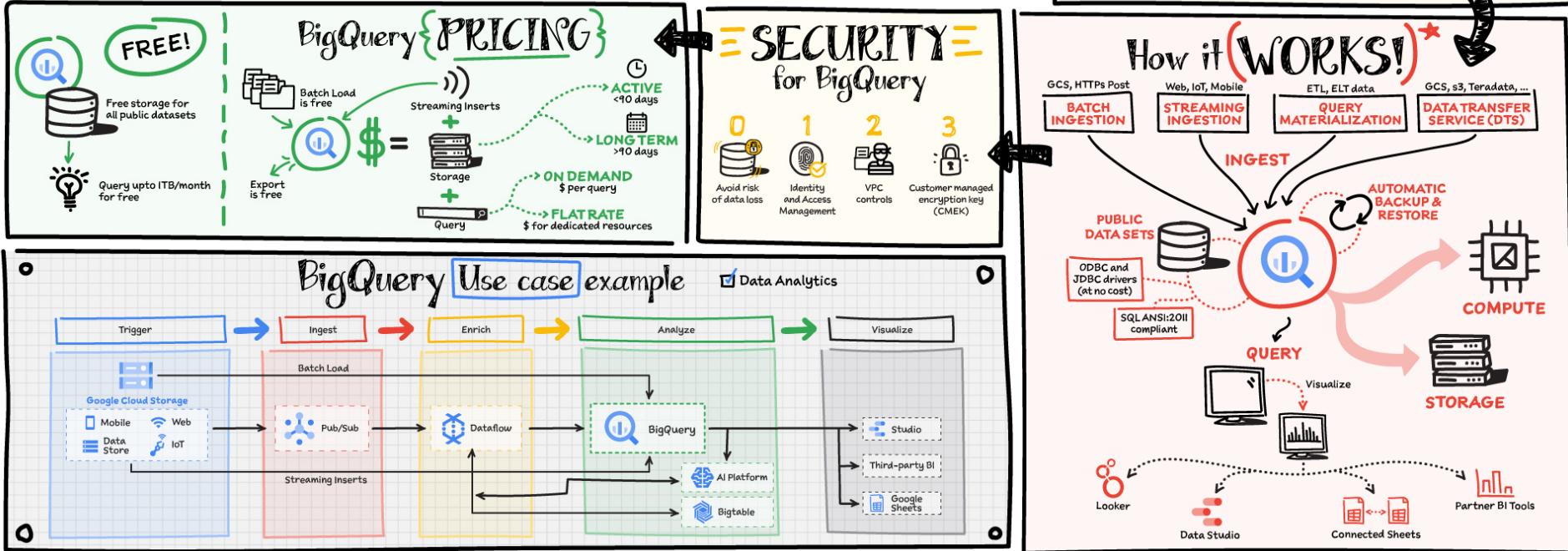
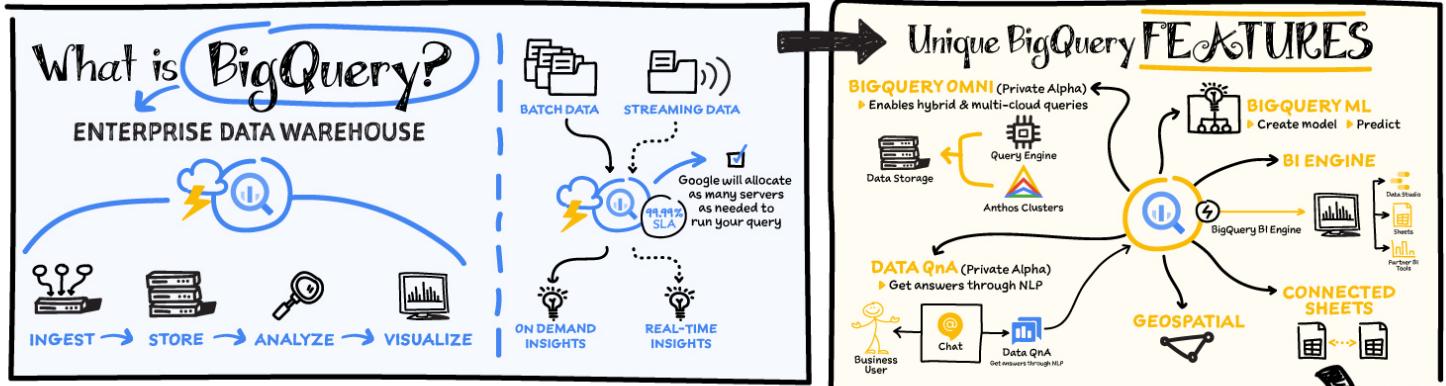
[@PVERGADIA](https://twitter.com/PVERGADIA) [THECLOUDGIRL.DEV](https://thecloudgirl.dev) 04.05.2021



BigQuery

#GCPsketchnote

  @PVERGADIA  THECLOUDGIRL.DEV 6.28.2021



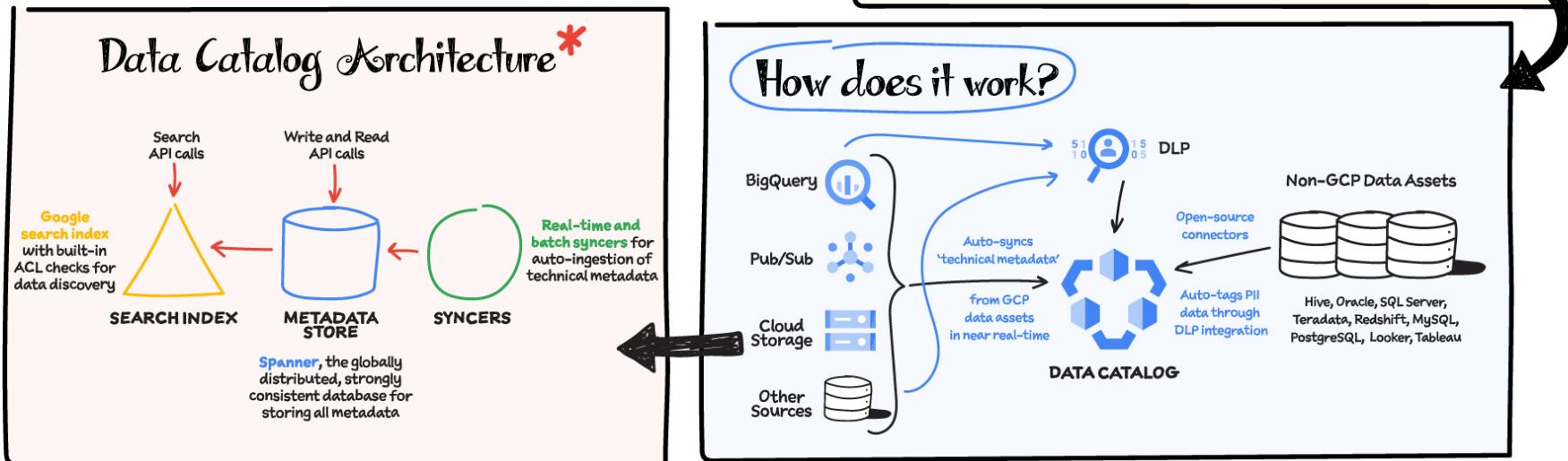
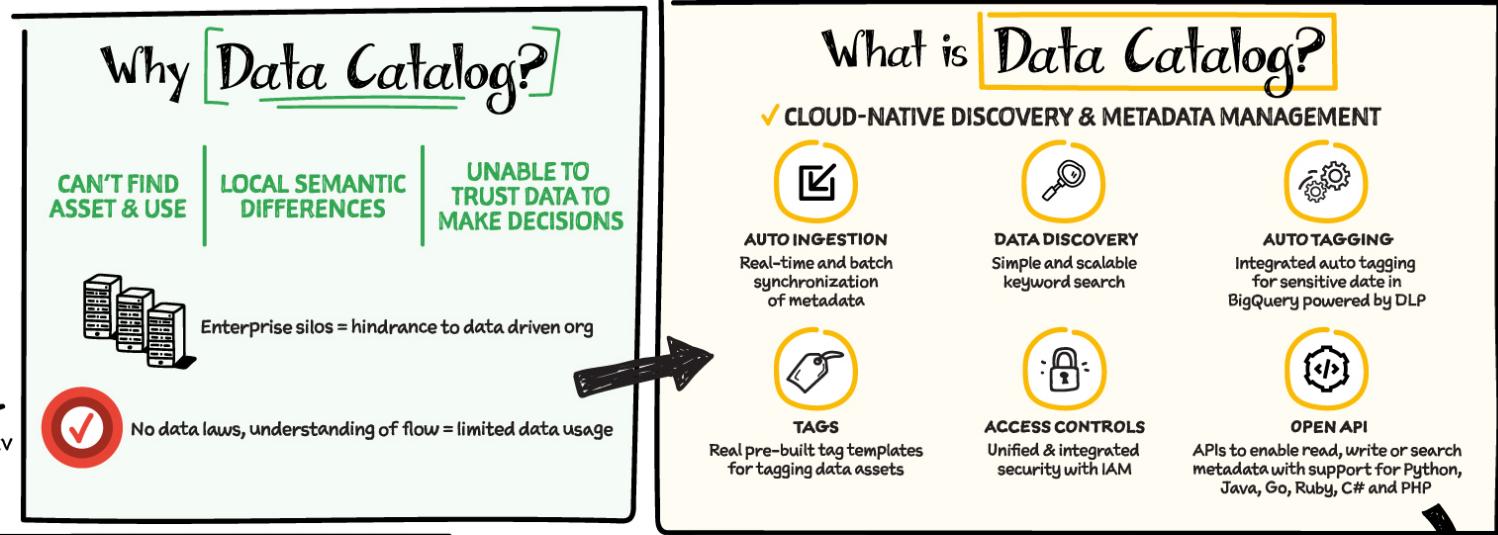


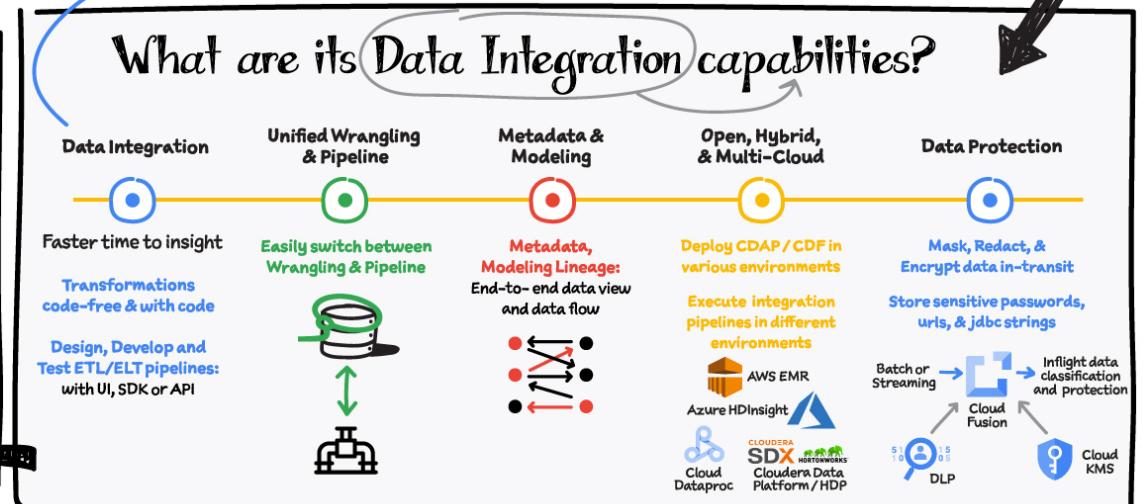
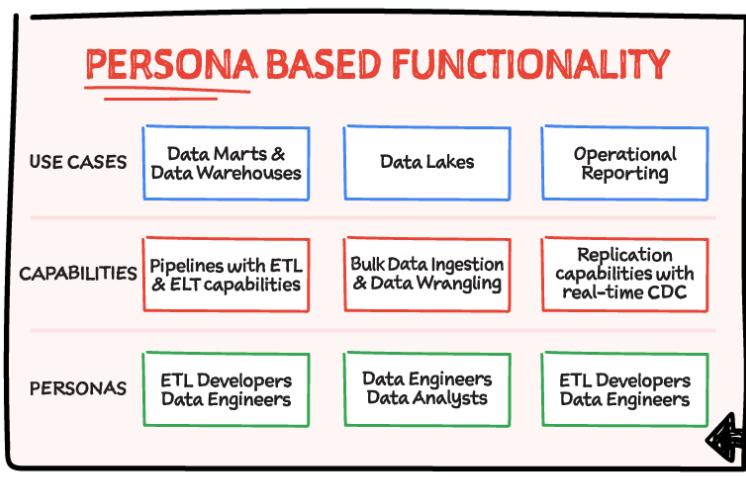
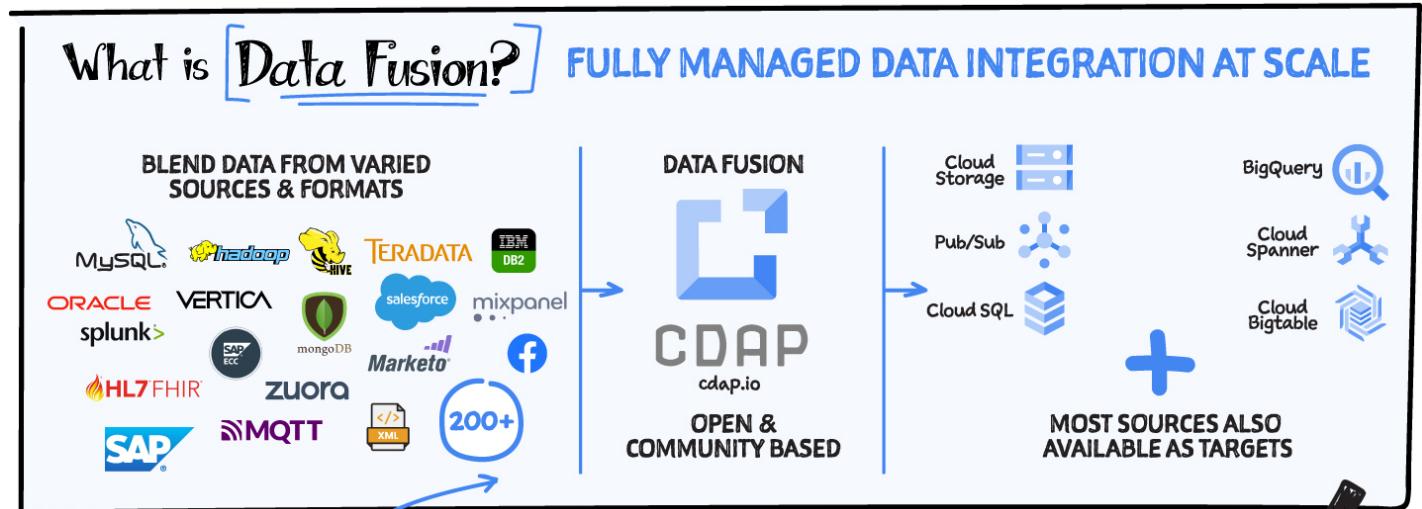
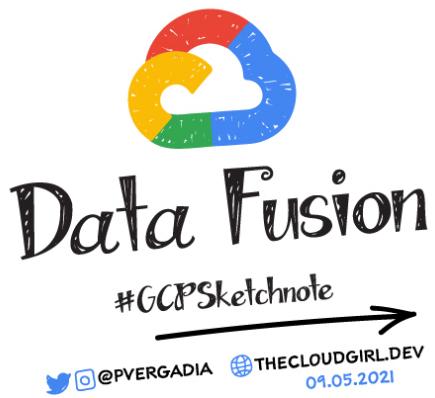
Data Catalog

#GCPSketchnote

[@PVERGADIA](https://twitter.com/pvergadia) [THECLOUDGIRL.DEV](https://thecloudgirl.dev)

7.7.2021







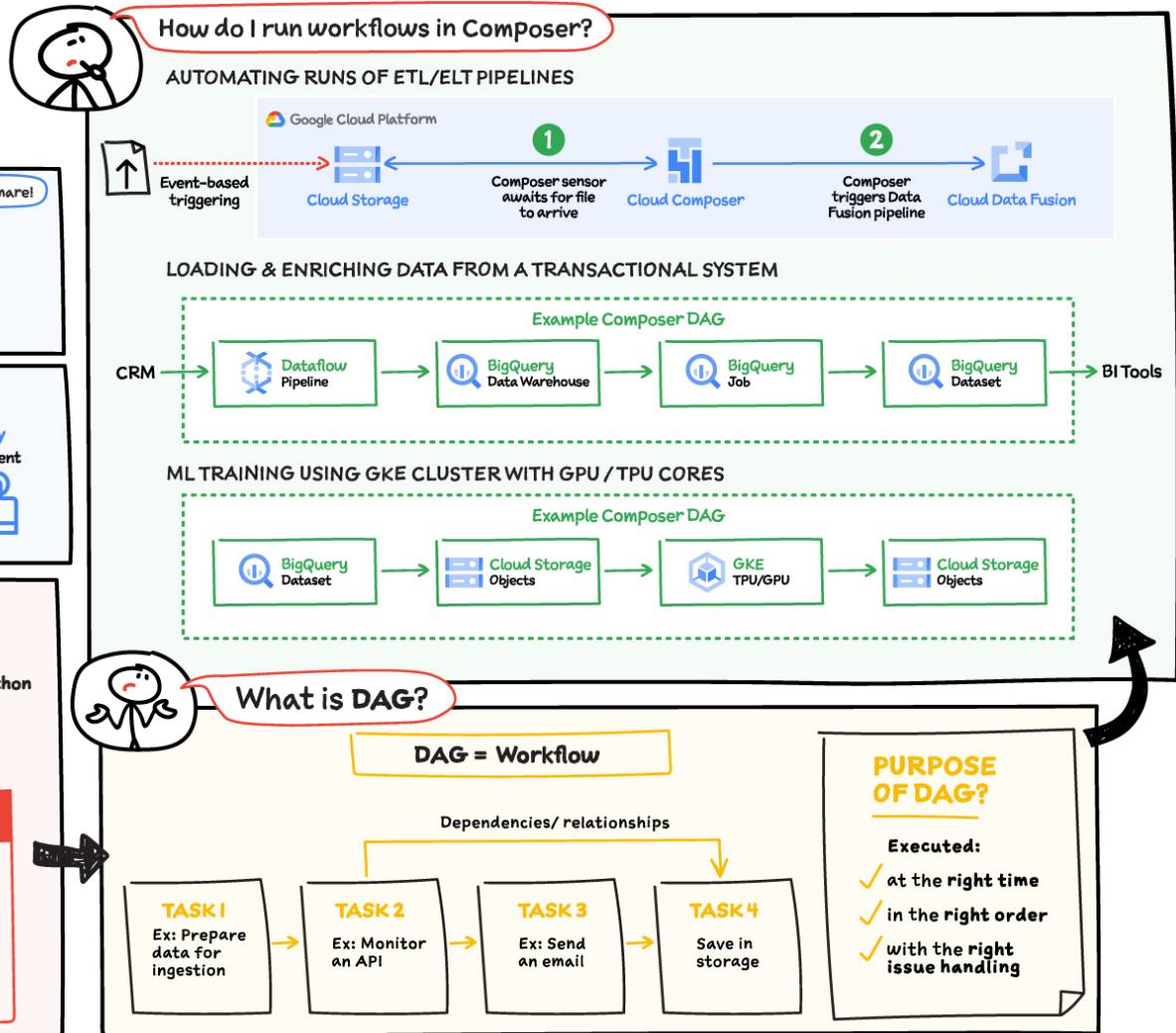
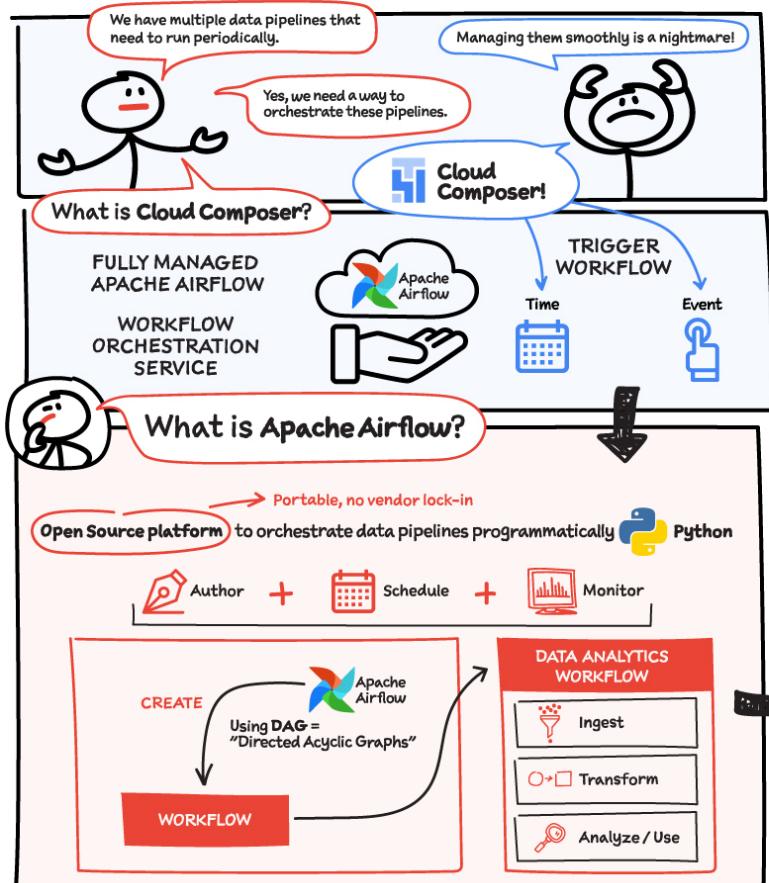
Cloud Composer

#GCPSketchnote

@PVERGADIA

THECLOUDGIRL.DEV

09.05.2021





#GCPSketchnote
@PVERGADIA THECLOUDGIRL.DEV 6.16.2021

Hey

Look! It's Looker



Real-time dashboards, reports, self service data exploration



Embed analytics into existing applications (CRMs, chat applications, etc)



Automate multicloud workflow through smart alerts and data delivery



Build a custom analytics experience

Modern BI & Analytics

Integrated Insights

Data-driven Workflows

Custom Applications



Looker

- API-first & cloud native for integrating into existing workflows
- Semantic modeling layer for enterprise-wide governance
- In-database architecture for access to real-time data

