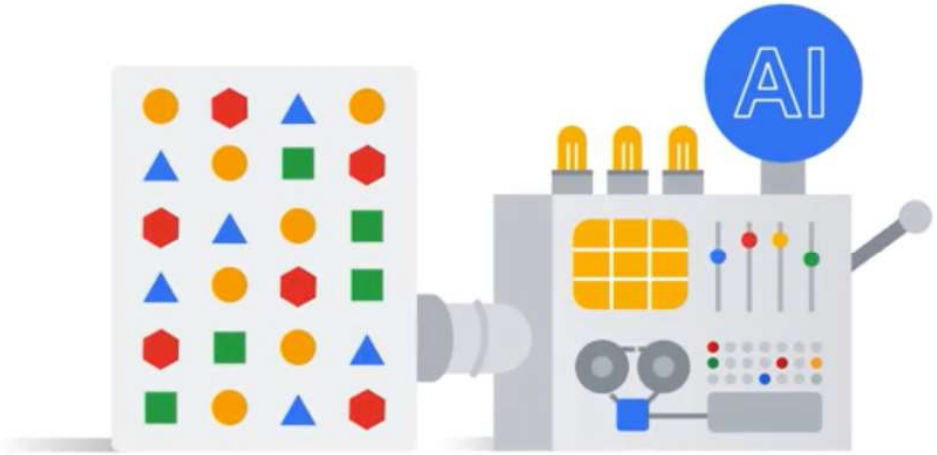
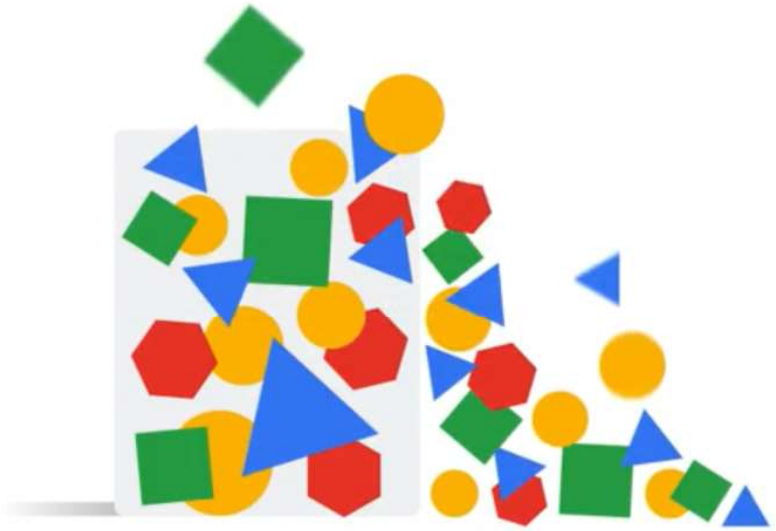
Google Cloud Big Data and Machine Learning Fundamentals



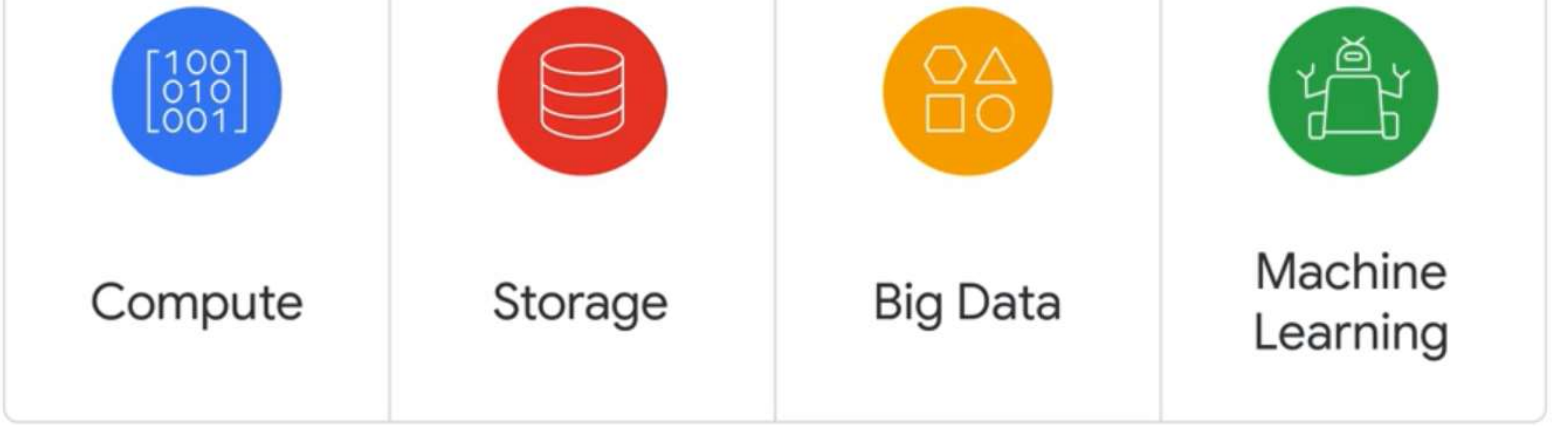
Data is the foundation of every application integrated with artifical intelligence



Without artifical intelligence, large amounts of data can be unmanageable or underutilized

Google

Google




Google Cloud
Big Data and ML Products
Compute
Storage
Networking & Security

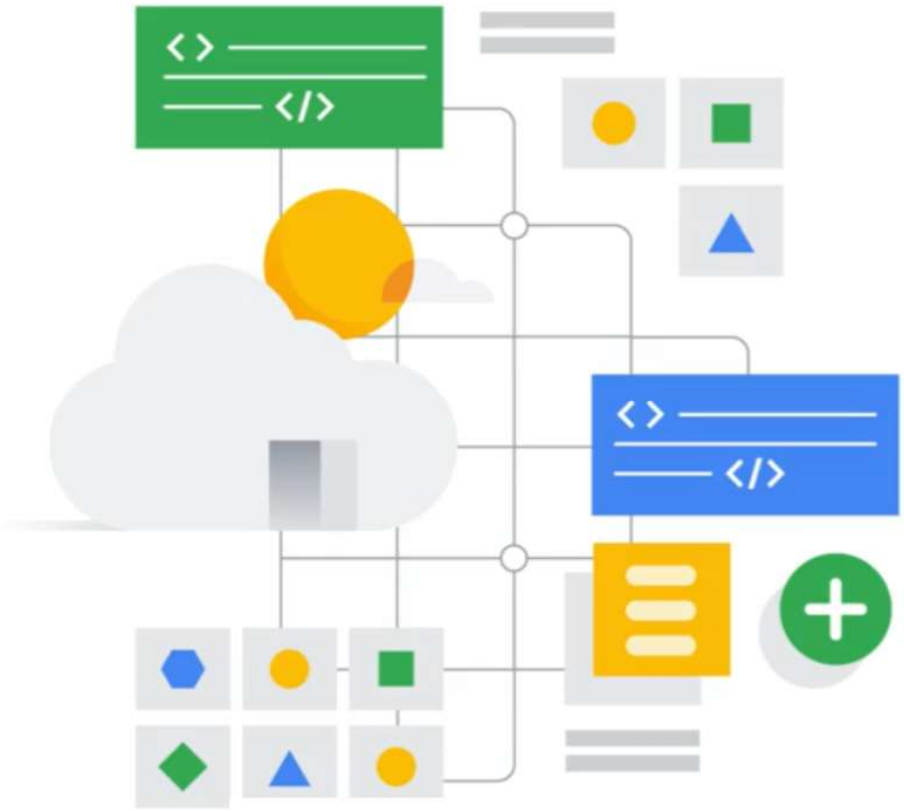

Google Cloud
Big Data and ML Products
Compute
Storage
Networking & Security

Data needs

Compute power

Compute Engine

- IaaS offering
- Raw compute capabilities
- Raw storage capabilities
- Raw network capabilities

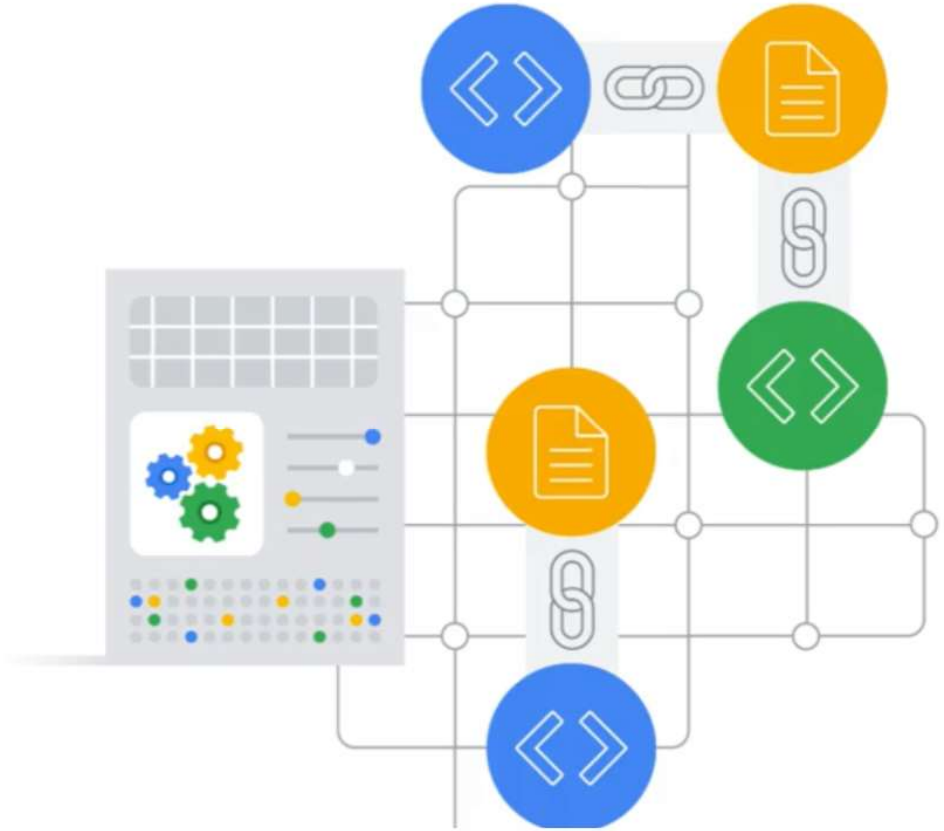


Google Kubernetes Engine

- Containerized applications
- Cloud environment

App Engine

- Fully managed PaaS offering
- Bind code to libraries
- Focused on application logic



Cloud Functions

Executes code in response to events

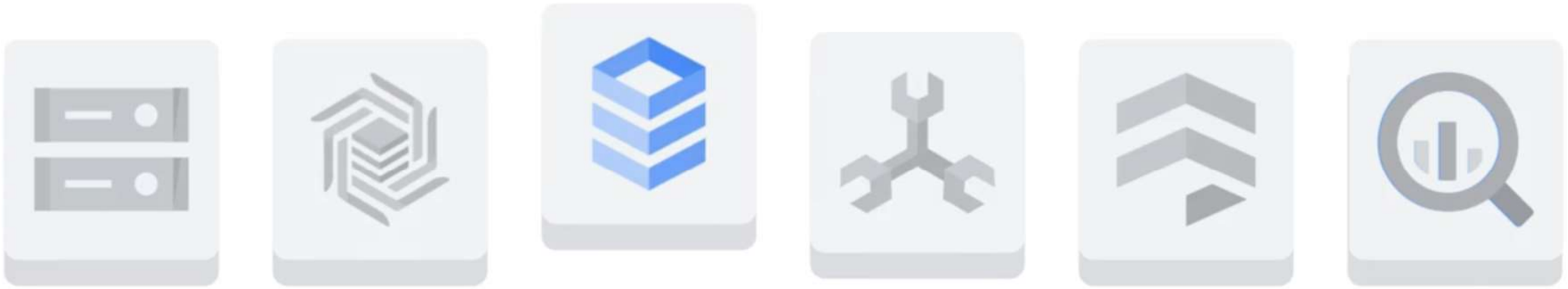
Cloud Run

- Fully managed platform
- Lets you focus on writing code
- Automatically scales up and down
- Charges you only for the resources you use

App Service, Cloud Services	App Engine
Azure Dedicated Host	Sole-Tenant Nodes
Azure GPU optimized VMs	Cloud GPUs
Azure Container Instances	Cloud Run
Azure Functions	Cloud Functions
Virtual Machines	Compute Engine
Virtual Machines Scale Sets	Instance groups
Azure Migrate	Migrate for Compute Engine
Azure Batch, Azure Spot VMs	Preemptible VMs
Azure SQL	Cloud SQL for SQL Server
Azure VMware Solution	Google Cloud VMware Engine

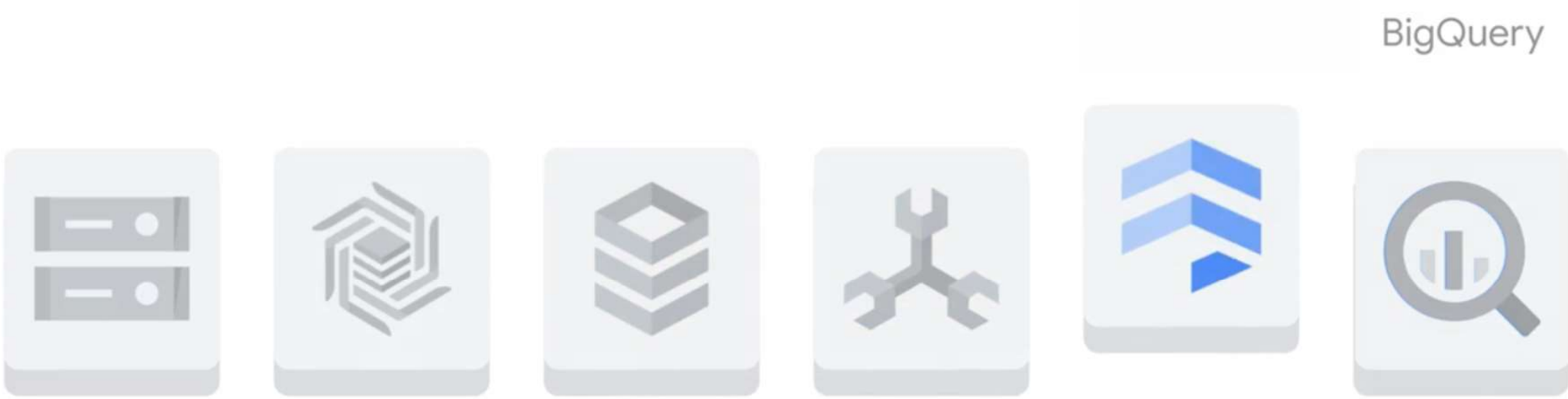
Cloud Storage

Cloud Bigtable


Cloud SQL


Cloud Spanner


Firestore

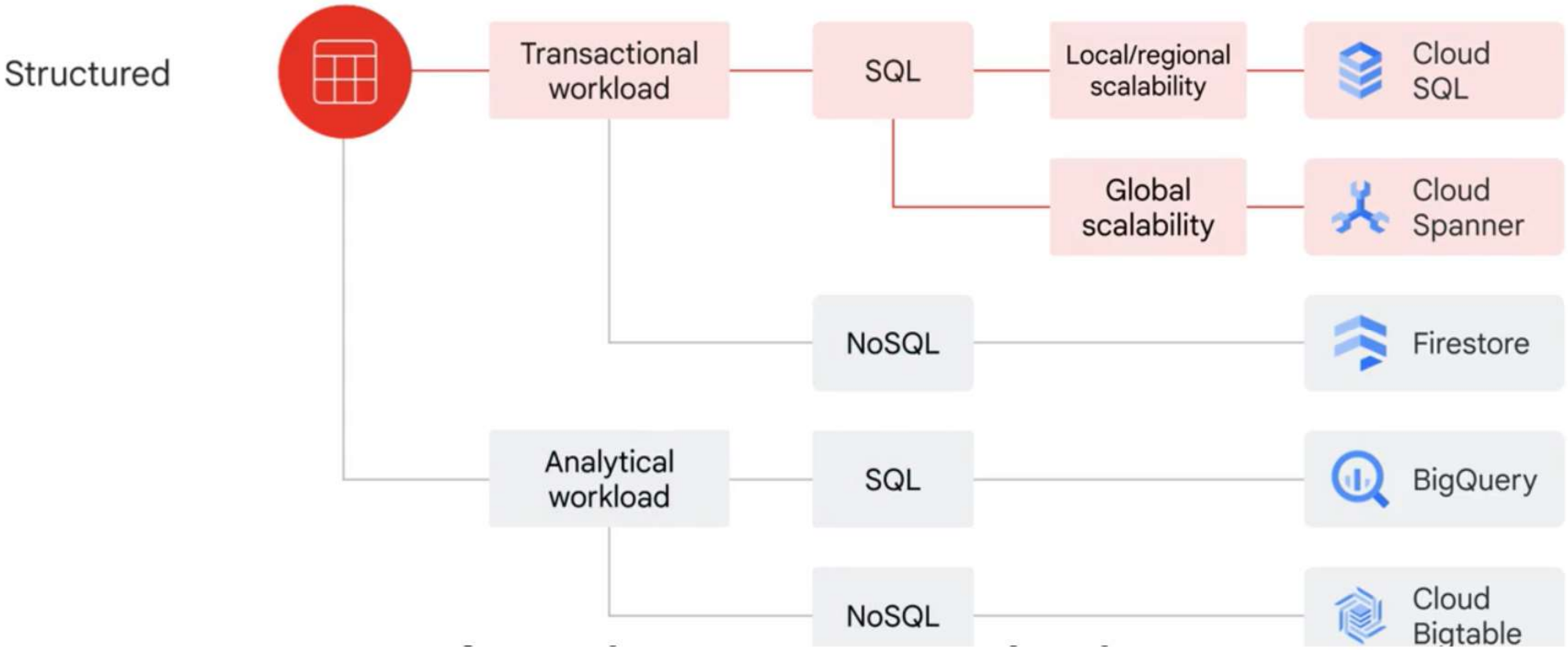

BigQuery

Azure Archive Storage	Cloud Storage Archive
Azure Blob Storage	Cloud Storage
Azure Cool Blob Storage	Cloud Storage Nearline and Cloud Storage Coldline
Azure Data Box	Transfer Appliance
Azure Files	Filestore
Azure Managed Disks	Persistent Disk, Local SSD

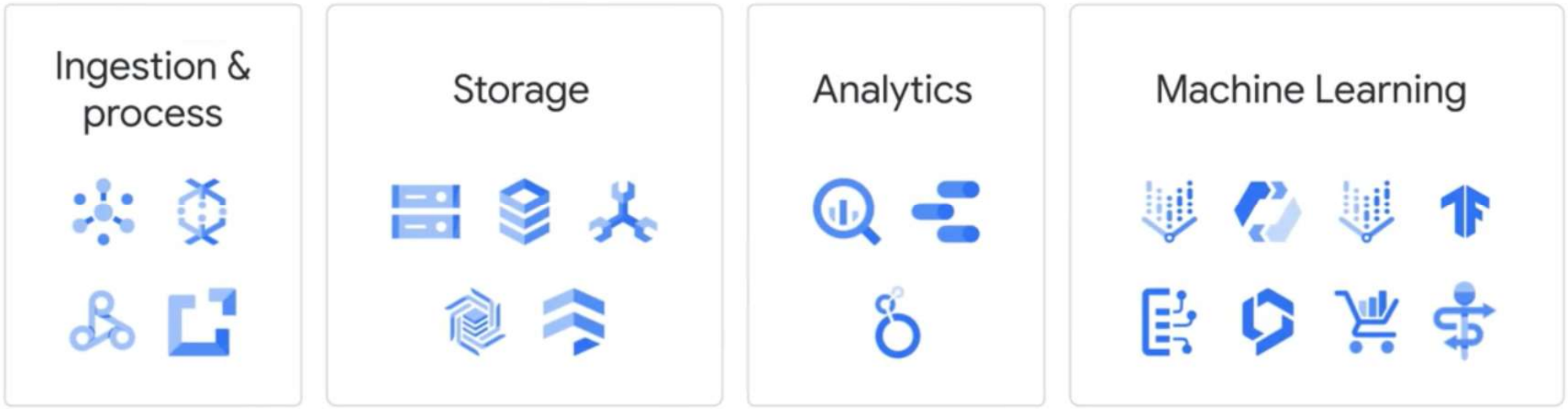
Azure Cache for Redis	Memorystore
Cosmos DB	Firestore
SQL Database	Cloud SQL, Cloud Spanner
Table Storage	Firestore, Cloud Bigtable

The goal of these products is to reduce the time and effort needed to store data

Standard Storage	Nearline Storage	Coldline Storage	Archive Storage
			
Hot data	Once per month	Once every 90 days	Once a year

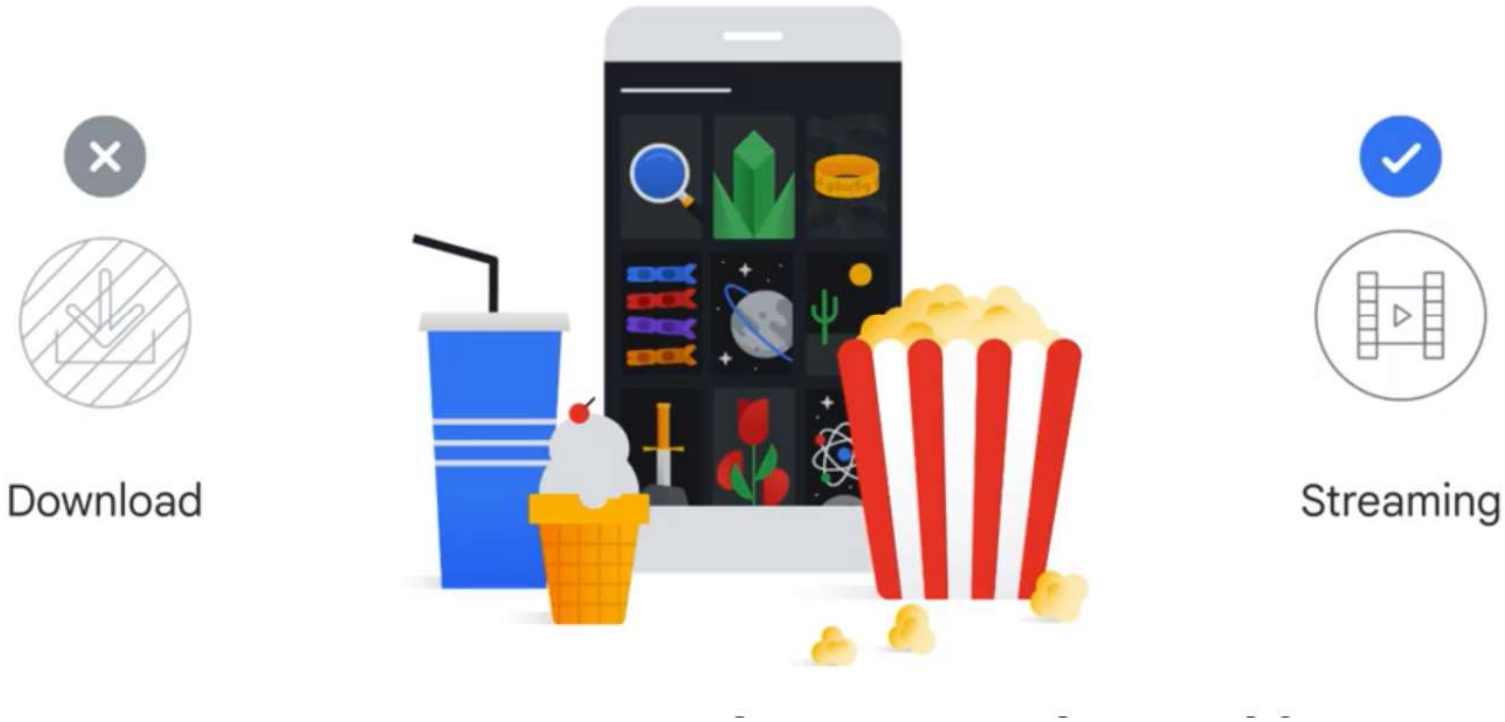

Structured
Transactional workload
SQL
Local/regional scalability
Cloud SQL
Global scalability
Cloud Spanner
NoSQL
Firestore
Analytical workload
SQL
BigQuery
NoSQL
Cloud Bigtable

As the internet grew, Google needed
to invent new data processing methods

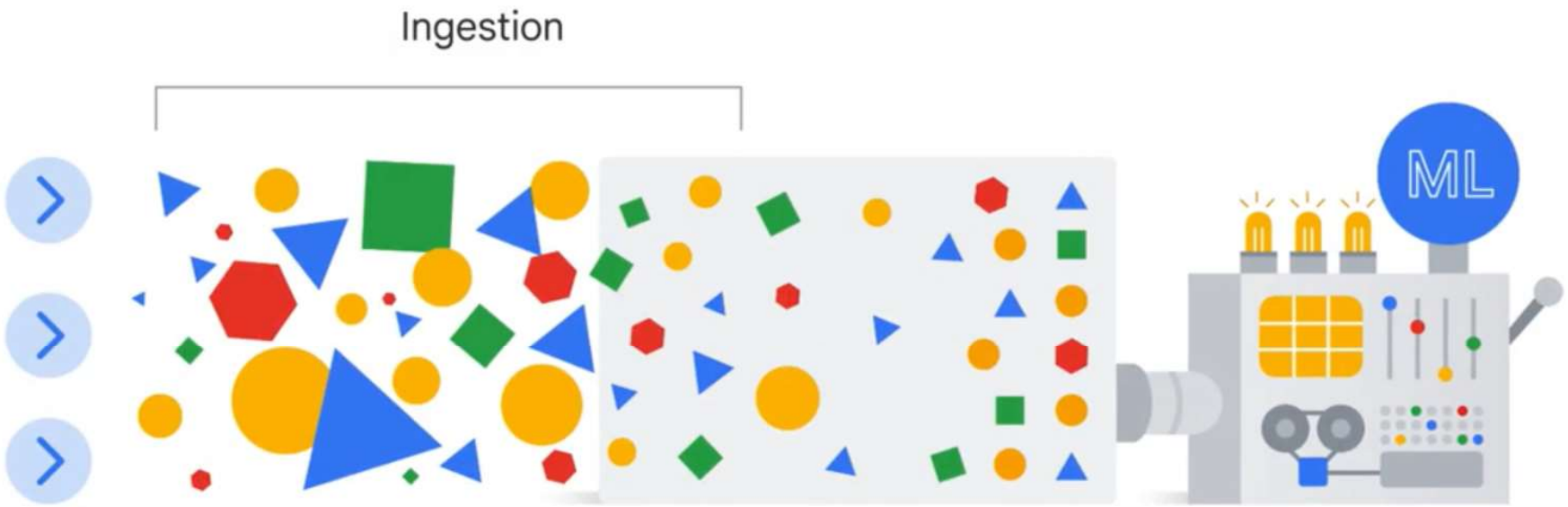

Ingestion & process
Storage
Analytics
Machine Learning

What streaming data is, how it differs from batch processing and why it's important

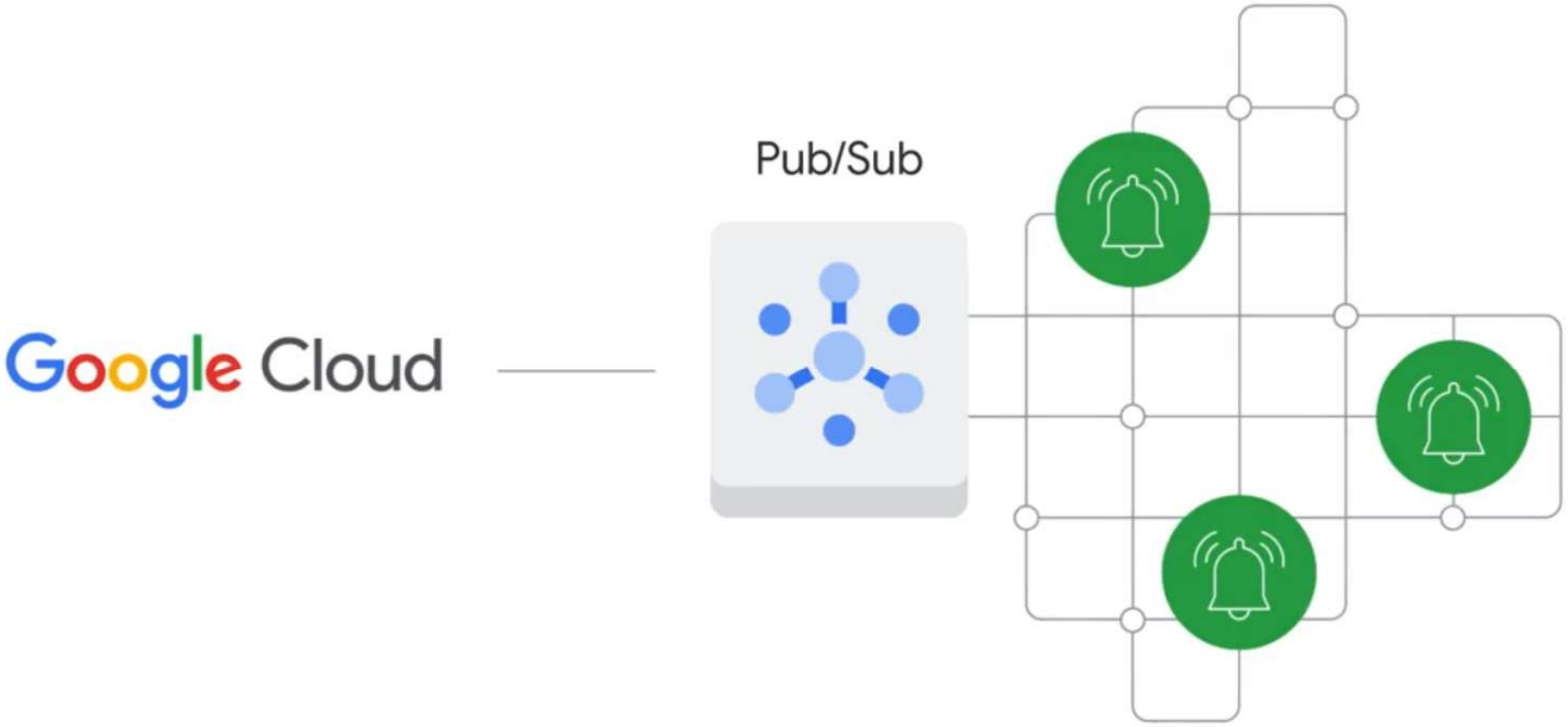
The data is analyzed in near real-time
and that actions will be taken on
the data as quickly as possible

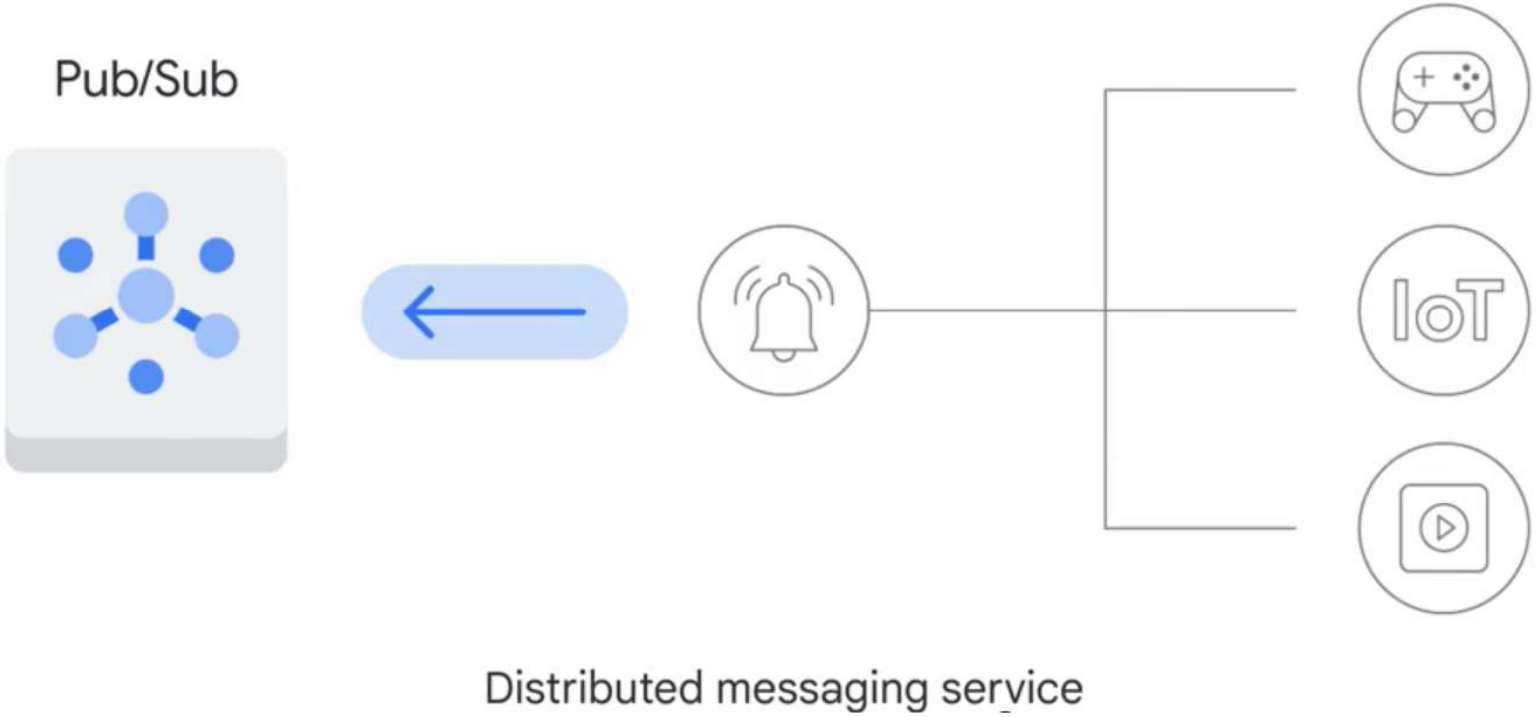

Download
Streaming


4Vs
Variety
Volume
Velocity
Veracity

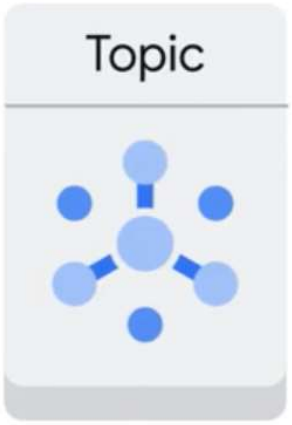


Data pipeline


Google Cloud
Pub/Sub


Pub/Sub
IoT
Distributed messaging service

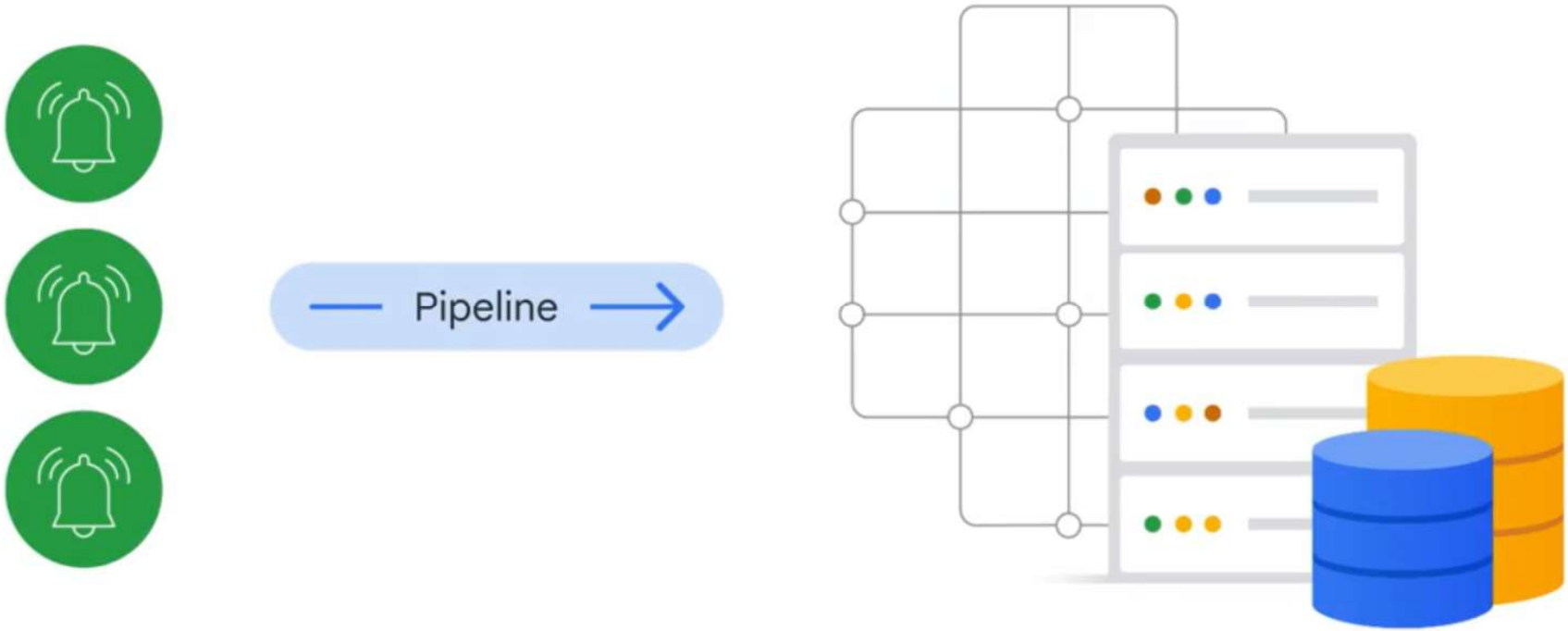
Pub/Sub

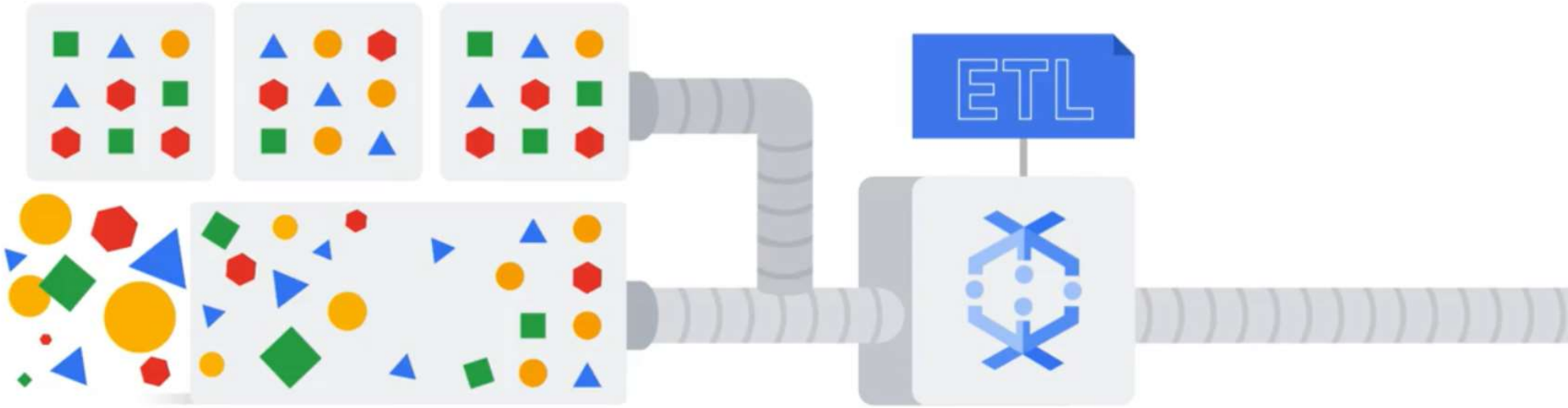


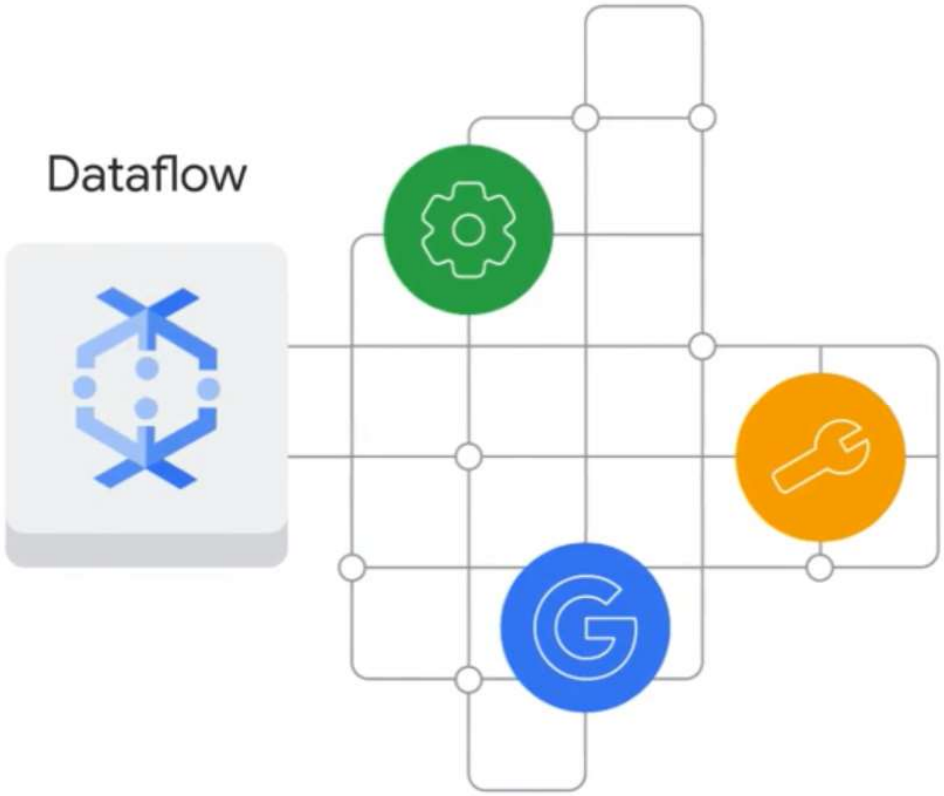
Pub/Sub

Radio

Pub/Sub supports many different inputs and outputs, and you can even publish a Pub/Sub event from one topic to another

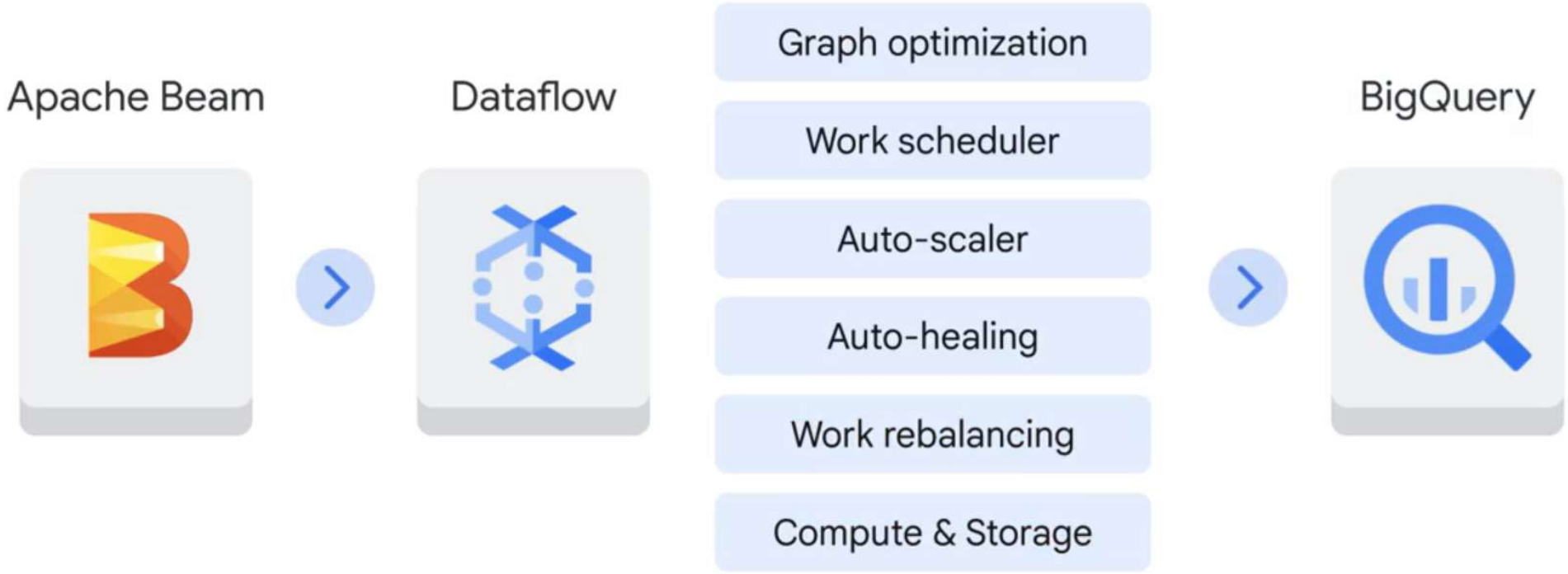

Pipeline


ETL

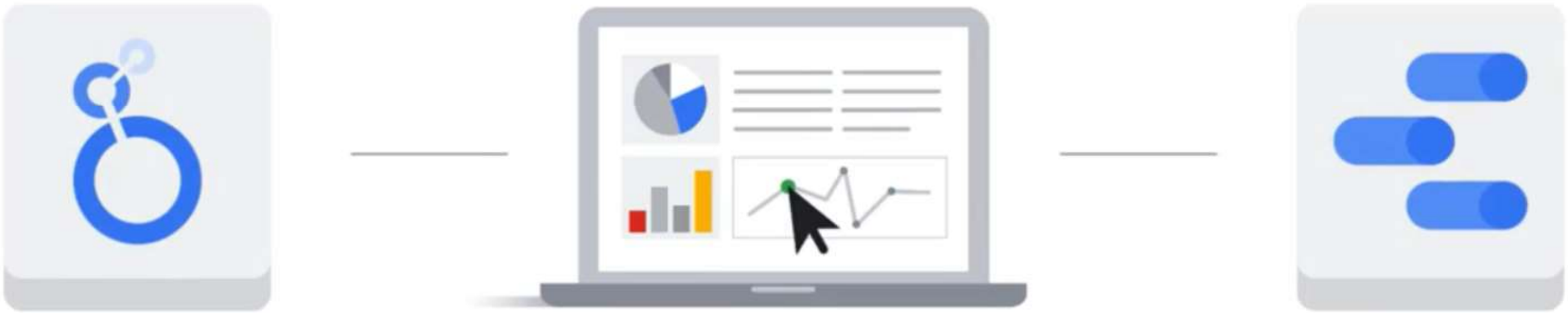


- Reliable auto scaling
- Meet data pipeline demands

Dataflow is serverless and NoOps


Apache Beam
Dataflow
Graph optimization
Work scheduler
Auto-scaler
Auto-healing
Work rebalancing
Compute & Storage
BigQuery

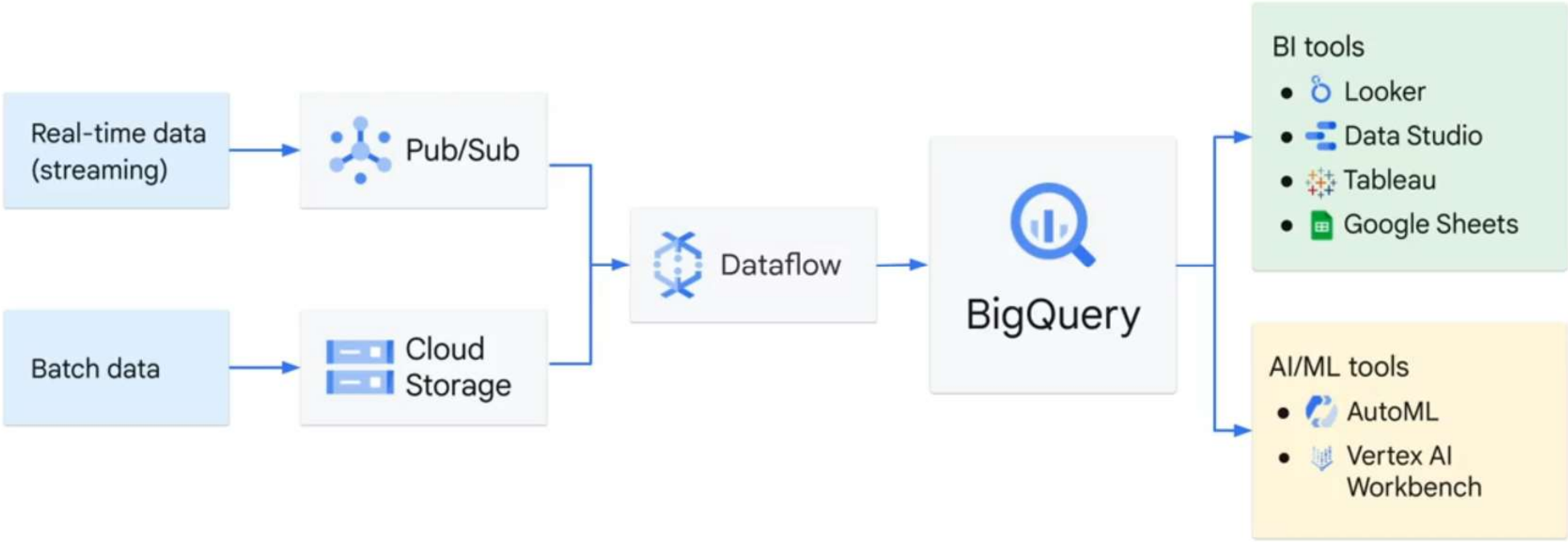
Google Cloud



Looker

Data Studio

BigQuery is a fully-managed data warehouse

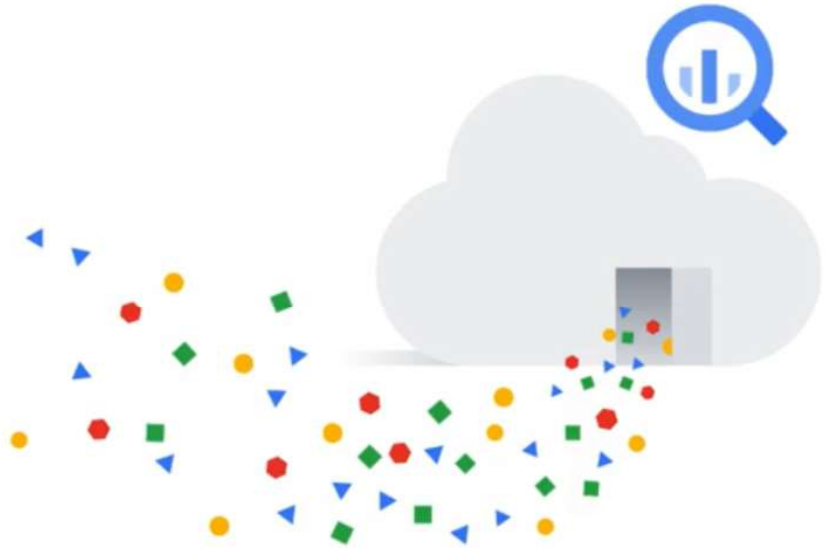

Real-time data (streaming)
Pub/Sub
Batch data
Cloud Storage
Dataflow
BigQuery
BI tools
Looker
Data Studio
Tableau
Google Sheets
AI/ML tools
AutoML
Vertex AI Workbench

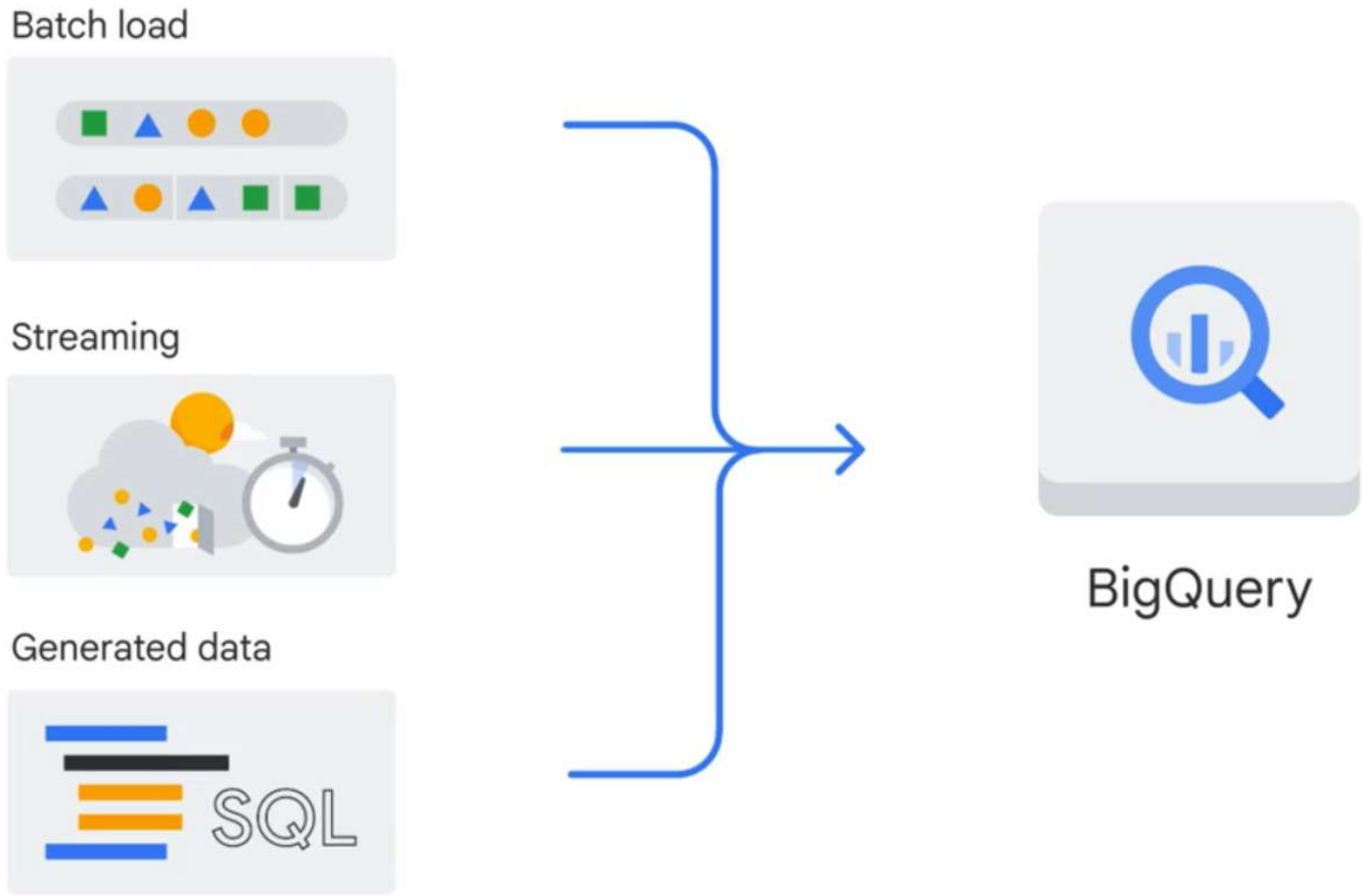
Internal data

External data

Multi-cloud data

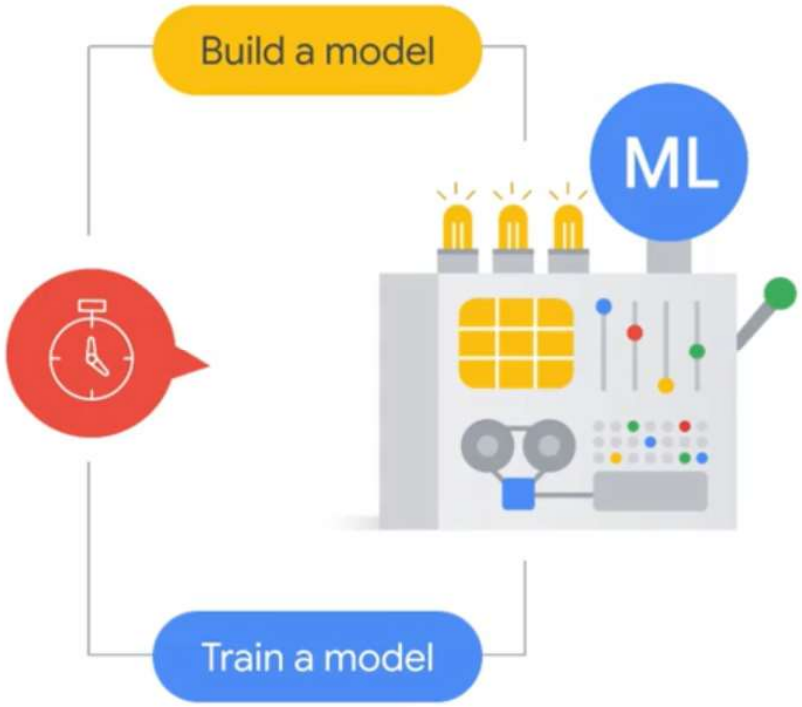
Public datasets




Batch load
Streaming
Generated data
SQL
BigQuery

BigQuery analytics features

- Ad hoc analysis
- Geospatial analytics
- Building machine learning models
- Building BI dashboards

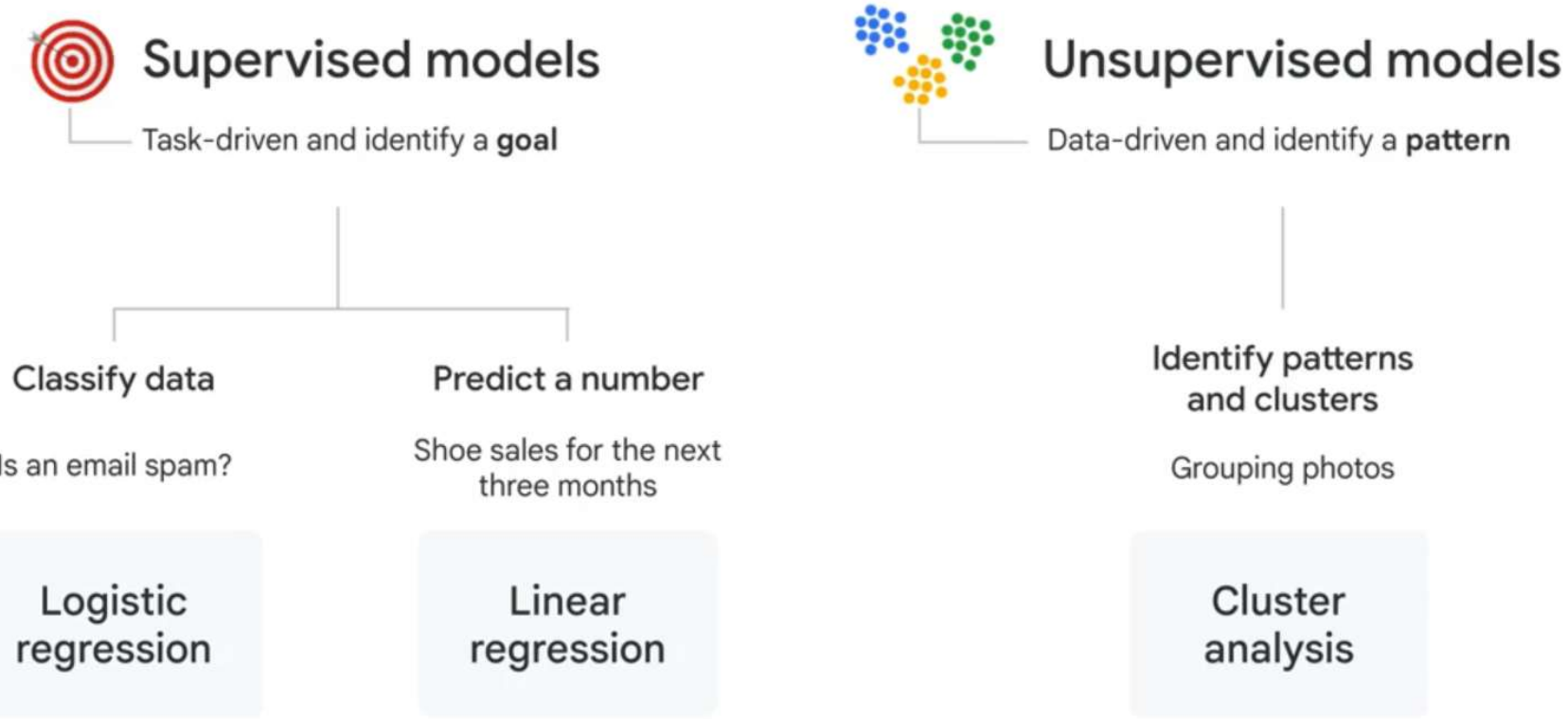


01 Export data from your datastore into an integrated development environment (IDE)

02 Transform the data and perform feature engineering

03 Build the model in TensorFlow and train it locally or on a virtual machine

\+ To improve model performance, you need to get more data and create new features


Supervised models
Task-driven and identify a goal
Classify data
Is an email spam?
Logistic regression
Predict a number
Shoe sales for the next three months
Linear regression
Unsupervised models
Data-driven and identify a pattern
Identify patterns and clusters
Grouping photos
Cluster analysis

Key phases of a machine learning project

01 Extract, transform, and load data into BigQuery

02 Select and preprocess features

03 Create the ML model inside BigQuery

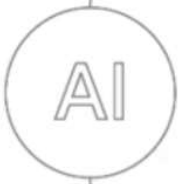
04 Evaluate the performance of the trained model

05 Use the model to make predictions

BigQuery ML

Why should I trust Google for artificial intelligence and machine learning?

Google



Google is an AI-first company.

A leader across industries thanks to contributions in AI and ML.

	01 BigQuery ML	02 Pre-built APIs	03 AutoML	04 Custom training
Data type	Tabular	Tabular, image, text, and video	Tabular, image, text, and video	Tabular, image, text, and video
Training data size	Medium to large	No data required	Small to medium	Medium to large
ML and coding expertise	Medium	Low	Low	High
Flexibility to tune hyperparameters	Medium	None	None	High
Time to train a model	Medium	None	Medium	Long


Familiar with SQL and have data in BigQuery
01
BigQuery ML
Have little ML expertise
02
Pre-built APIs
Want to build custom models with your own training data with minimal coding
03
AutoML
Want full control of the ML workflow
04
Custom training

Pre-built APIs

Offered as services

Building datasets

Curating datasets

Training datasets

Vision API

Based on

Google's image datasets

Speech-to-Text API

Trained on

YouTube captions

Translation API

Built on

Google's neural machine translation technology


AutoML
Automated
machine
learning

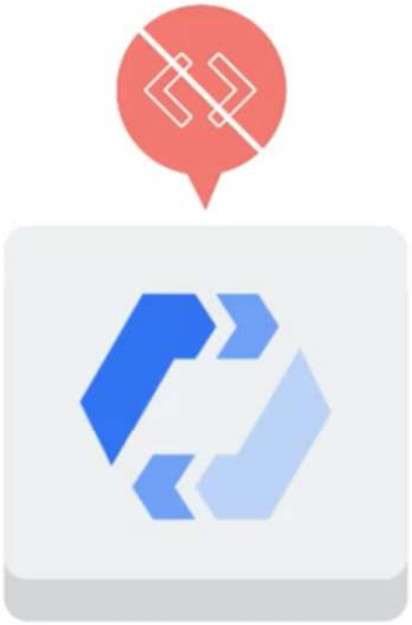
Transfer learning

Train custom ML models with

- Minimal effort
- Little machine learning expertise

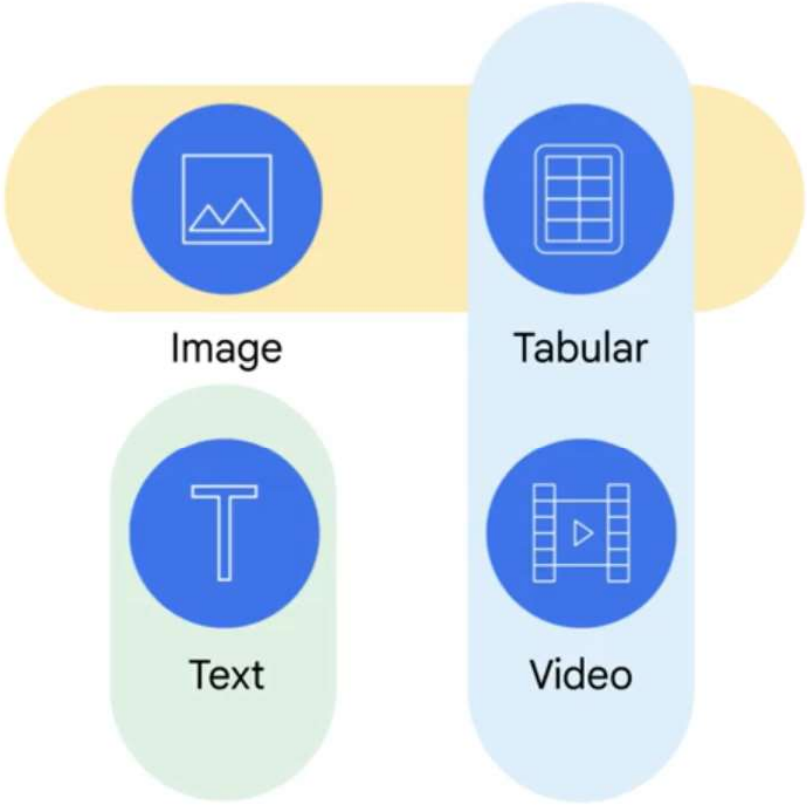
Allows data scientists to focus on

- Defining business problems
- Evaluating and improving model results



AutoML

AutoML



Vertex AI

- AutoML
- Custom training

Machine learning

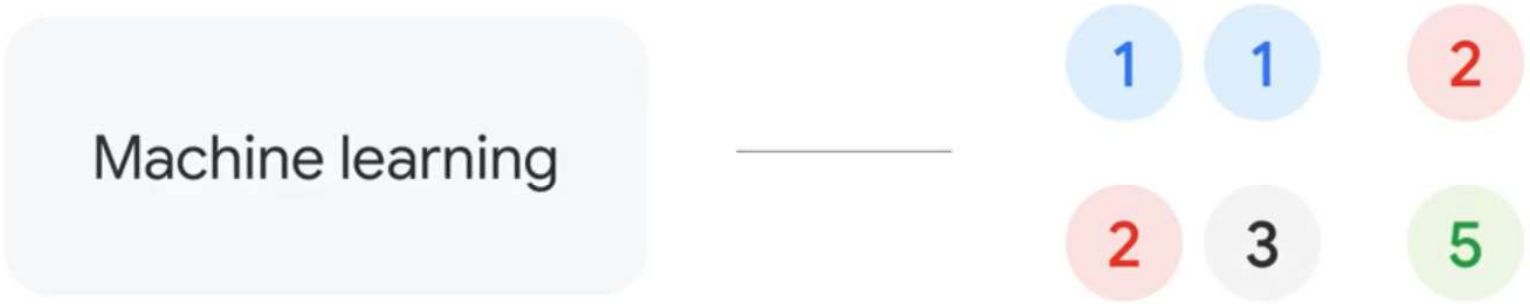
vs

Traditional programming

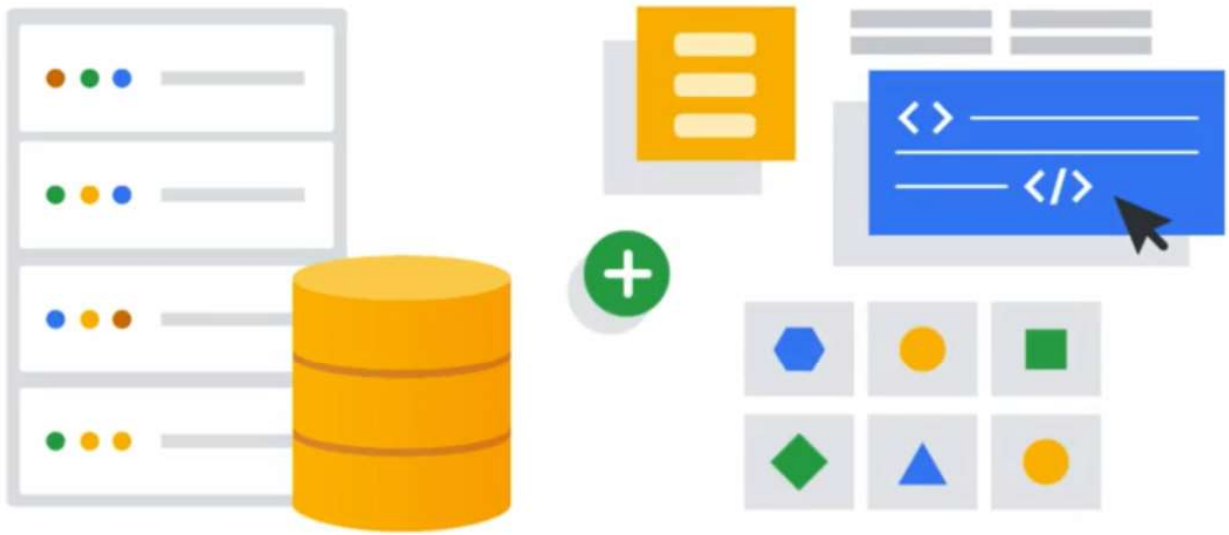
1 + 1 = 2

Data + rules = answers

Traditional programming


Machine learning
1
1
2
2
3
5

How is it possible that a machine can actually learn to solve puzzles?



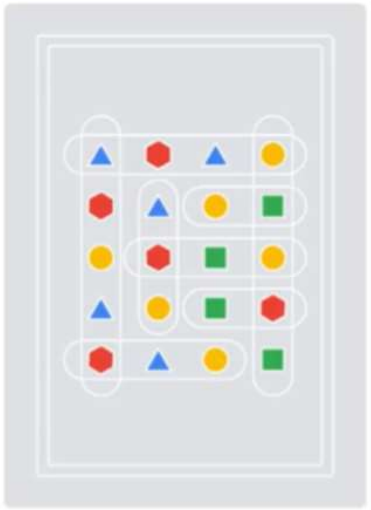
Lots of storage

Fast calculations

Cloud Storage

- Search functionality
- Photo albums

Data preparation



Two steps

- Data uploading
- Feature engineering

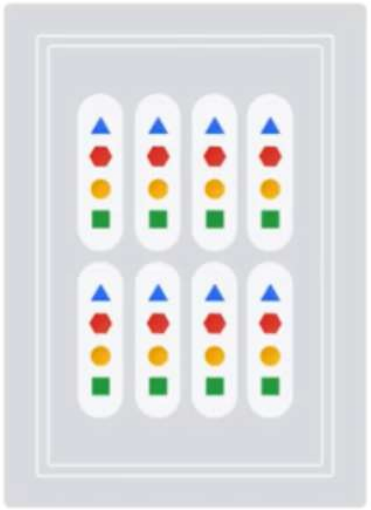
Data types

- Steaming vs. batch data
- Structured vs. unstructured data

Model training

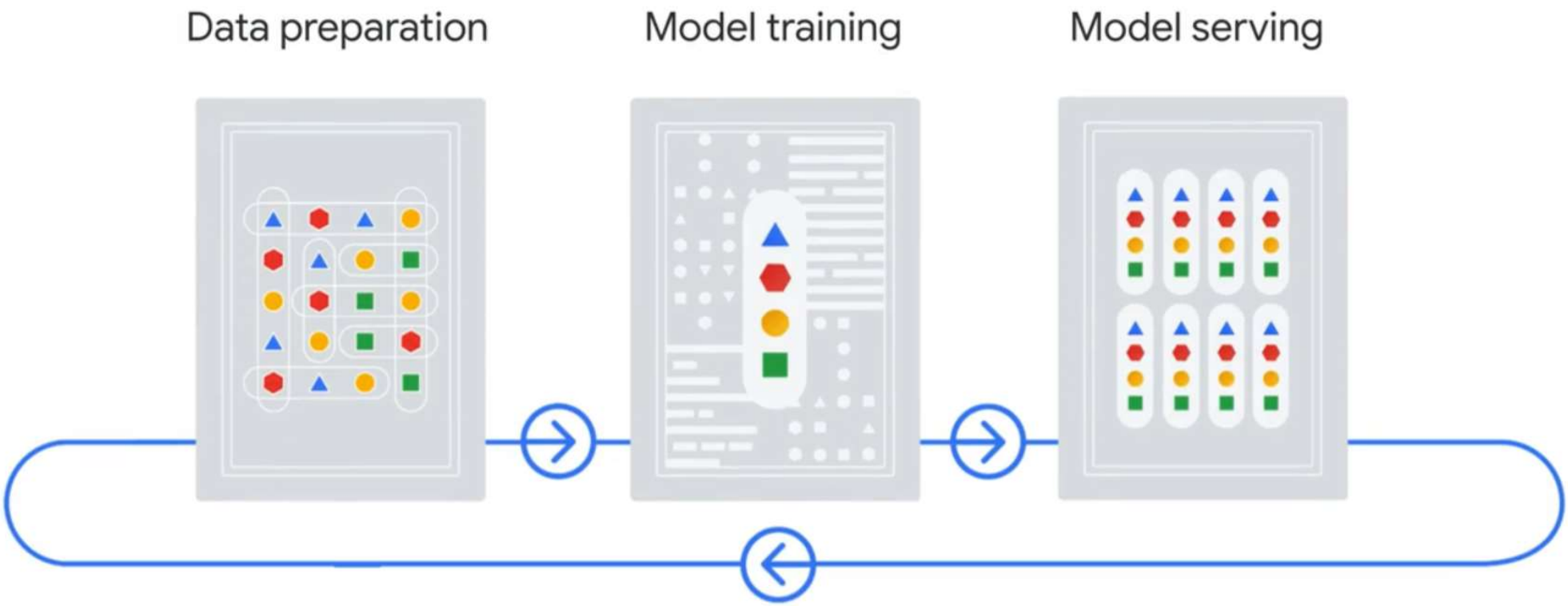
A model needs a tremendous amount of iterative training

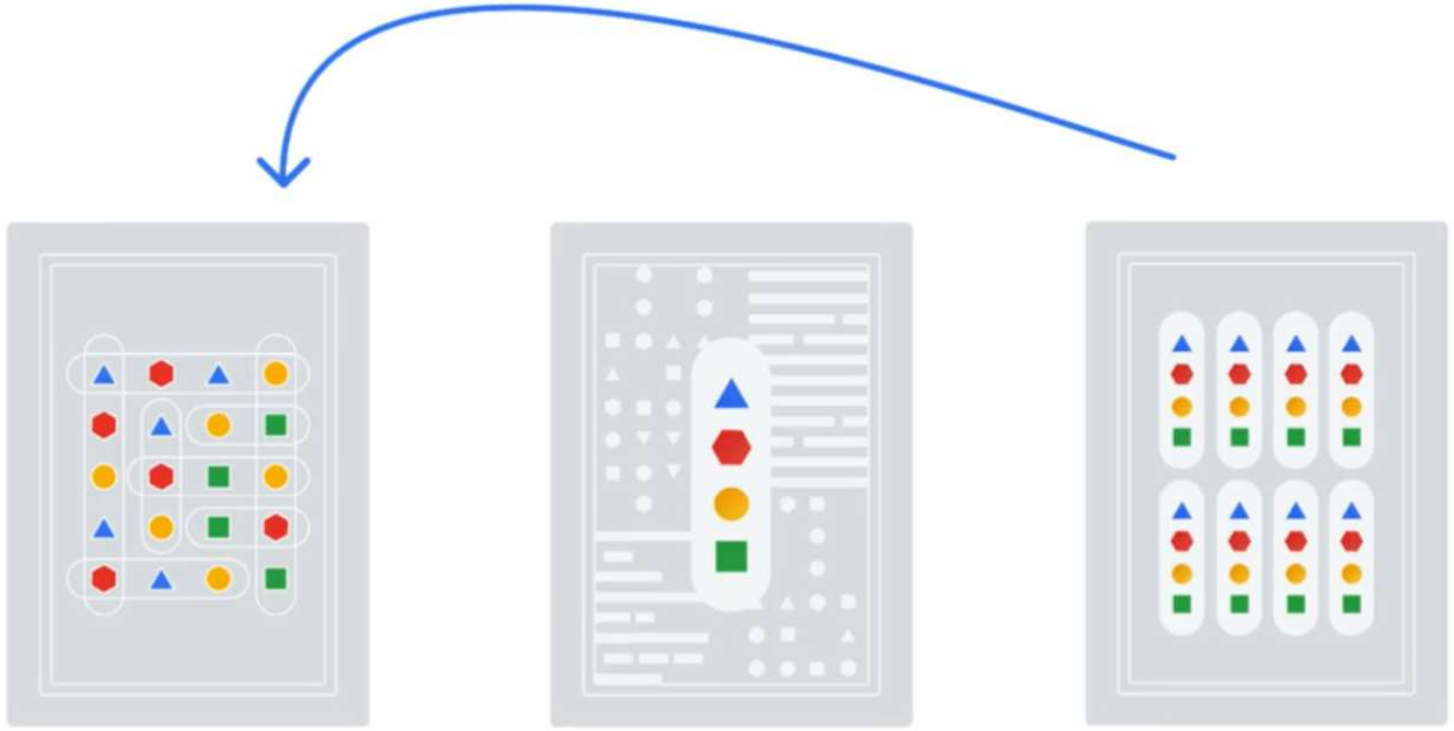
Model serving



A model needs to actually be used in order to predict results.

- Deployed
- Monitored
- Managed


Data preparation
Model training
Model serving



Vertex AI

1. AutoML: codeless
2. Custom training: code-based

Vertex AI

Feature Store

A centralized repository for organizing, storing, and serving features to feed to training models

Vizier

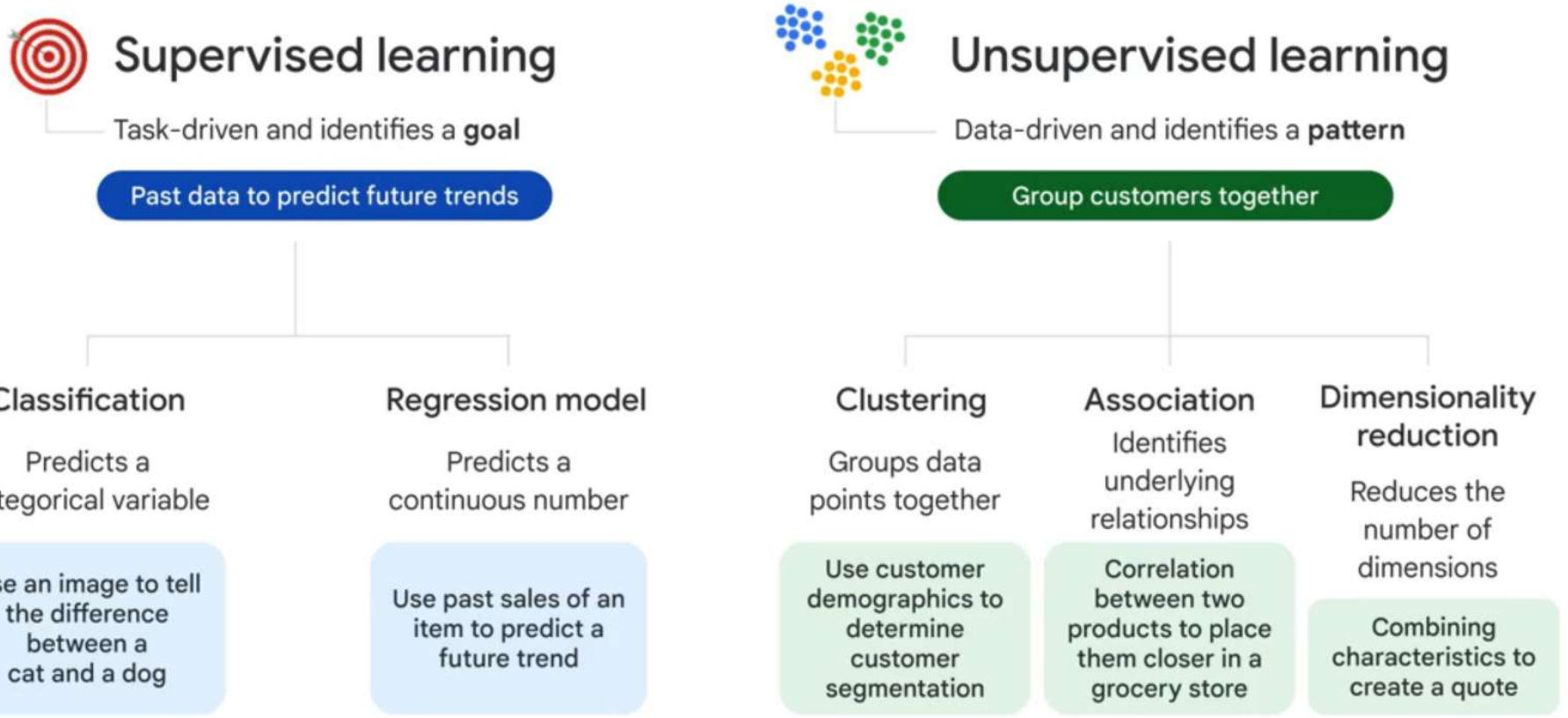
Helps tune hyperparameters in complex machine learning models

Explainable AI

Helps with things like interpreting training performance

Pipelines

Help monitor the ML production line

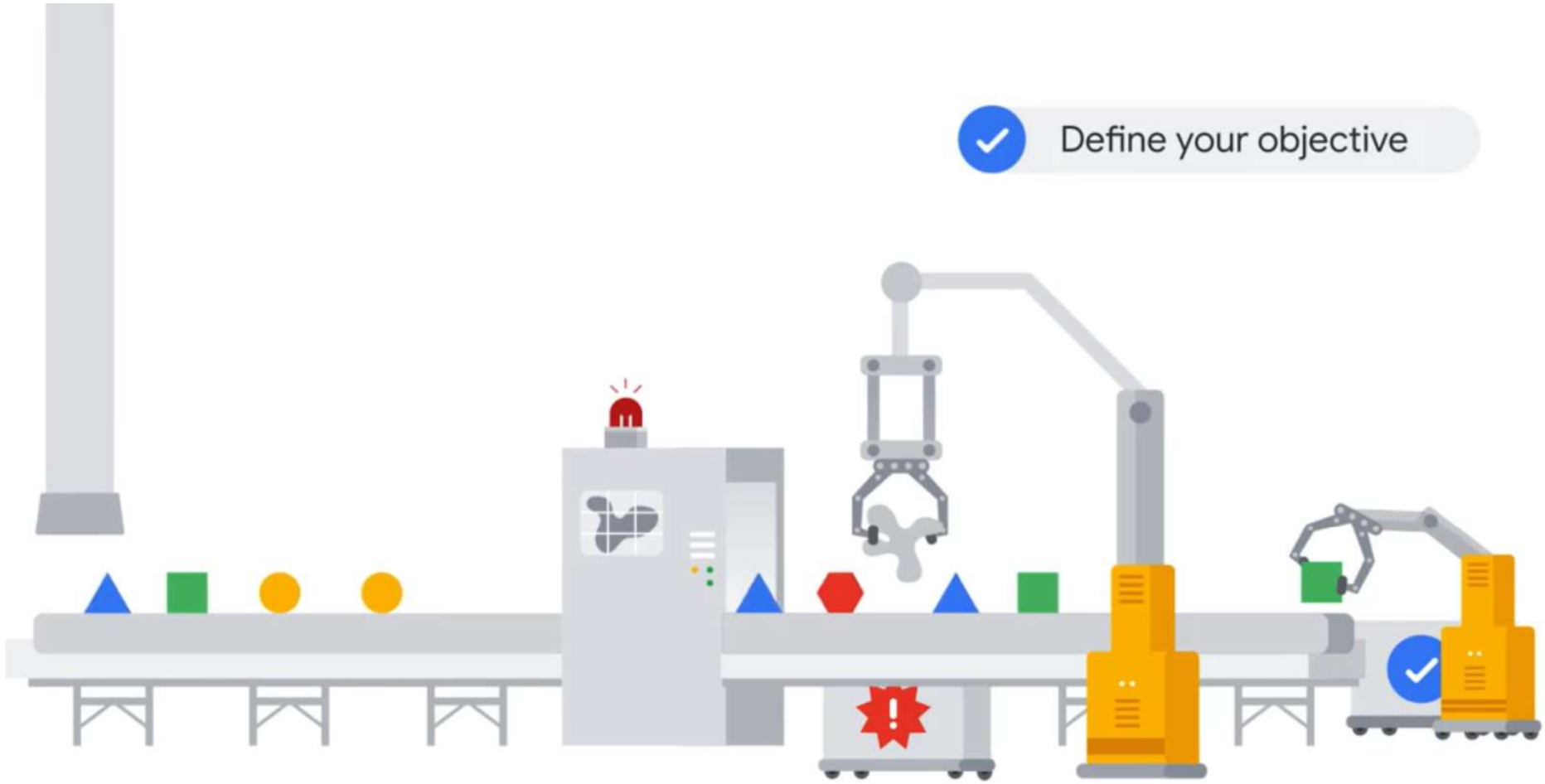

Supervised learning
Task-driven and identifies a goal
Past data to predict future trends
Classification
Predicts a categorical variable
Use an image to tell the difference between a cat and a dog
Regression model
Predicts a continuous number
Use past sales of an item to predict a future trend
Unsupervised learning
Data-driven and identifies a pattern
Group customers together
Clustering
Groups data points together
Use customer demographics to determine customer segmentation
Association
Identifies underlying relationships
Correlation between two products to place them closer in a grocery store
Dimensionality reduction
Reduces the number of dimensions
Combining characteristics to create a quote

BigQuery ML

AutoML

Pre-built APIs

Custom training


Define your objective

Model training

Experiment with recipes

01 Model training

02 Model evaluation

Vertex AI

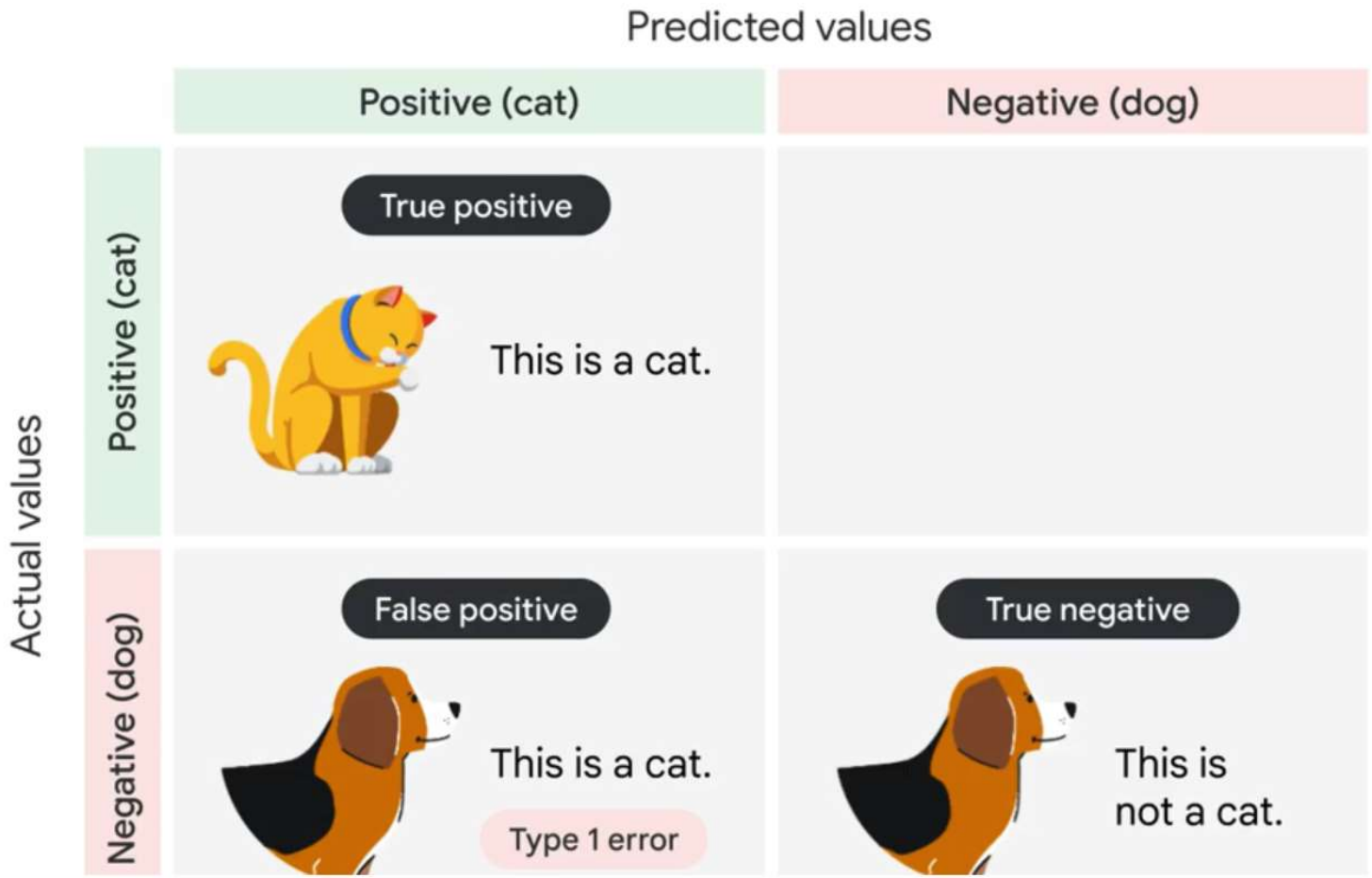
Evaluation metrics

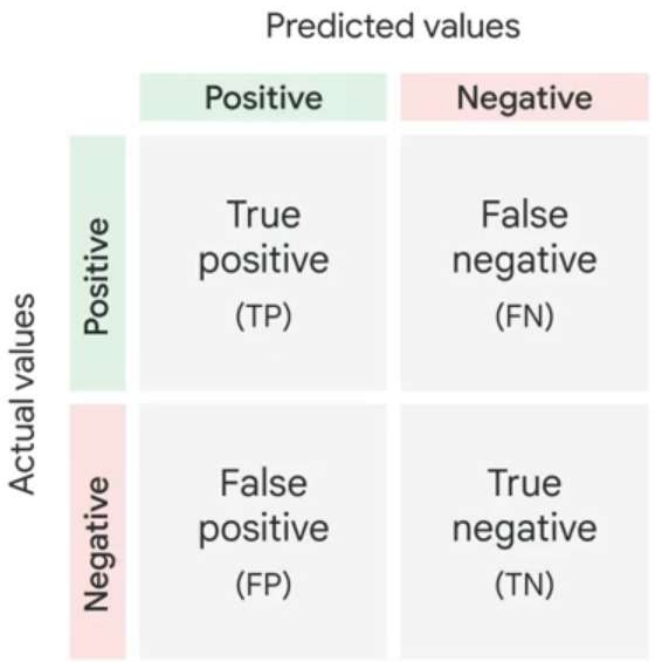
Confusion matrix

Recall

Precision

Feature importance


Predicted values
Positive (cat)
Negative (dog)
Actual values
Positive (cat)
Negative (dog)
True positive
This is a cat.
False positive
This is a cat.
Type 1 error
True negative
This is
not a cat.

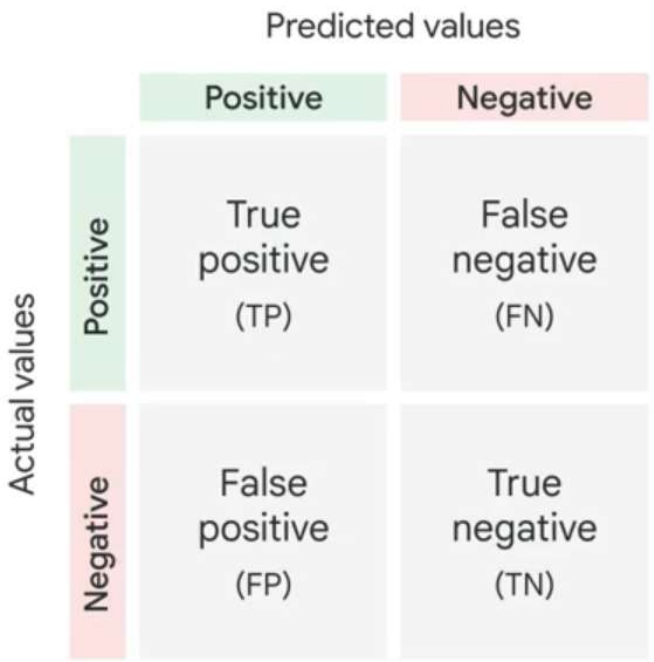


$$\text{Recall} = \frac{TP}{TP+FN}$$

Recall

Refers to all the positive cases, and looks at how many were predicted correctly.

Precision



Recall

Refers to all the positive cases, and looks at how many were predicted correctly.

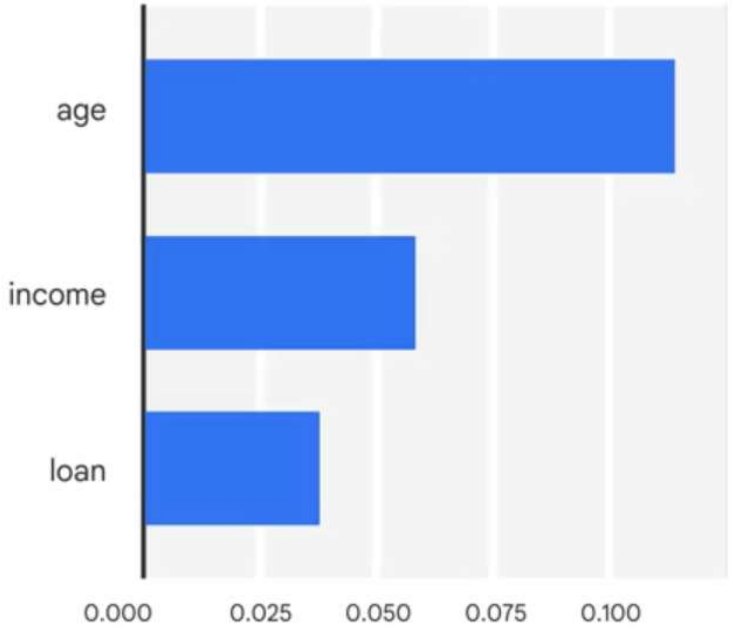
Precision

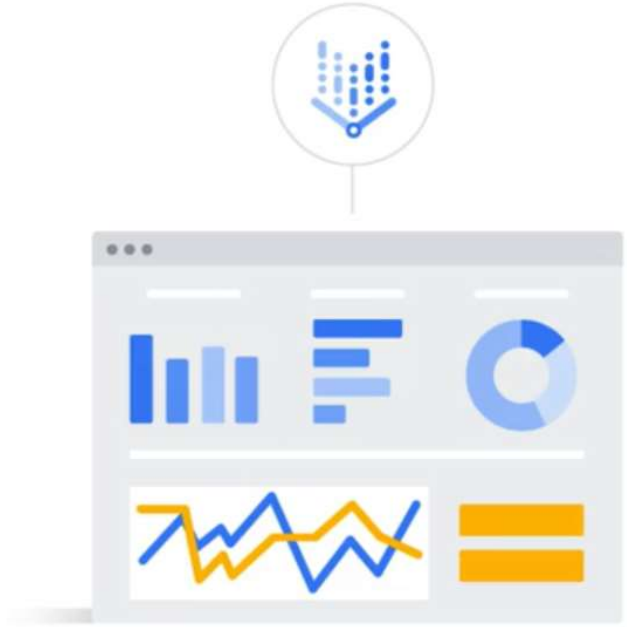
Refers to all the cases predicted as positive, and how many are actually positive.

$$\text{Precision} = \frac{TP}{TP+FP}$$

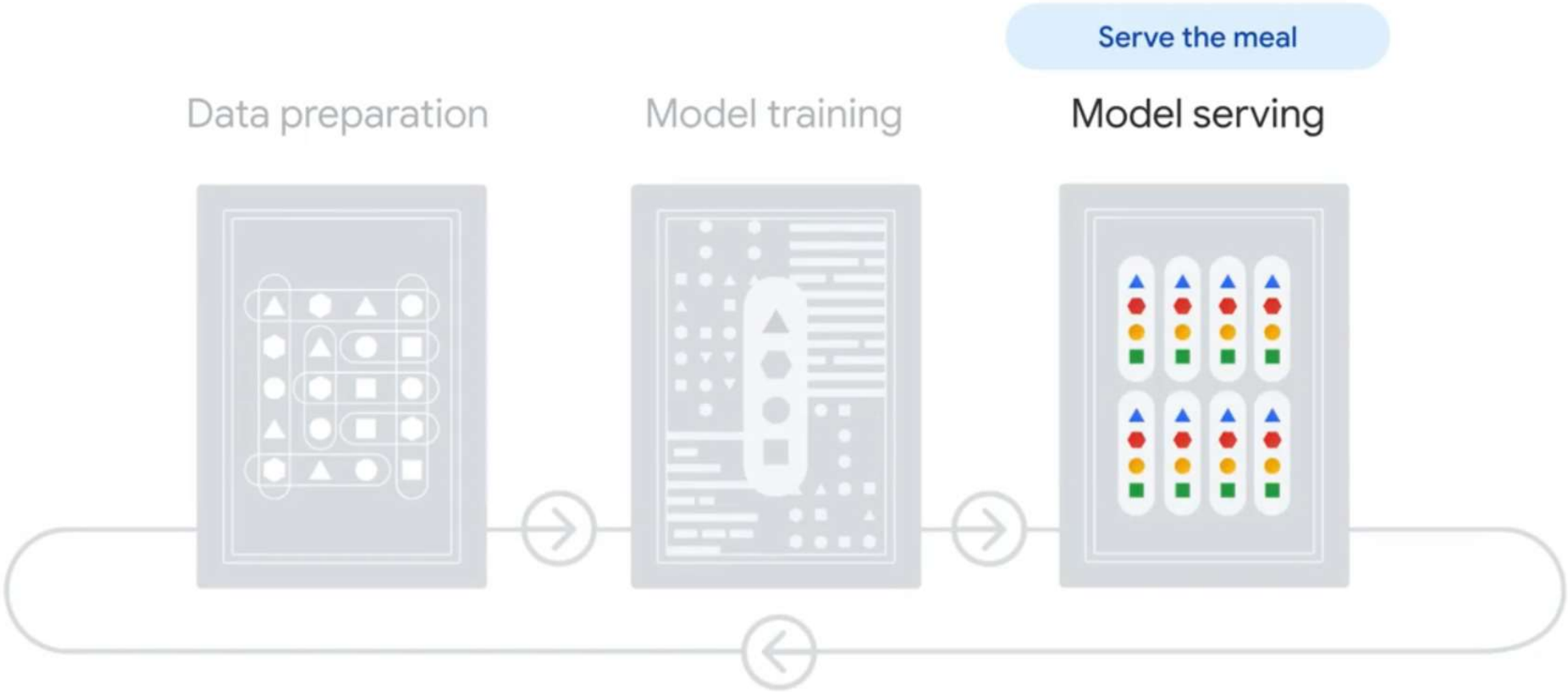
Confusion matrix

Feature importance





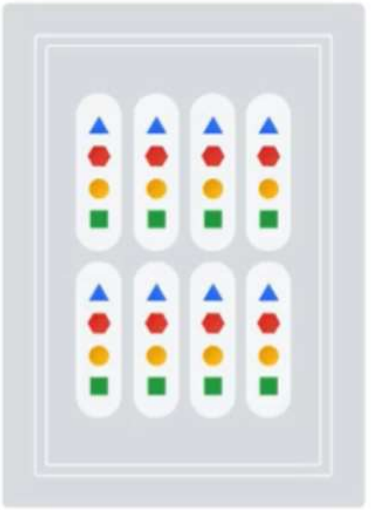
Explainable AI


Serve the meal
Data preparation
Model training
Model serving

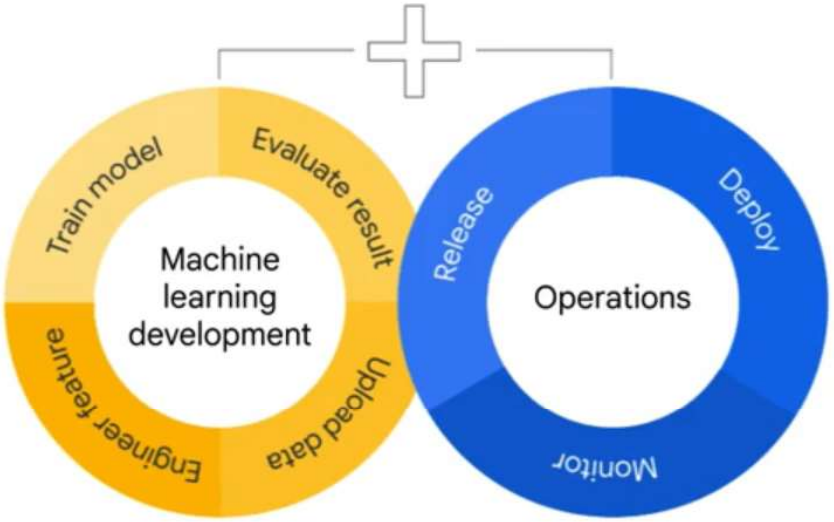
01 Model deployment

02 Model monitoring

Model serving



- Solve production challenges
- Building an integrated ML system
- Operating it in production



Machine learning operations

What does MLOps have
to do with model serving?

01

Model deployment

Endpoint

Best when immediate results with low latency are needed.

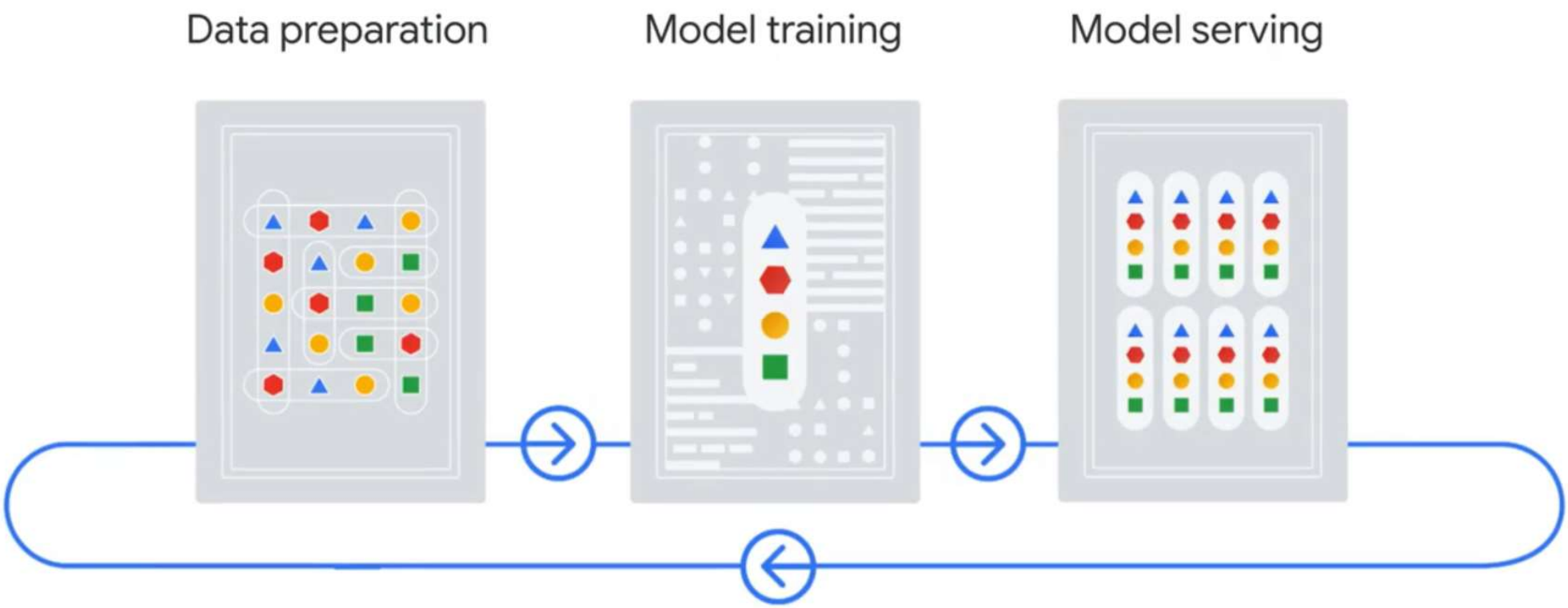
Must be deployed to an endpoint before that model can be used to serve real-time predictions.

Batch prediction

Best when no immediate response is required, and accumulated data should be processed with a single request.

Offline prediction

Best when the model should be deployed in a specific environment off the cloud.


Data preparation
Model training
Model serving

Thanks