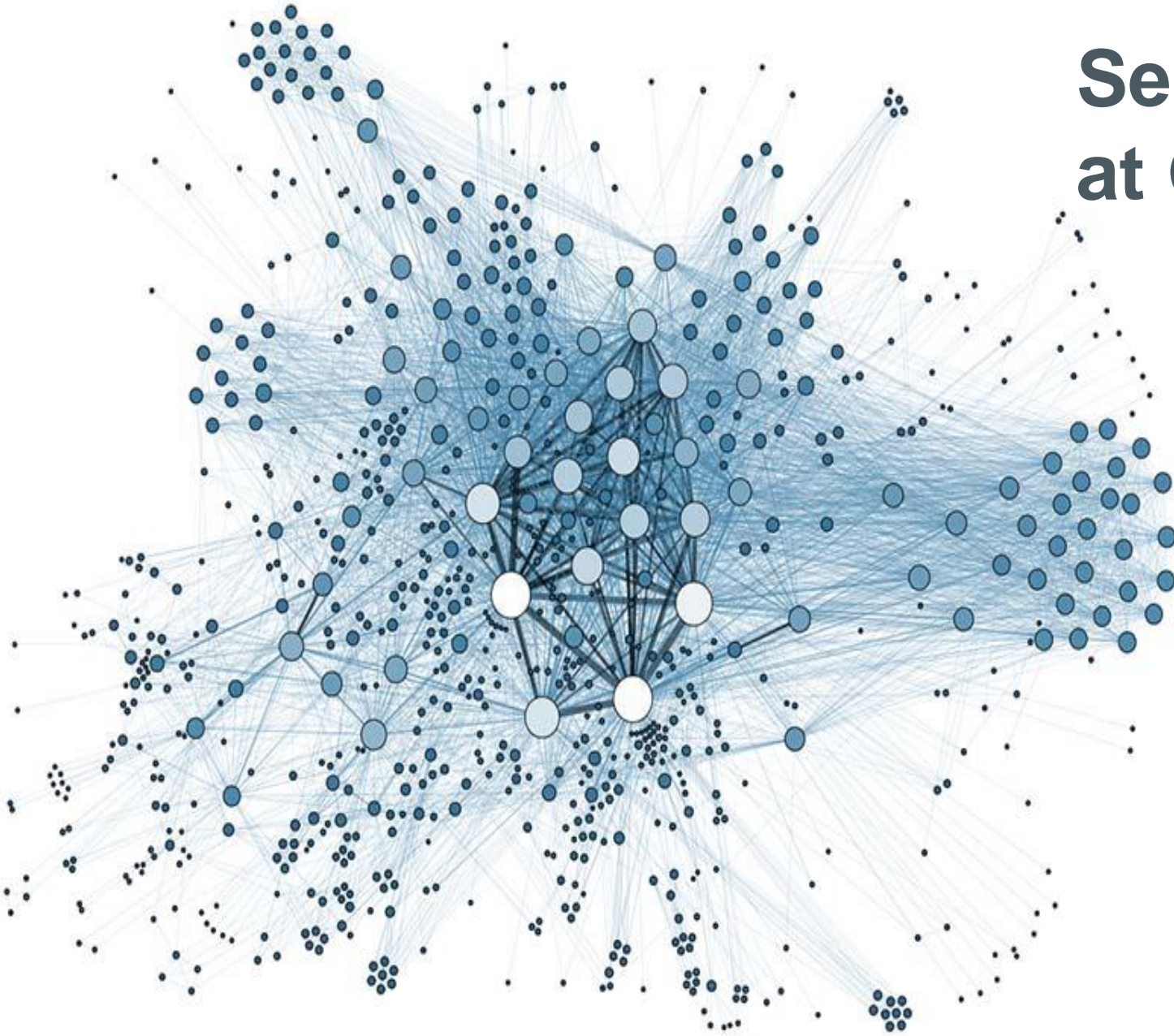# Services for Data Science at Google Cloud

Todi Thanasi & Ozge Koroglu
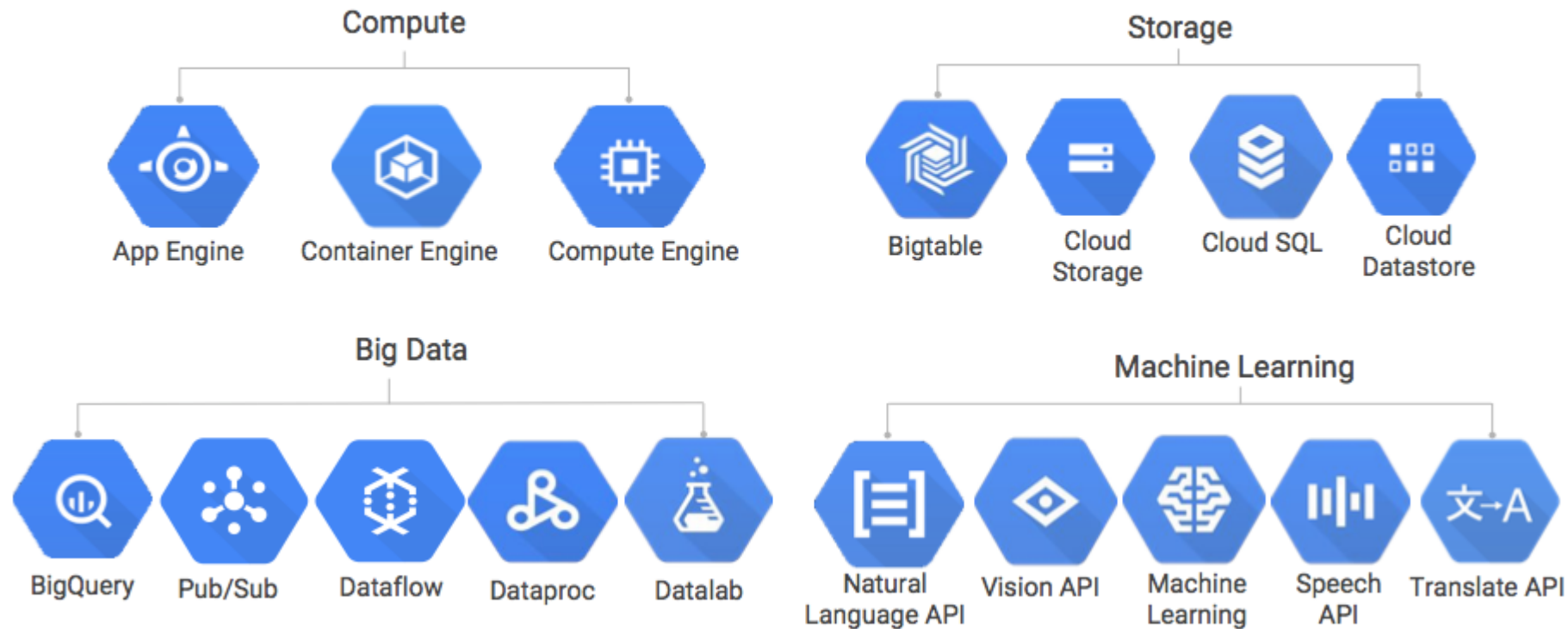
Why Google Cloud?

- Good UX & Clean API

- Implements community tools (Apache Beam, Jupyter, ..)

- Integrates with Google tools (TenserFlow, Kubernetes,..)

- Good price (storage) and per minute billing

- ISO security certifications

Data Science On Google Cloud

Google Cloud Platform

Compute
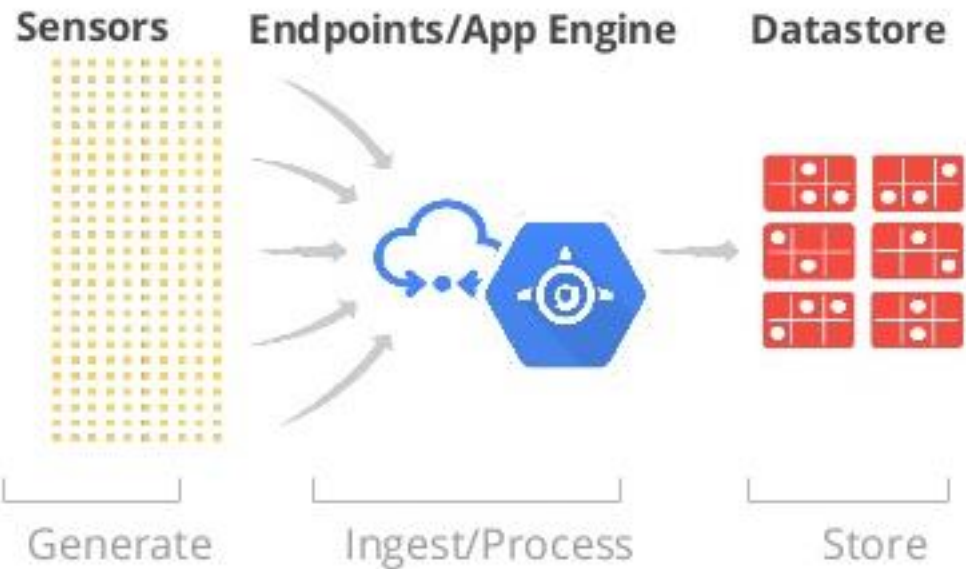- App Engine
- Container Engine
- Compute Engine

Storage
- Bigtable
- Cloud Storage
- Cloud SQL
- Cloud Datastore

Big Data
- BigQuery
- Pub/Sub
- Dataflow
- Dataproc
- Datalab

Machine Learning
- Natural Language API
- Vision API
- Machine Learning
- Speech API
- Translate API

Data Collection

App Engine

Sensors — **Endpoints/App Engine** — **Datastore**

Generate — Ingest/Process — Store

Data Collection

Cloud Pub/Sub

From Twitter to Pub/Sub

ABSORBS SHOCKS AND CHANGES

? → Pub/Sub Topic

RELIABLE AND SCALABLE DELIVERY

DECOUPLES PRODUCER AND CONSUMER(S)

Subscription A → Consumer A
Subscription B → Consumer B
Subscription C → Consumer C

Data Collection

Big Query

Data Collection

Google Cloud Storage
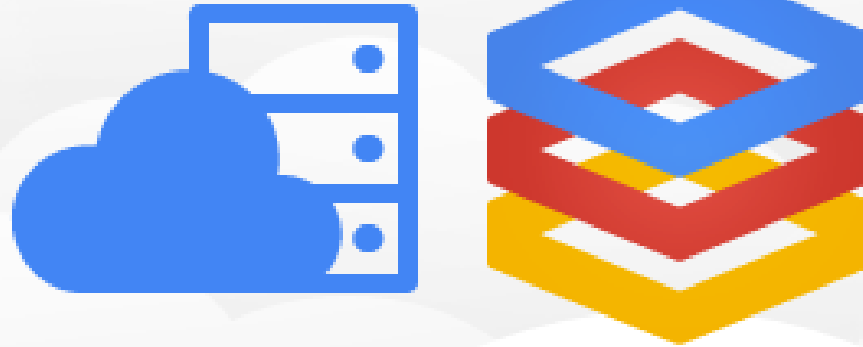
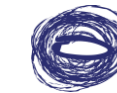Cloud Storage

Data Collection

Stackdriver Logging

Data Cleaning

Google Cloud Dataprep
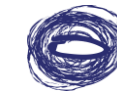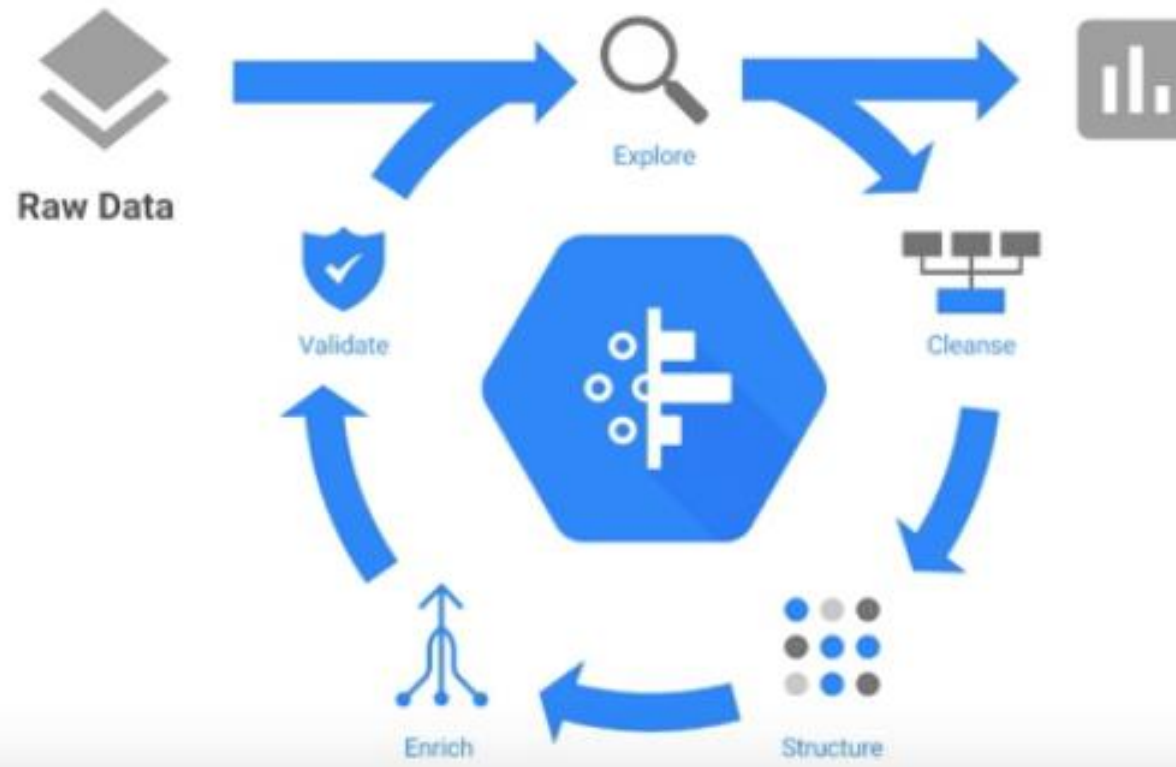
- Visually Exploring, Cleaning & Preparing Data

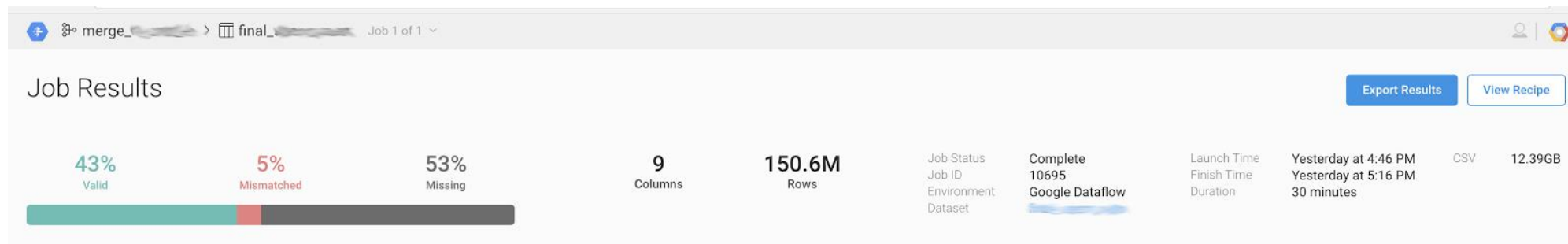- Schema, Type & Missing Data Detection

- Preparing Big Data For Training Machine Learning Models
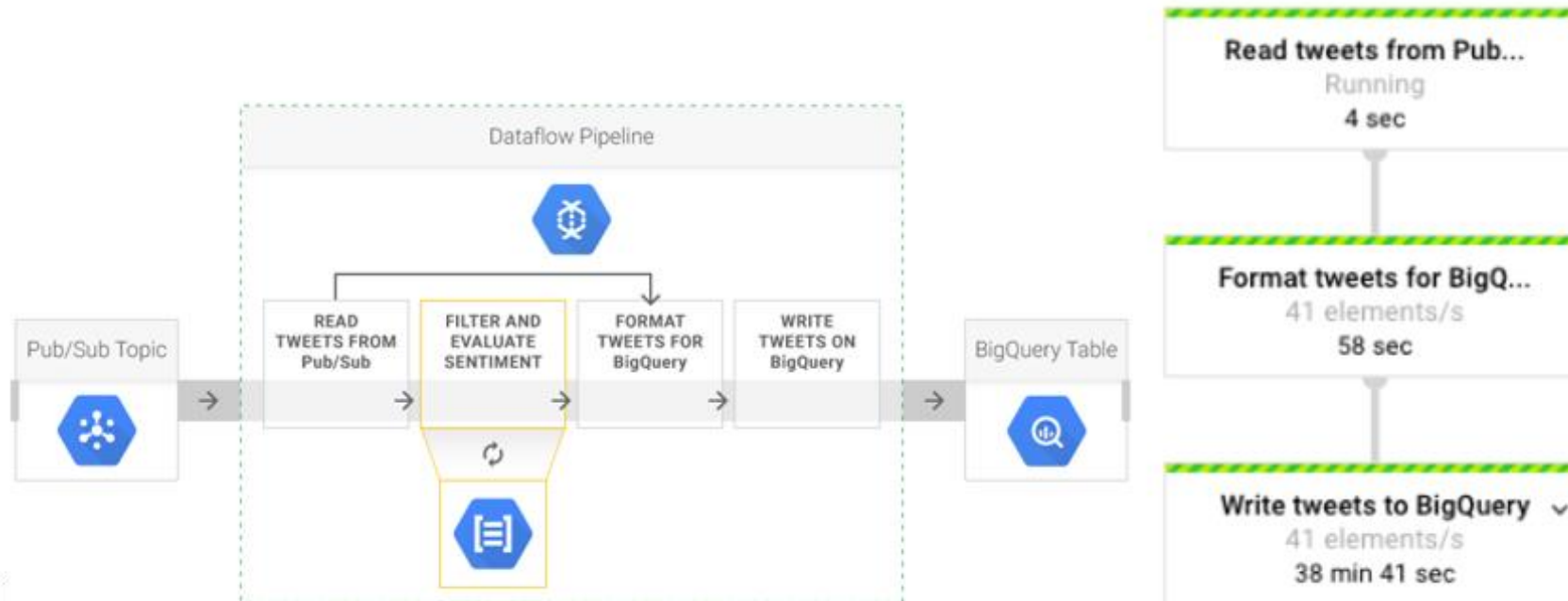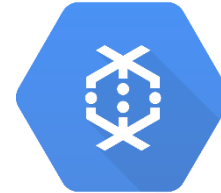
Data Cleaning

Google Cloud Dataprep
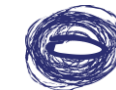
Google Cloud Platform

Data Processing

Google Cloud Dataflow

- Data Processing Pipeline
- Automatized Process
- Serverless

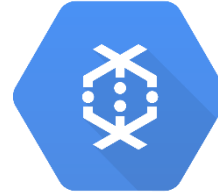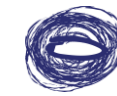# Data Processing

Google Cloud Dataflow

FILTERED

FILTERED & GROUPED

FILTERED, GROUPED & WINDOWED

ANY COMBINATION OF
PRIMITIVE & CUSTOM TRANSFORMATIONS
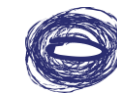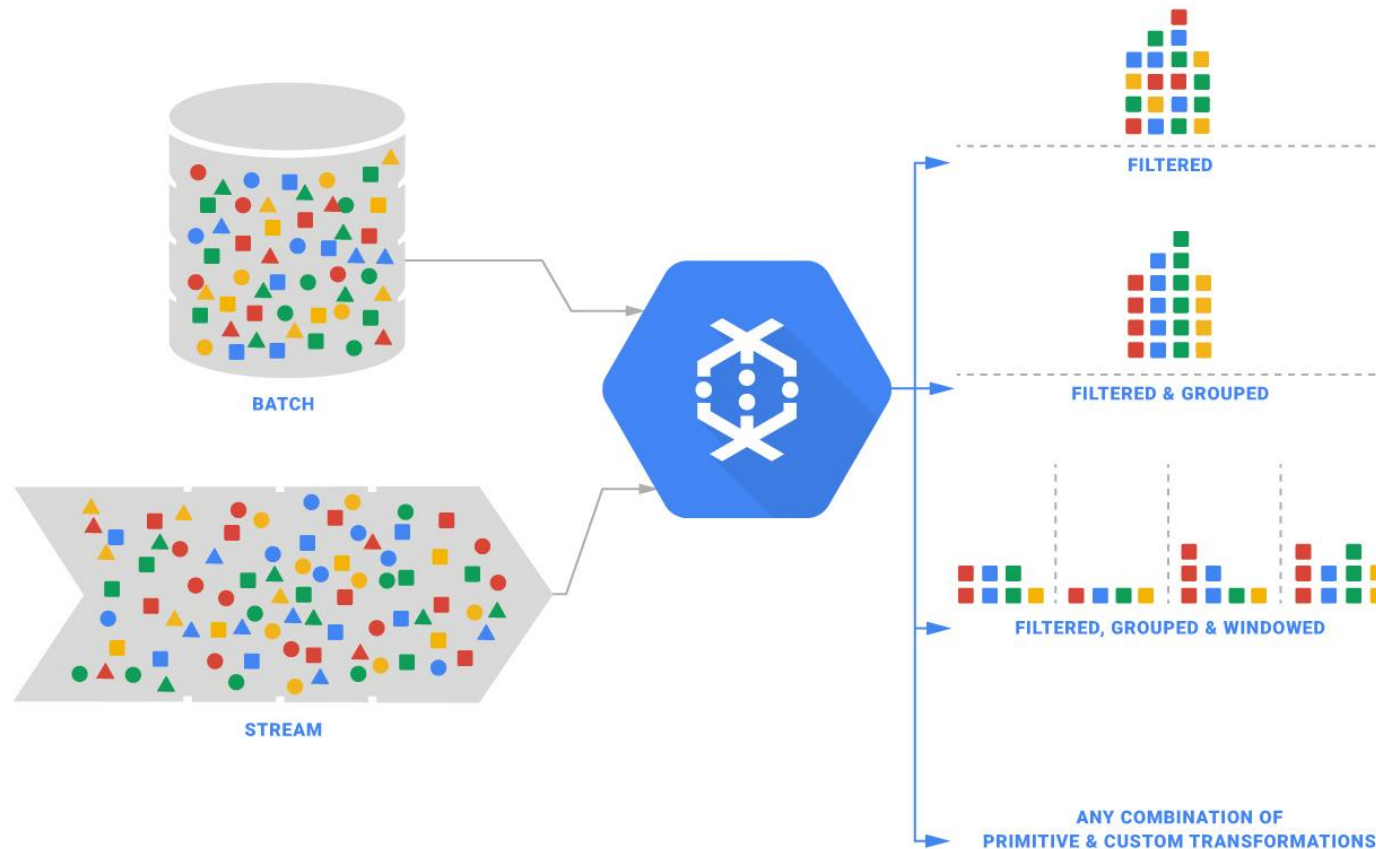
BATCH

STREAM

Accelerate development for **batch** & **streaming**

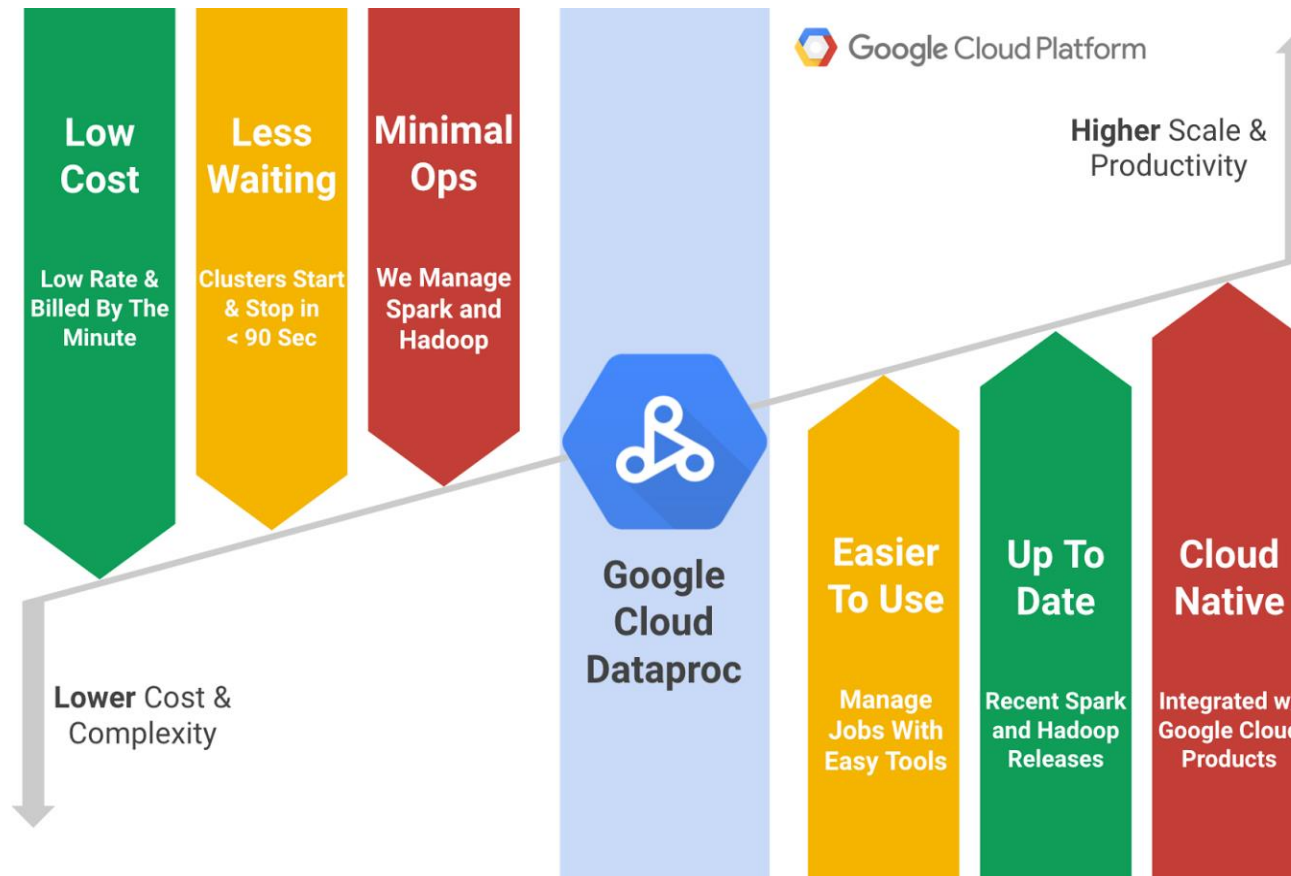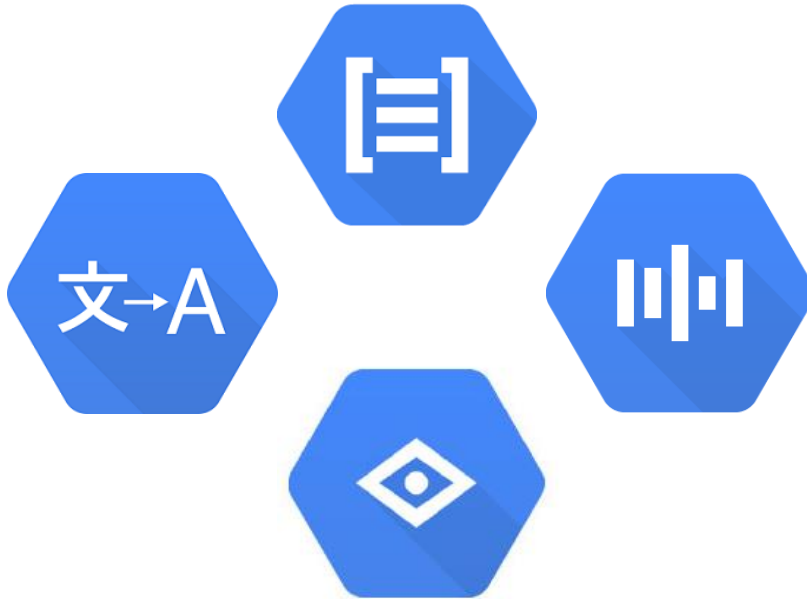Useful For Data Science

Data Processing

Google Cloud Dataproc

- Runs Hadoop & Spark Jobs
- Stateless Hadoop & Spark clusters

Low Cost — Low Rate & Billed By The Minute

Less Waiting — Clusters Start & Stop in < 90 Sec

Minimal Ops — We Manage Spark and Hadoop

Google Cloud Dataproc

Easier To Use — Manage Jobs With Easy Tools

Up To Date — Recent Spark and Hadoop Releases

Cloud Native — Integrated w/ Google Cloud Products

Higher Scale & Productivity

Lower Cost & Complexity

Google Cloud Platform

Data Extraction and Transformation

Powerful Image Analysis

Powerful Speech Recognition

Powerful Text Analysis

Fast, Dynamic Translation

Machine Learning API's
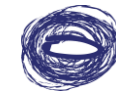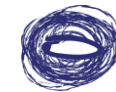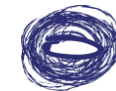
Google Cloud Translation API

文→A

Google Cloud Platform

Hello你好

- Translate Many Languages
- Language Detection
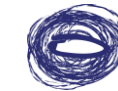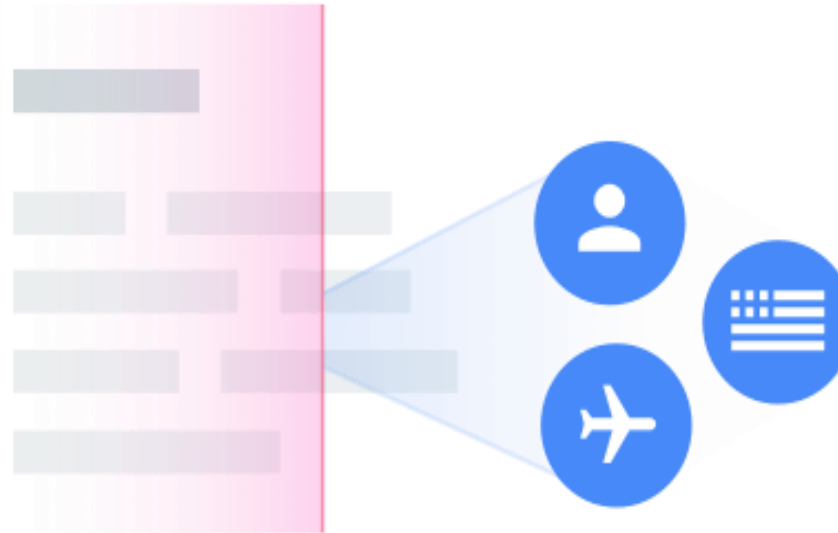- Simple Integration
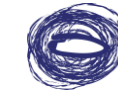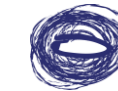
Data Extraction and Translation

Google Cloud Natural Language API
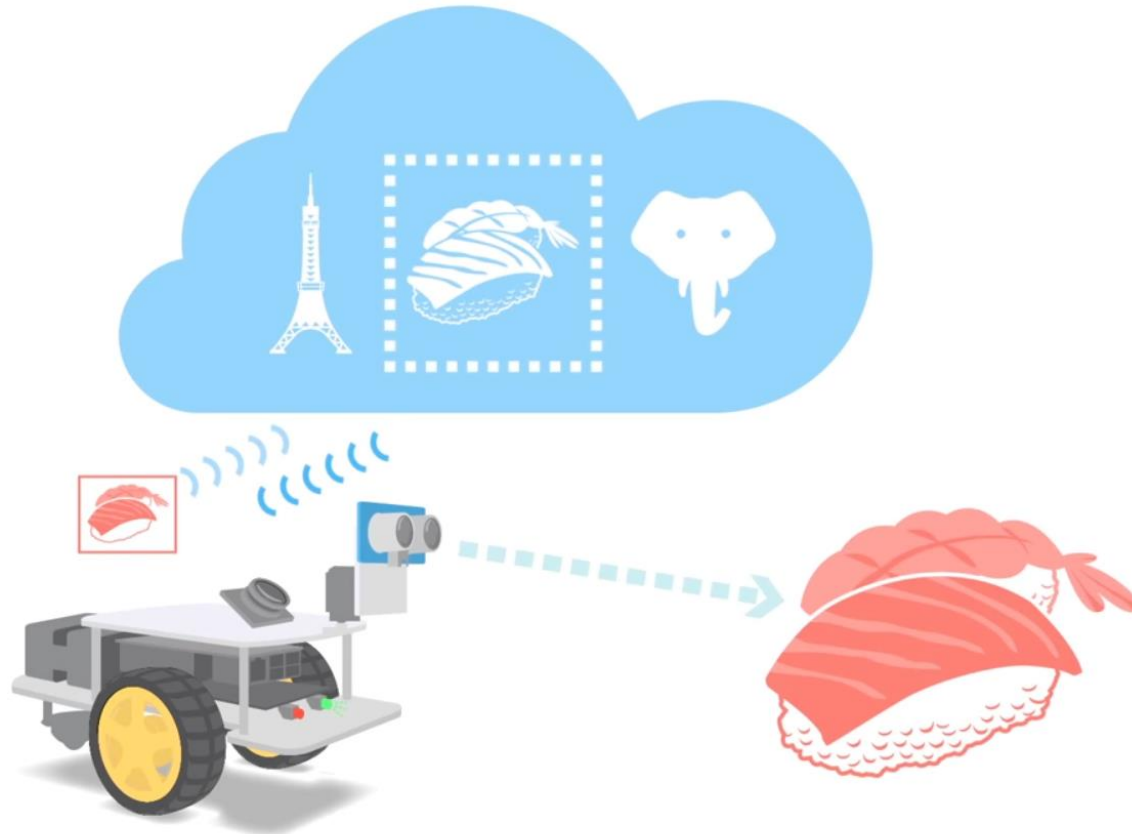
Google Cloud Platform
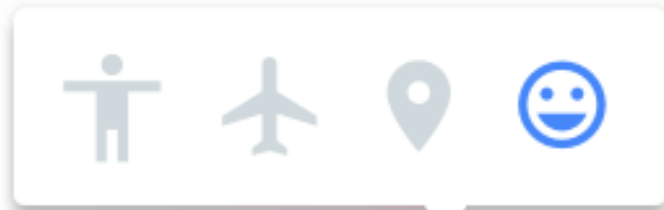
- Extracting Entities
- Sentimental Analysis
- Syntax Analysis

Data Extraction and Translation

Google Cloud Vision API
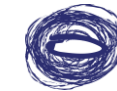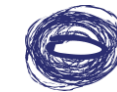
Google Cloud Platform
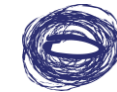
- Facial Detection
- Label Detection
- Landmark Detection
- Explicit Content Detection

Data Extraction and Translation

Google Cloud Vision API
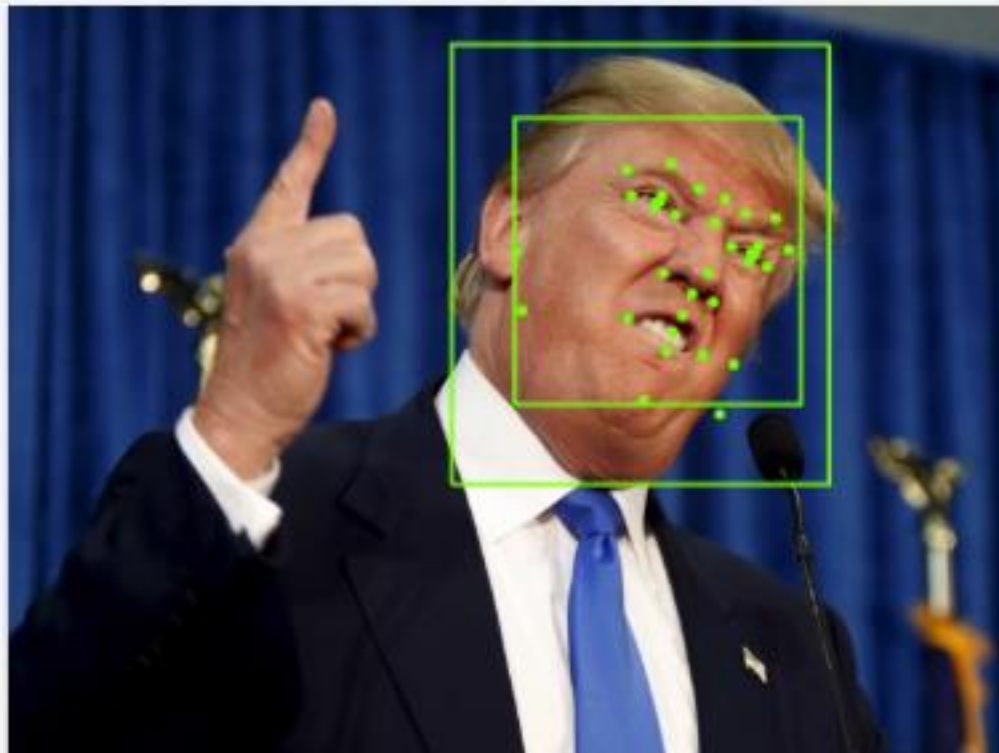
Joy		Very Unlikely
Sorrow		Very Unlikely
Anger		Possible
Surprise		Very Unlikely
Exposed		Very Unlikely
Blurred		Very Unlikely
Headwear		Unlikely

Roll: 20°   Tilt: 21°   Pan: 20°

Confidence   85%

trump-600x450.jpg

Data Extraction and Translation

Google Cloud Speech-to-Text

hello

Thanks!

Any questions?

Google Cloud Platform