Happy Aloha Friday!

Workshop #1
Data Analytics & Visualization

Daniel Liu, Google

April 24, 2020

hacc.hawaii.gov

Happy Aloha Friday!

Welcome from ETS

Marc Masuno
Cyber Security Manager

April 24, 2020

hacc.hawaii.gov

Logistic

Google Cloud

Google Meet

Option 1:

[Join Hangouts Meet](#)

Meeting ID
[meet.google.com/odb-krud-dvu](#)

Option 2:
Phone Numbers   (US) [475-329-7374](#)
PIN: 438 611 547#

Google Meet

Add people

Daniel Liu (You)

Daniel Liu

Daniel Liu

You 11:11 AM
I have a question

I have another question !

Turn on captions   Present now

The Google Team

Daniel Liu
Cloud Customer Engineer

danieltliu@google.com

Amanda Stange
Account Executive

amandastange@google.com

Rob Grace
Cloud customer Engineer

robgrace@google.com

Google Cloud

The HACC Committee
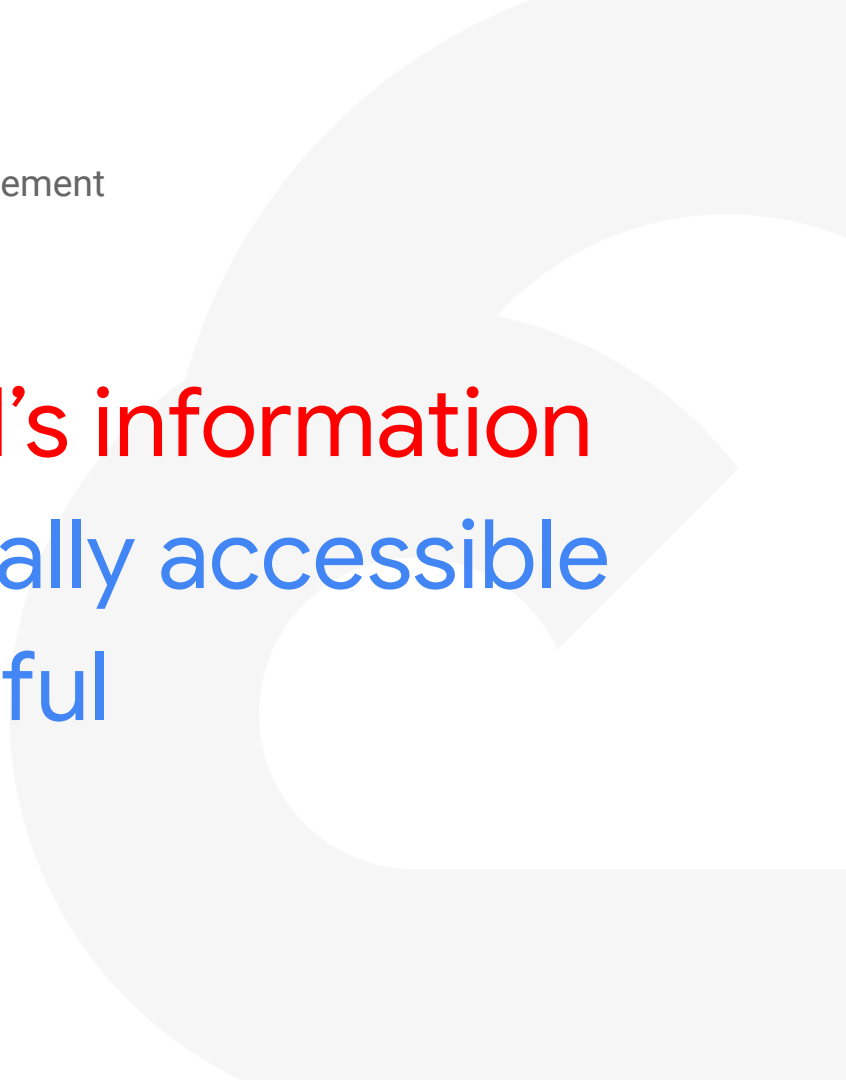
Agenda

Google Cloud

Google Mission Statement

Organize the world's information and make it universally accessible and useful

Google Mission Statement

Organize the world's information and make it universally accessible and useful

Data Volume Growth

Digital Information Measurement Unit

Byte	Value	Name	Value
1,000	1.E+03	kilobyte	(KB)
1,000,000	1.E+06	megabyte	(MB)
1,000,000,000	1.E+09	gigabyte	(GB)
1,000,000,000,000	1.E+12	terabyte	(TB)
1,000,000,000,000,000	1.E+15	petabyte	(PB)
1,000,000,000,000,000,000	1.E+18	exabyte	(EB)
1,000,000,000,000,000,000,000	1.E+21	zettabyte	(ZB)
1,000,000,000,000,000,000,000,000	1.E+24	yottabyte	(YB)

Google Cloud

Data Volume Growth

Survey **in** **2009**

Byte	Value	Name	Value
1,000	1.E+03	kilobyte	(KB)
1,000,000	1.E+06	megabyte	(MB)
1,000,000,000	1.E+09	gigabyte	(GB)
1,000,000,000,000	1.E+12	terabyte	(TB)
1,000,000,000,000,000	1.E+15	petabyte	(PB)
1,000,000,000,000,000,000	1.E+18	exabyte	(EB)
1,000,000,000,000,000,000,000	1.E+21	zettabyte	(ZB)
1,000,000,000,000,000,000,000,000	1.E+24	yottabyte	(YB)

- 2K - A typewritten page
- 5M – The complete works of Shakespeare
- 10M – One minute of high fidelity sound
- 2T – Information generated on YouTube in one day
- 10T – 530,000,000 miles of bookshelves at Library of congress
- 20P – All hard disk drives in 1995
- 700P – Data of 700,000 companies with Revenues less than $200M
- 1E – Combined Fortune 1000 company database (1P each)
- 1E – Next 9000 world company databases (average 100T each)
- 1Z – 1000E (Zettabyte–Grains of sand on beaches)
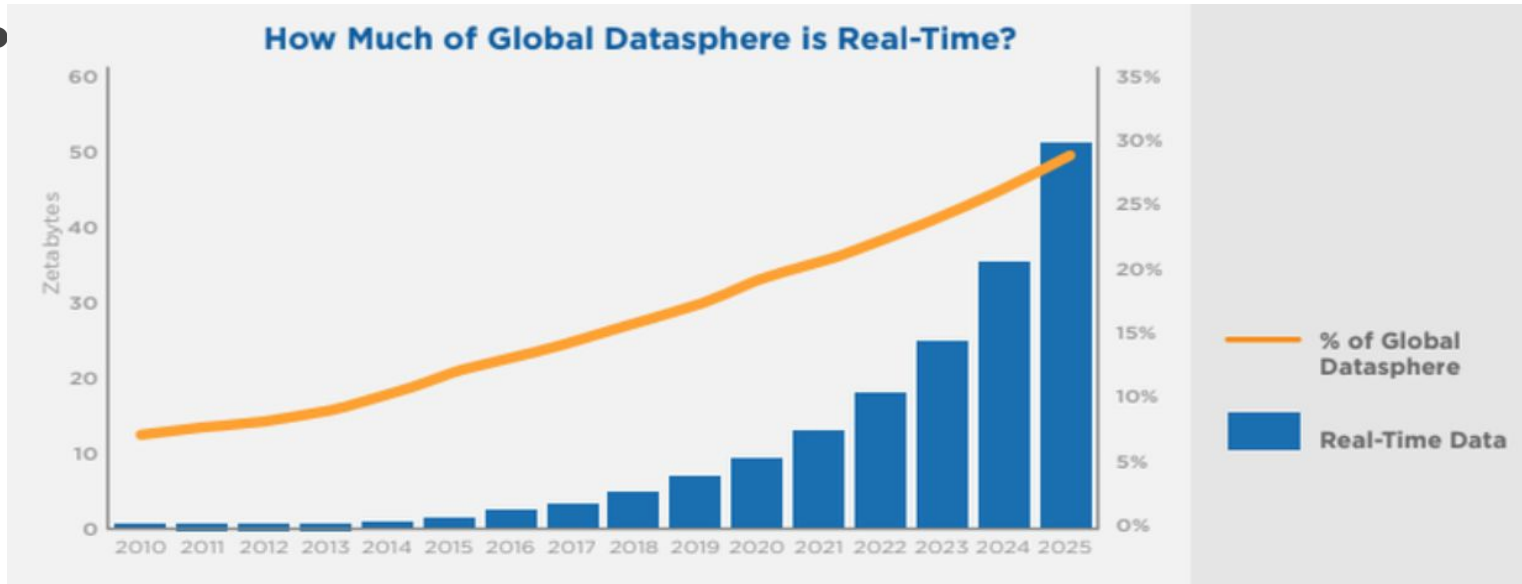- 100Y –Yottabytes – Addressable memory 128 -bit

Google Cloud

Global Datasphere

Survey by IDC

- IDC defines the "global datasphere" as "the quantification of the amount of data created, captured, and replicated across the world."

-

How Much of Global Datasphere is Real-Time?

Google Cloud

Organize the world's information and make it universally accessible and useful

Core tenets

1 If users can't spell, it's our problem.

2 If they don't know how to form the query, it's our problem.

3 If they don't know what words to use, it's our problem.

4 If they can't speak the language, it's our problem.

5 If there is not enough content on the web, it's our problem.

6 If the web is too slow, it's our problem.

Machine Learning is the new ground for gaining competitive edge & creating business value

Competitive advantage ranked as top goal of machine-learning projects for 46% of IT leaders & 50% of adopters can quantify ROI

| 2X more data-driven decisions | 5X faster decisions than others | 3X faster execution |

First Step in This Journey Begins with

Data

"Every Company will be a Data Company"
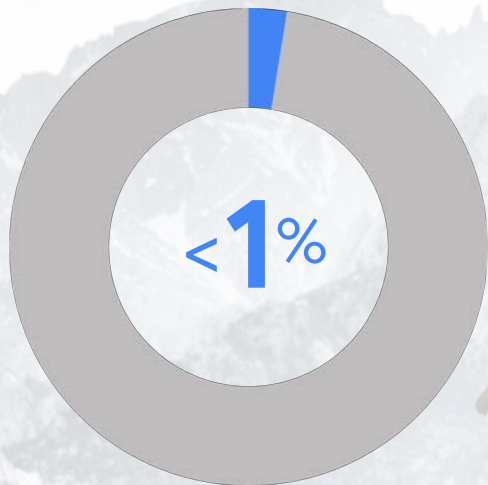
Source: Wired, Bloomberg, Fortune, McKinsey

Data Challenges

Data is Everything

Companies win or lose based on how do they use it
Governments make the right and wrong decisions based on the data they processed
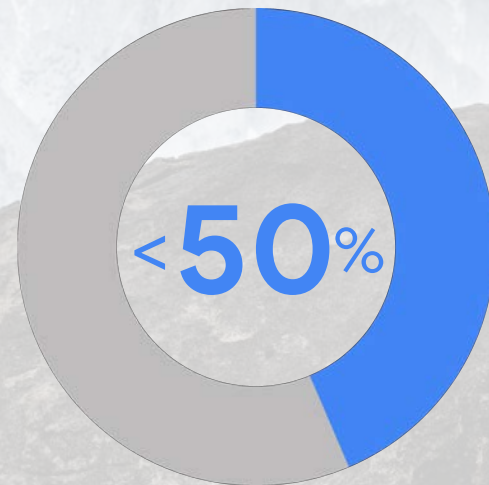You make your personal decision based on the data you collected

Data analytics is still too hard

<1%

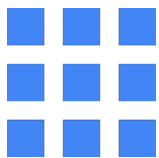Less than 1% of unstructured data is analyzed or used at all*

Less than 50% of structured data is used to make decisions*

<50%

Unstructured Data

Structured Data

Google Cloud

Data complexities

Unstructured data accounts for 90% of enterprise data

Legacy applications

1011101
0100101
11010101
0111100
10001101

Data silos everywhere

Changing view on value of data

Regulatory environment

Limited skills, hard to recruit

Challenges with Big Data Projects

1 Complexity of building and maintaining a Big Data system with consistent ease of use

2 Capture and store all data for all business functions

3 Continuously accommodating greater data volumes and new data sources

4 Finding value in existing data very easily

5 Reducing the time from data collection to action

6 Hurdles to innovate and iterate with Big Data

7 Collaboration within or across organizations

8 Keep your data secure

9 Keep system reliable/running

World Map   NEW   **U.S. Map**   **Critical Trends**

COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Total Confirmed
2,766,611

Confirmed Cases by Country/Region/Sovereignty

- 883,826 US
- 219,764 Spain
- 192,994 Italy
- 159,495 France
- 154,111 Germany
- 144,632 United Kingdom
- 104,912 Turkey
- 88,194 Iran
- 83,885 China
- 68,622 Russia
- 51,073 Brazil
- 44,293 Belgium
- 43,559 Canada
- 36,727 Netherlands

Admin0 Admin1 Admin2

Last Updated at (M/D/YYYY)
4/24/2020, 10:31:35 AM

NORTH AMERICA
North Pacific Ocean
North Atlantic Ocean
EUROPE
ASIA
North Pacific Ocean
AFRICA
SOUTH AMERICA
South Pacific Ocean
South Atlantic Ocean
Indian Ocean
AUSTRALIA

Esri, FAO, NOAA

Cumulative Confirmed Cases | Active Cases | Incidence Rate | Case-Fatality Ratio | Testing Rate | Hospitalization Rate

185 countries/regions

Lancet Inf Dis Article: Here. Mobile Version: Here.
Lead by JHU CSSE. Automation Support: Esri Living Atlas team and JHU APL. Contact US. FAQ.

Data sources: WHO, CDC, ECDC, NHC, DXY, 1point3acres, Worldometers.info, BNO, the COVID Tracking Project (testing and ...

Total Deaths
194,456

- 25,969 deaths Italy
- 22,524 deaths Spain
- 22,245 deaths France
- 19,506 deaths United Kingdom
- 16,388 deaths New York City **New York** US
- 6,679 deaths Belgium
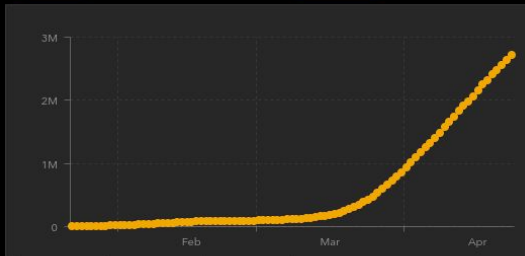- 5,632 deaths

Deaths | Recovered

Total Test Conducted in U.S.
4,692,797

- 695,920 tested New York US
- 482,097 tested California US
- 301,147 tested Florida US
- 225,078 tested Texas US
- 200,148 tested New Jersey US
- 195,076 tested Massachusetts US
- 178,708 tested Pennsylvania US

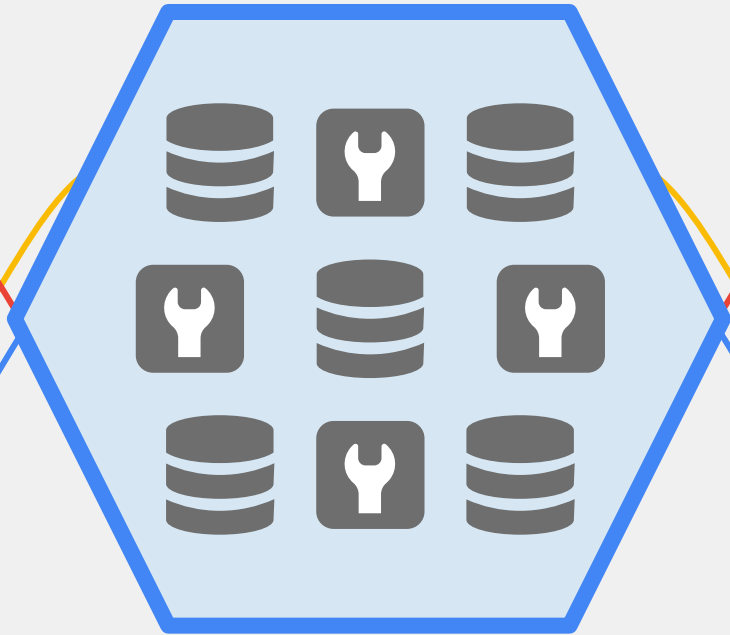US Tested | US Hospitalization

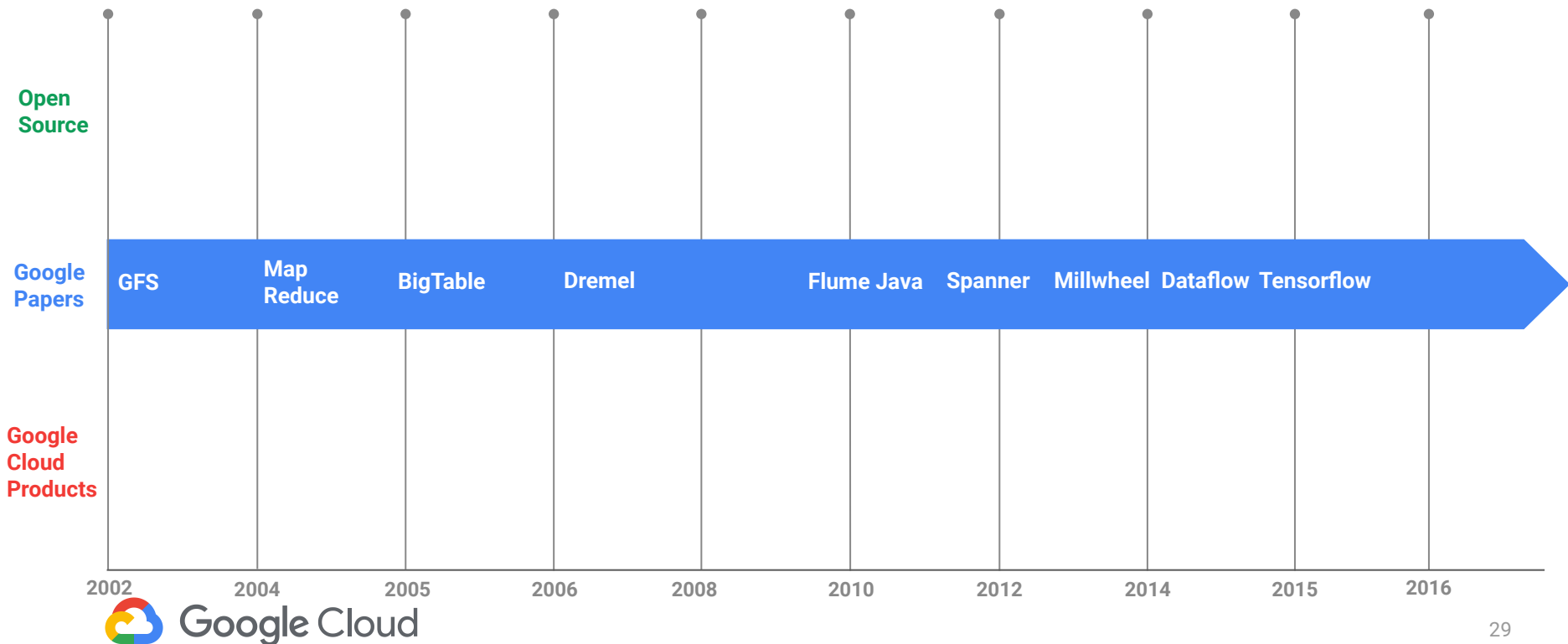Confirmed | Logarithmic | Daily Cases

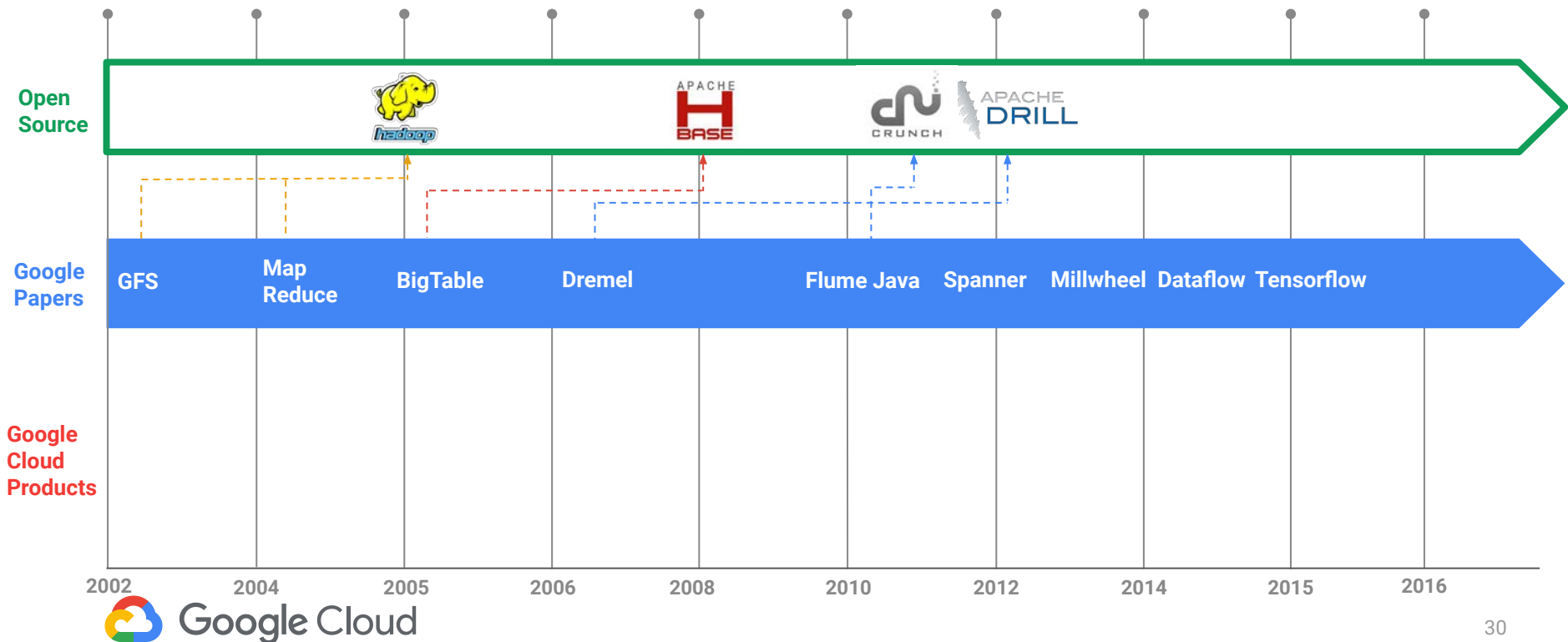If you want to unlock the power of your data, you need a **customer data platform**, not just new tools.

" **If Your Organization Isn't Good at Analytics,** "
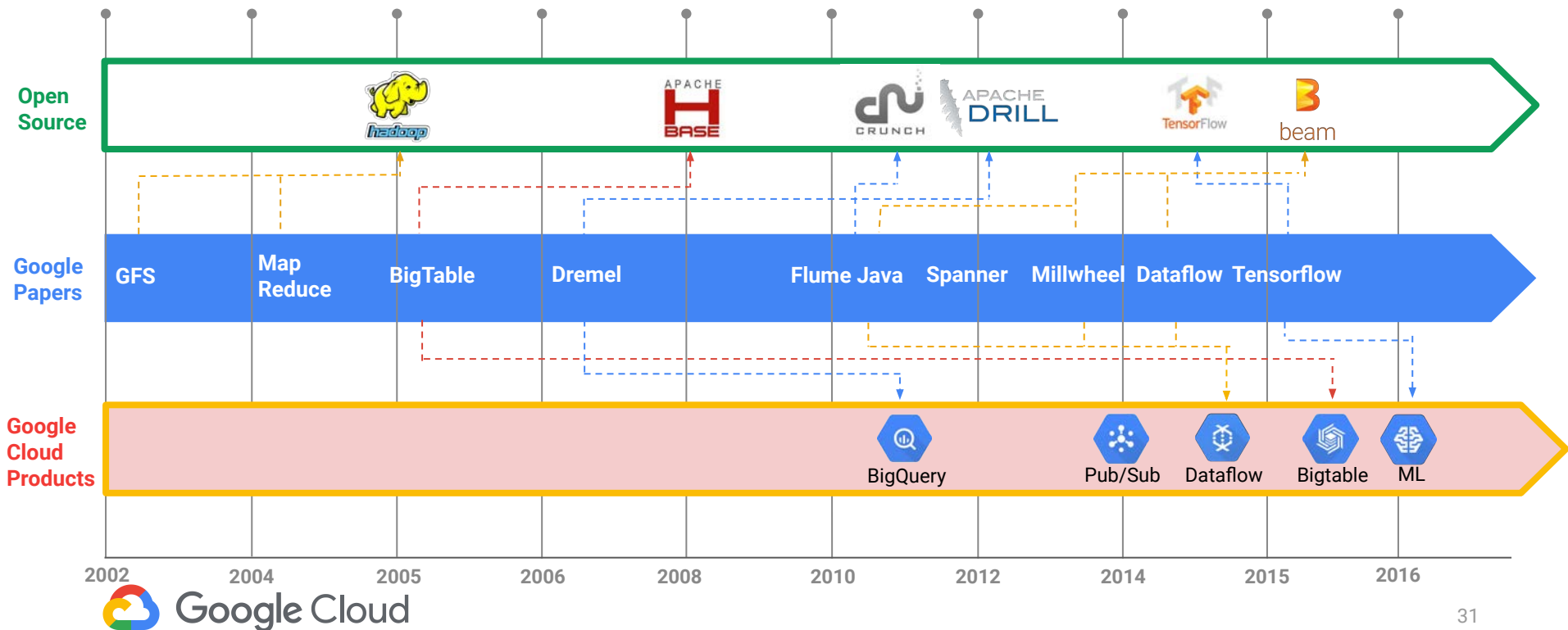**It's Not Ready for AI**

Google Cloud

Our Approach to Data Analytics

Google Cloud

15+ Years of Tackling Big Data Problems

Open Source

Google Papers

| GFS | Map Reduce | BigTable | Dremel | Flume Java | Spanner | Millwheel | Dataflow | Tensorflow |

Google Cloud Products

2002   2004   2005   2006   2008   2010   2012   2014   2015   2016

Google Cloud

15 Years of Tackling Big Data Problems

Open Source

Google Papers

GFS   Map Reduce   BigTable   Dremel   Flume Java   Spanner   Millwheel   Dataflow   Tensorflow

Google Cloud Products

2002   2004   2005   2006   2008   2010   2012   2014   2015   2016

Google Cloud

30

15 Years of Tackling Big Data Problems

Open Source

Google Papers

GFS   Map Reduce   BigTable   Dremel   Flume Java   Spanner   Millwheel   Dataflow   Tensorflow

Google Cloud Products

BigQuery   Pub/Sub   Dataflow   Bigtable   ML

2002   2004   2005   2006   2008   2010   2012   2014   2015   2016

Google Cloud

Serverless data analytics

From infrastructure to platform for insights

The traditional data analytics platform

Monitoring

Performance tuning

Utilization improvements

Deployment & configuration

Analysis and insights

Resource provisioning

Handling growing scale

Reliability
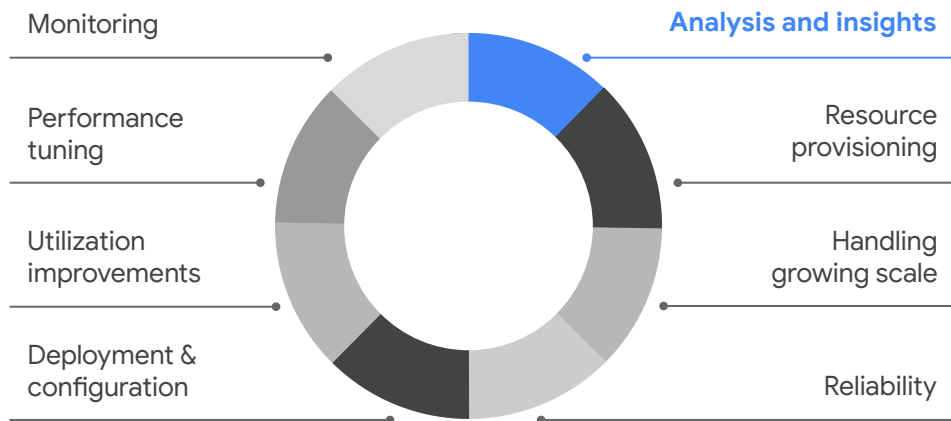
The serverless data analytics model

Analysis and insights

Google Cloud

Enterprise Challenges in Data to ML Journey

Data Silos and Legacy Systems

Limits decision-making and is time consuming

Missing Out on Real-Time Insights

Rear-view approach causes business anxiety

Lacks How-To Predict Business Outcomes

Depends on guts for predicting the unknown

Key Solutions Powered by Google Cloud

Cloud Data Warehouse

Modern Data Warehousing which builds foundation for AI

Streaming Data Analytics

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

Predictive Analytics / ML

Anticipate customer needs and automate delivery with Machine Intelligence

Complete foundation for data lifecycle

Data ingestion at any scale
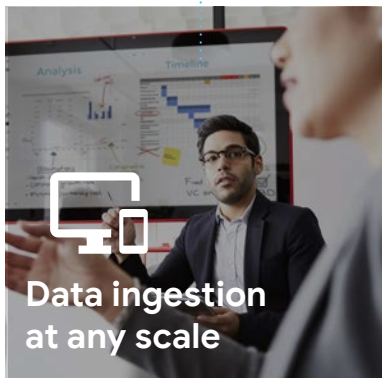
Cloud Pub/Sub   Data Transfer Service

Cloud IoT Core

Reliable streaming data pipeline
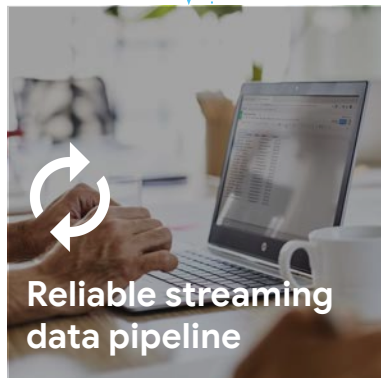
Cloud Dataflow   Cloud Dataproc (Hadoop, Spark)

Apache Beam   Cloud Dataprep (Trifacta)

Data warehousing and data lake
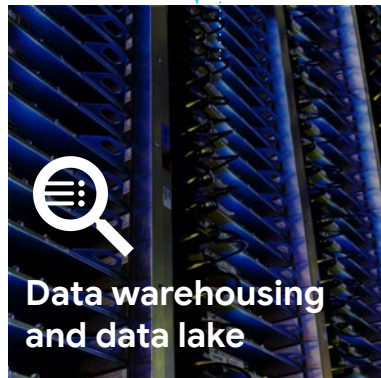
BigQuery   Cloud Storage

Advanced analytics
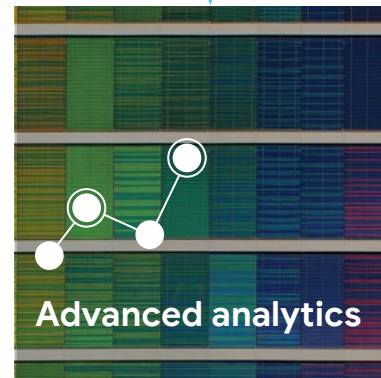
Cloud ML Engine   Google Data Studio

Tensorflow   Sheets

Cloud Composer (Apache Airflow) = Cloud Composer + Apache Airflow + Kubernetes Engine + Cloud Storage

Google Cloud

35

Modernize Your Data Warehouse

Get all your business data in one place for faster and comprehensive analysis

Google Cloud

Data warehousing for AI-driven business

90's

00's

Now

Next

Data warehouses

From 1st-gen EDWs, increased data collection and analysis has helped build more data-driven businesses.

BI foundations

Data warehousing formed the foundation of reporting and business intelligence.

Cloud data warehousing

BigQuery represents a fundamentally different approach to cloud data warehousing.

AI foundations

We're working to make BigQuery the foundation for organizations that will leverage machine intelligence in their businesses.

Google Cloud

Google Cloud Data Warehouse: Four Typical Flows
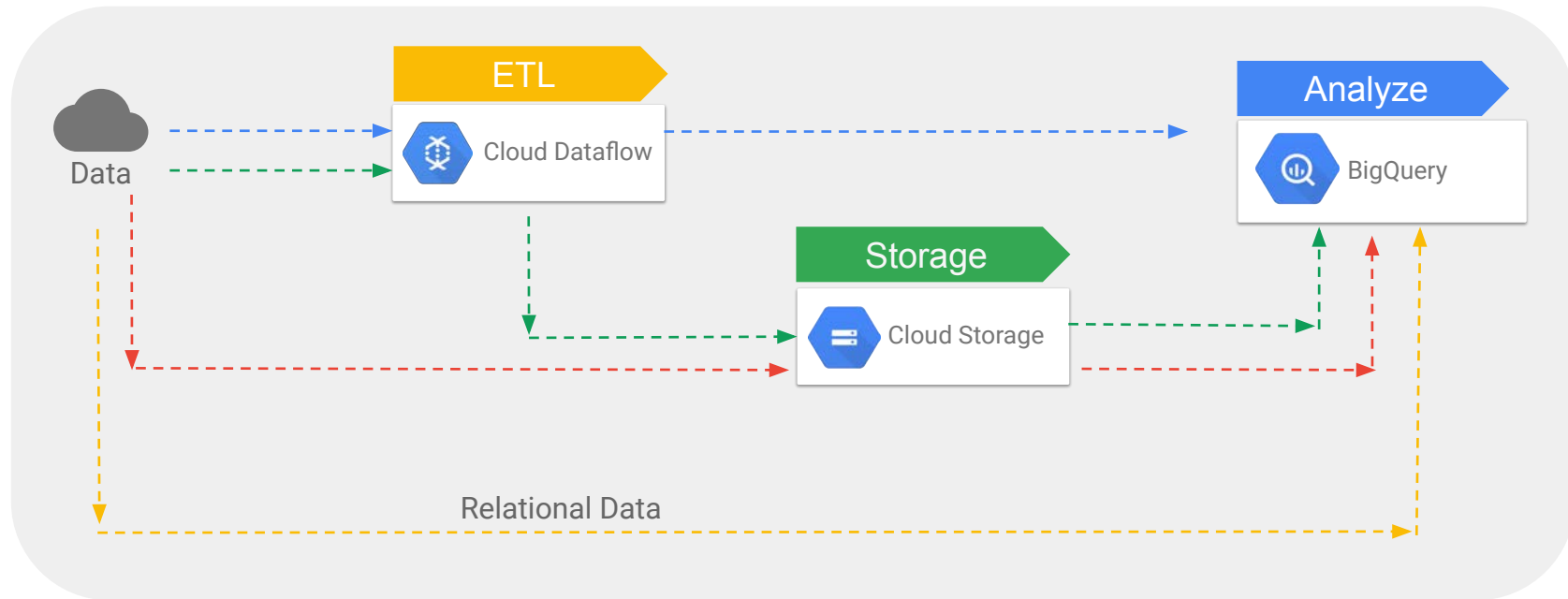
What is BigQuery?

Google Cloud Platform's enterprise data warehouse for analytics

Petabyte-scale storage and queries

Convenience of standard SQL

Encrypted, durable and highly available

Fully managed and serverless

Real-time analytics on streaming data

Google Cloud

BigQuery: architecture

Serverless. Decoupled storage and compute for maximum flexibility.
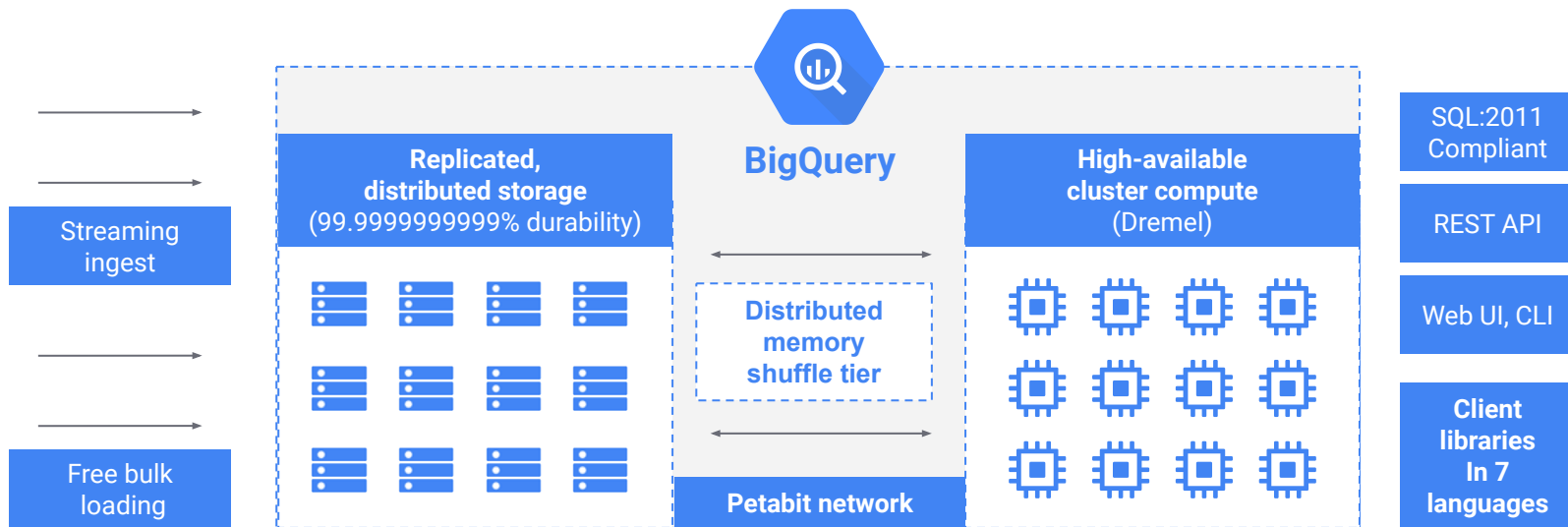
Introducing BigQuery ML

Making machine learning accessible

Google Cloud

BigQuery ML empowers data analysts and data scientists

Execute ML initiatives without moving data from BigQuery

Iterate on models in SQL in BigQuery to increase development speed

Automate model selection, and hypertuning

Google Cloud

Analyze GIS data in BigQuery with familiar SQL

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

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

Google Cloud

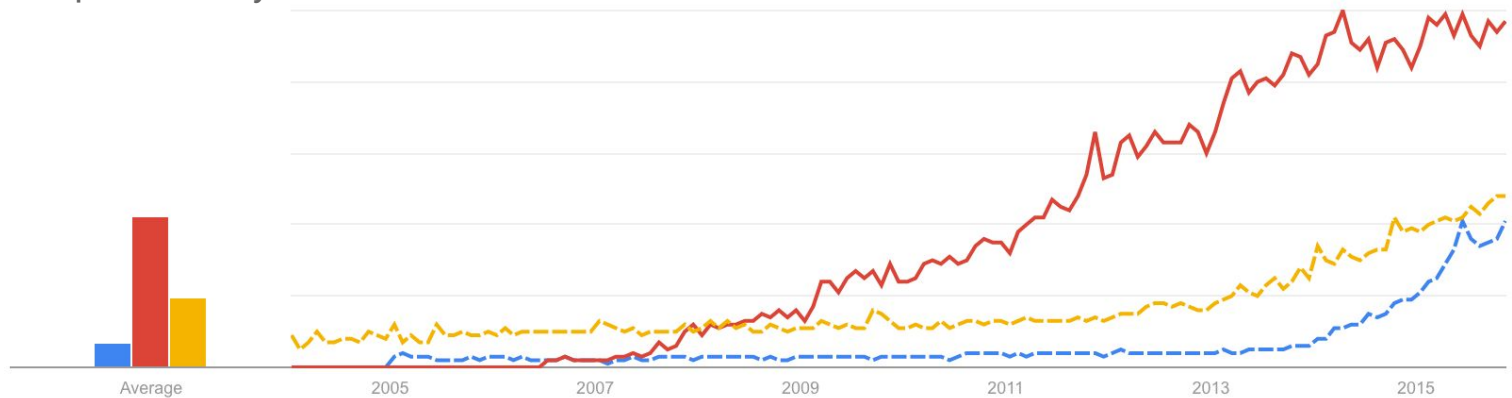# You can use BigQuery to build a modern marketing data warehouse
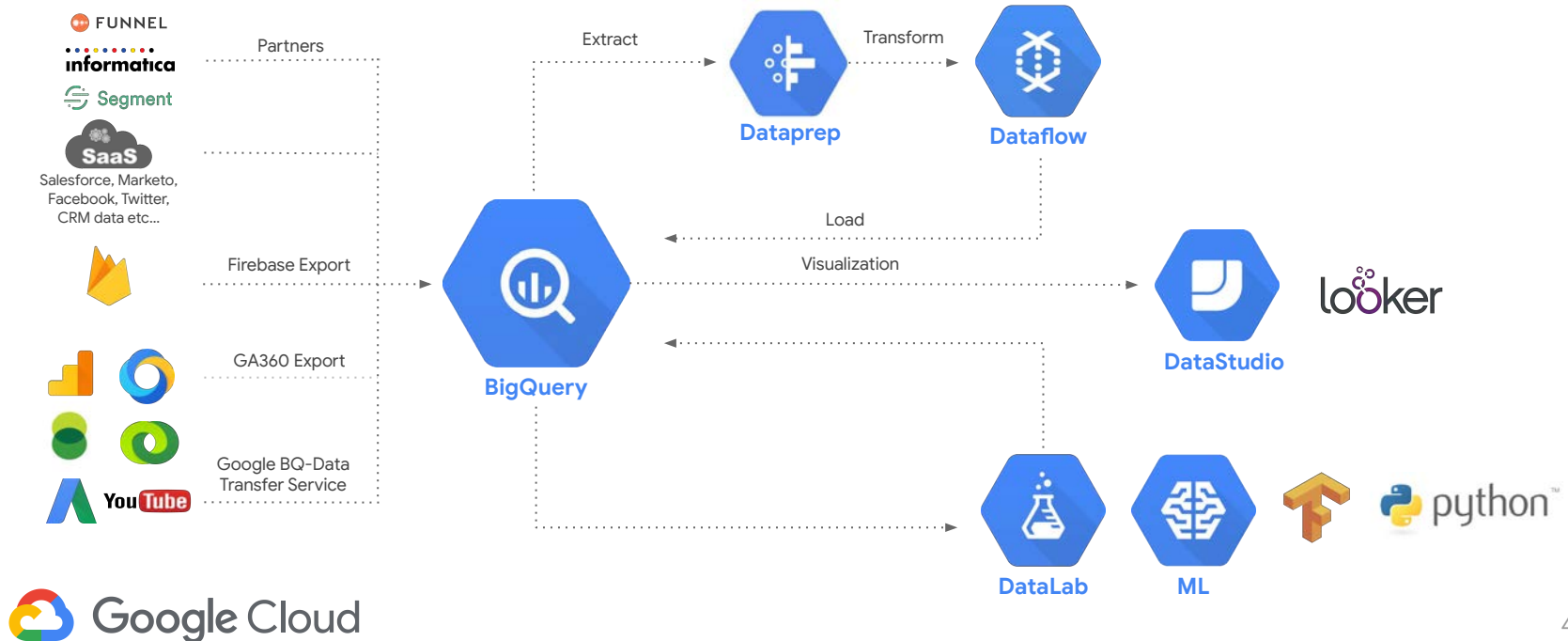
Big Data & Hadoop

Hadoop Ecosystem is a Popular Choice

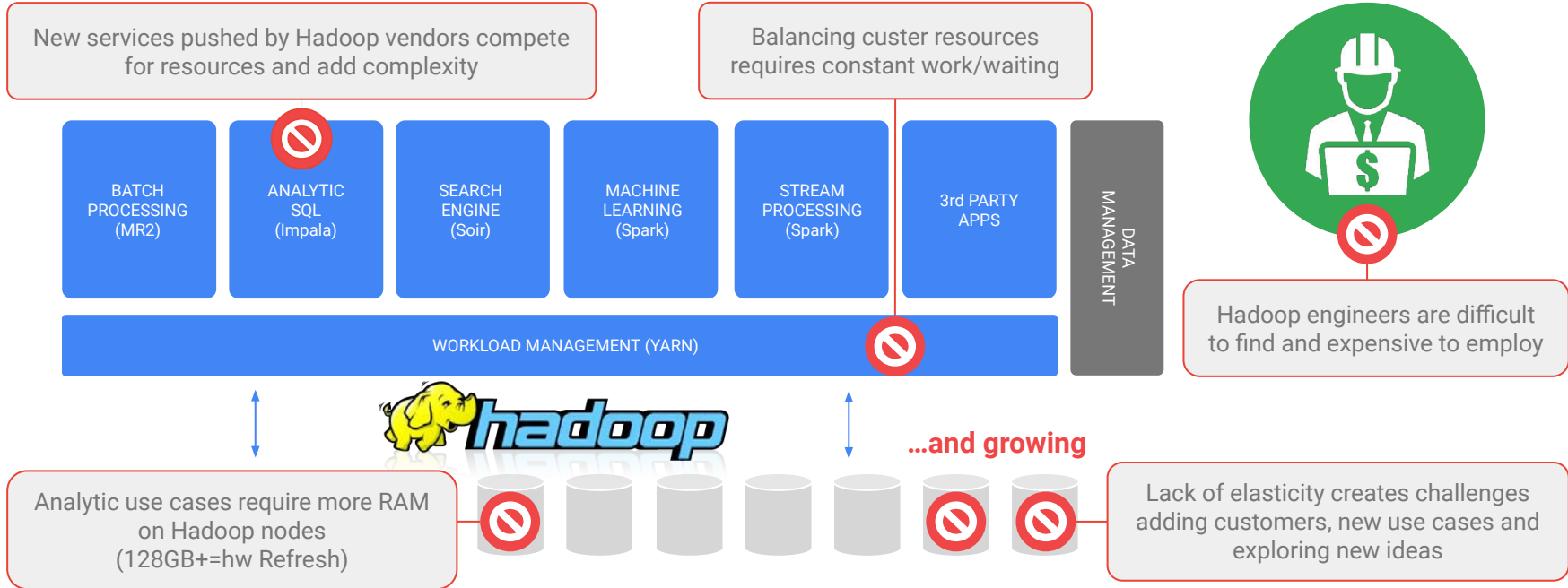Production implementations of hadoop, spark, and other components (like hive) are growing steadily over time

Hadoop will likely be a $50b market by 2020

Hadoop is used by 80% of the Fortune 500

Source: Lorem ipsum, 2017

Google Cloud

Managing Hadoop can be … Complex

New services pushed by Hadoop vendors compete for resources and add complexity

Balancing custer resources requires constant work/waiting

| BATCH PROCESSING (MR2) | ANALYTIC SQL (Impala) | SEARCH ENGINE (Soir) | MACHINE LEARNING (Spark) | STREAM PROCESSING (Spark) | 3rd PARTY APPS | DATA MANAGEMENT |

WORKLOAD MANAGEMENT (YARN)

Hadoop engineers are difficult to find and expensive to employ

...and growing

Analytic use cases require more RAM on Hadoop nodes (128GB+=hw Refresh)

Lack of elasticity creates challenges adding customers, new use cases and exploring new ideas

How can you simplify Hadoop management?

Hadoop to Cloud: options

managed by you

managed by Google

Complex

Simple

BigQuery (Dremel - Cloud DW)

Dataflow (Apache Beam - ETL, Streaming)

PubSub (Apache Kafka - Global-scale Messaging)

Dataproc (Hadoop, Spark - BigData)

Bigtable (NoSQL Database - HBase, Cassandra)

HDFS → Storage

Compute Engine
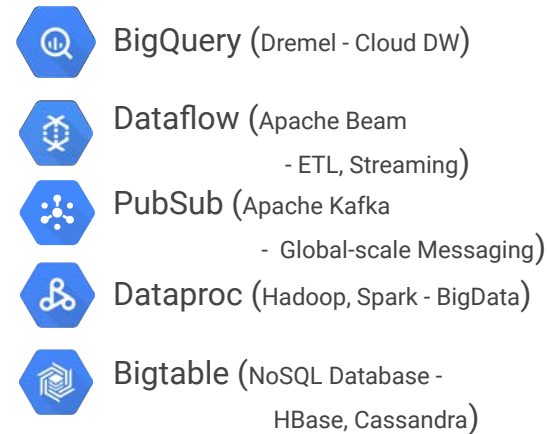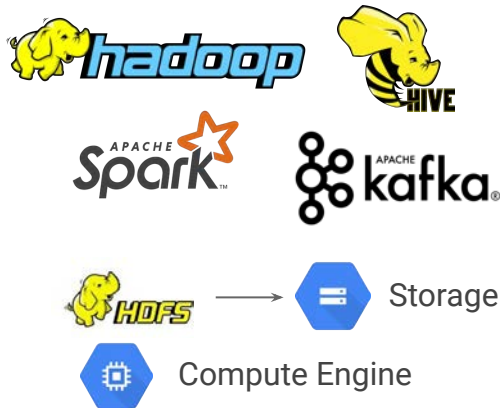
Google Cloud Platform

Google Cloud

Faster and easier Spark & Hadoop jobs with **Cloud Dataproc**

Cloud Dataproc

It is the simpler, more cost-efficient way to make your **Apache Spark & Hadoop (MapReduce)** deployments a success

It's flexible

Create and resize managed Hadoop and Spark clusters in less than 90 seconds

It's easy

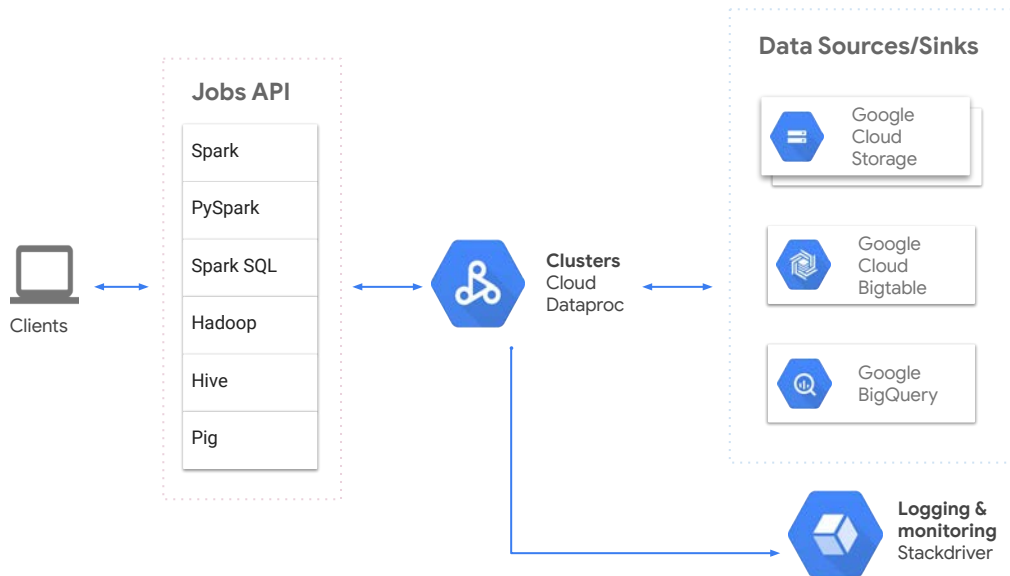Lift and shift existing projects or ETL pipelines, no redevelopment necessary

It's cost effective

Easily process large datasets at low cost, pay only for the resources you use (by the minute)

It's open

Leverage tools, libraries, and documentation from the Spark and Hadoop ecosystem

Google Cloud

Integrates easily with other Cloud Platform services

Data Sources/Sinks

Jobs API

Spark
PySpark
Spark SQL
Hadoop
Hive
Pig

Clients

Clusters
Cloud Dataproc

Google Cloud Storage

Google Cloud Bigtable

Google BigQuery

Logging & monitoring
Stackdriver

Share data across the platform
Connectors available to read/write from BigQuery, Cloud Bigtable, and Cloud Storage

Match processing with data
Bring the right processing engine to the workload (and at right cost) on the same storage

Monitor & alert with Stackdriver
Integration with Stackdriver, GCP's logging & monitoring framework, lets you identify/diagnose issues

Google Cloud

Analyze Streaming Data in Real Time

Gain real-time business insights and
make your business more responsive

Google Cloud

Stream data analytics on Google Cloud Platform

Ingest

Ingest and distribute data reliably

Cloud Pub/Sub

Transform

Fast, correct computations quickly and simply

Cloud Dataflow

Analyze

Machine learning & data warehouse

BigQuery

Cloud Machine Learning

Cloud Translation API

Cloud Vision API

Cloud Natural Language API

Google Cloud

Unified streaming and batch data processing with **Cloud Dataflow**

Google Cloud

Cloud Dataflow

The fully-managed data processing service that simplifies development and management of stream and batch pipelines

Accelerate development for *streaming & batch*

Fast, simplified data pipeline development via expressive Java and Python APIs in the Apache Beam SDK

Simplified management and operations

Remove operational overhead by letting Cloud Dataflow auto-manage performance, scaling, availability, security and compliance.

Build on a foundation for machine learning

Add TensorFlow-based Cloud Machine Learning models and APIs to your data processing pipelines for real-time predictions

Google Cloud

Visually prepare data for analysis or ML projects with **Cloud Dataprep**

Fully-managed and serverless

Team-based data wrangling
Share & copy flows to collaborate in real-time, reuse custom samples in a recipe, and audit individual user wrangling steps

Data analyst productivity
Use shortcuts for popular transforms (pivots, joins, unions), simplify date formatting, and parameterize scheduled flows with dynamic data

Comprehensive design refresh
Ensure smooth onboarding, see activity organized by recency and navigate easily between different stages of the workflow

Google Cloud
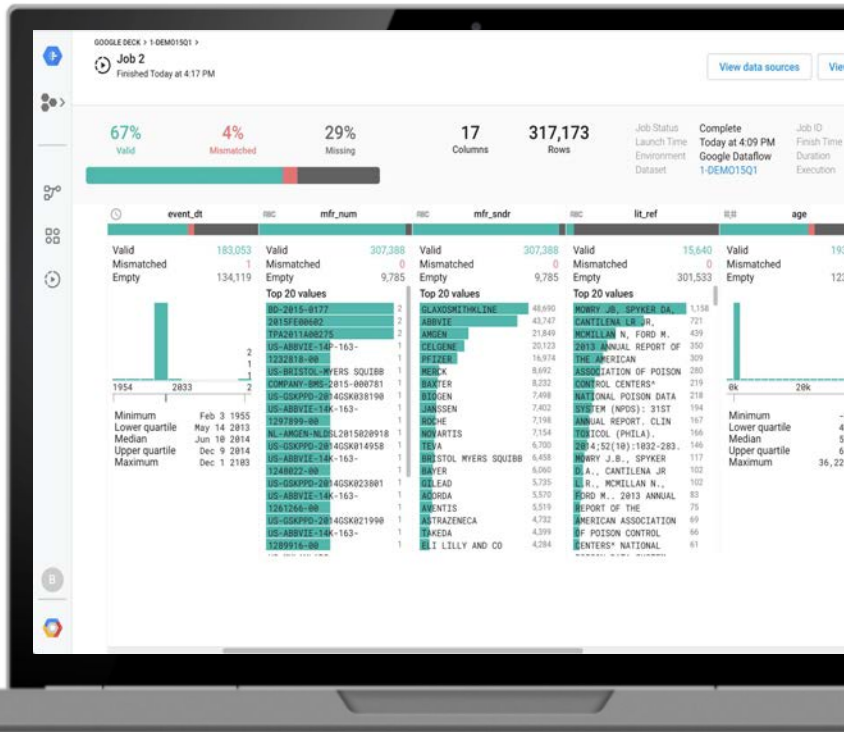
Collaborative Business Intelligence with Data Studio

One-click visualization

Visualize BigQuery data in a click with Data Studio Explorer

Data Studio data blending

Join across data sources with a simple right-click

Data Studio template gallery

Get started fast using Google-built or community-built templates.

Data Studio custom Visualizations DEVELOPER PREVIEW

Build custom visualizations using the popular D3.js framework.

Google Cloud

61
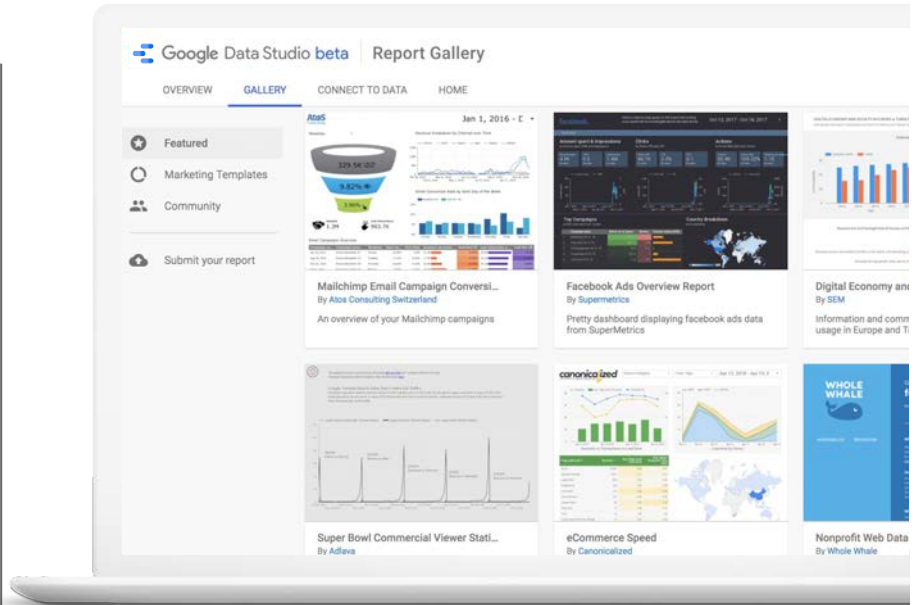
More from the tools you already use

Data ingest	Data integration	Data management	Prep / processing	BI/visualization	Services partners
Fivetran	Informatica	Waterline Data	TRIFACTA	tableau	Pythian
Stitch	talend	DATAGUISE	atscale	looker · Qlik Q	pwc
striim	SAP Data Services	protegrity	tamr	MicroStrategy	accenture
FUNNEL	pentaho	UNIFI	kyvos insights	DOMO · ZOOMDATA	Capgemini
Infoworks	software AG	Reltio	CoolaData	SAP Analytics Cloud	slalom
Dbvisit	alooma		ClearStory	IBM Cognos Analytics	tectonic
TREASURE DATA	xplenty		Lytics	Yellowfin · CHARTIO	Cloud Technology Partners
SKYVIA	Segment		Anodot	iCharts · Dundas BI	SADA Systems
Switchboard Software	MATILLION		Trend	OWOX · MODE	TRACE3
BLEND	StreamSets		Ideata	bime by zendesk · SISENSE	moser consulting
RIVERY	snapLogic		indicative	XCALAR · incorta	
EXAPIE	ironSource			re·dash	
fluentd				liquidlandscape	
				PYRAMID ANALYTICS	

Google Cloud

Serverless platform & auto-optimized usage across the entire data lifecycle

Capture	Process	Store	Analyze	Use

Capture
- Cloud Pub/Sub
- Storage Transfer Service
- Adwords
- DoubleClick
- YouTube
- Firebase
- Stackdriver
- GA 360

Process
- Cloud Dataprep
- Cloud Dataflow
- Cloud Dataproc

Store
- BigQuery storage (tables)
- Cloud Bigtable (NoSQL)
- Cloud Storage (files)

Analyze
- BigQuery analytics — SQL
- Cloud Dataflow
- Cloud Dataproc
- ML Engine

Use
- Data Studio
- Looker
- MicroStrategy
- tableau software
- Qlik
- redash
- iCharts
- Cloud Datalab

Business Analysts

Data Scientists

informatica Segment xplenty talend alooma striim

ST **Cloud Composer (orchestration)**

Google: a leader in insight platforms-as-a-service

"Our evaluation identified one vendor as a Leader based on the strength of its PaaS strategy, advanced tools for batch and real-time solutions, and machine learning and AI offerings."
— *The Forrester Report*

- Google has the highest scores in the Current Offering and Strategy categories.

- Noted as the only vendor in the evaluation to offer insight execution features like full machine learning automation with hyperparameter tuning, container management, and API management.

- Receives recognition for advanced platform features like autoscaling for most of its services, efforts at integrating leading Hadoop cloud services and its data flow service works on both batch and streaming data.

Google Cloud

Data Visualization

Google Cloud

06

Google Cloud's cloud-native Enterprise BI Platform enabling **secure access** to **near real-time data when and where** you need it.

Predictive Analytics & Machine Learning

Google Cloud

What is Machine Learning?

Machine Learning systems take inputs (data) to make useful predictions and decisions about previously unseen pieces of data.

Machine learning is a specific field of AI where a system learns to find patterns in examples in order to make predictions.

Computers learning how to do a task without being explicitly programmed to do so.

Machine Learning systems might:

- Label or classify data

- Predict numerical values

- Cluster similar pieces of data together

- Infer association patterns in data

- Create complex outputs

Examples of Machine Learning

Automating drone-based wildlife surveys saves time and money

Machine Learning could be used for early dementia diagnosis
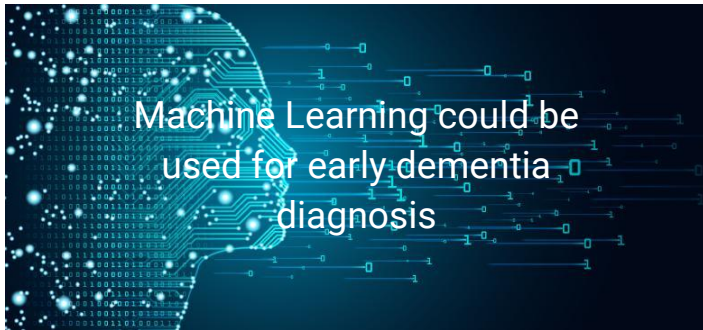
Read a couple of news articles involving applications of ML.

1. Would a traditional programming solution be more efficient?

2. Could a human perform the same task in less time?

3. What are the benefits of a Machine Learning model in these instances?

Machine Learning & Bias

Discussion points:

- Initial thoughts?
- How can we be mindful of this moving forward?
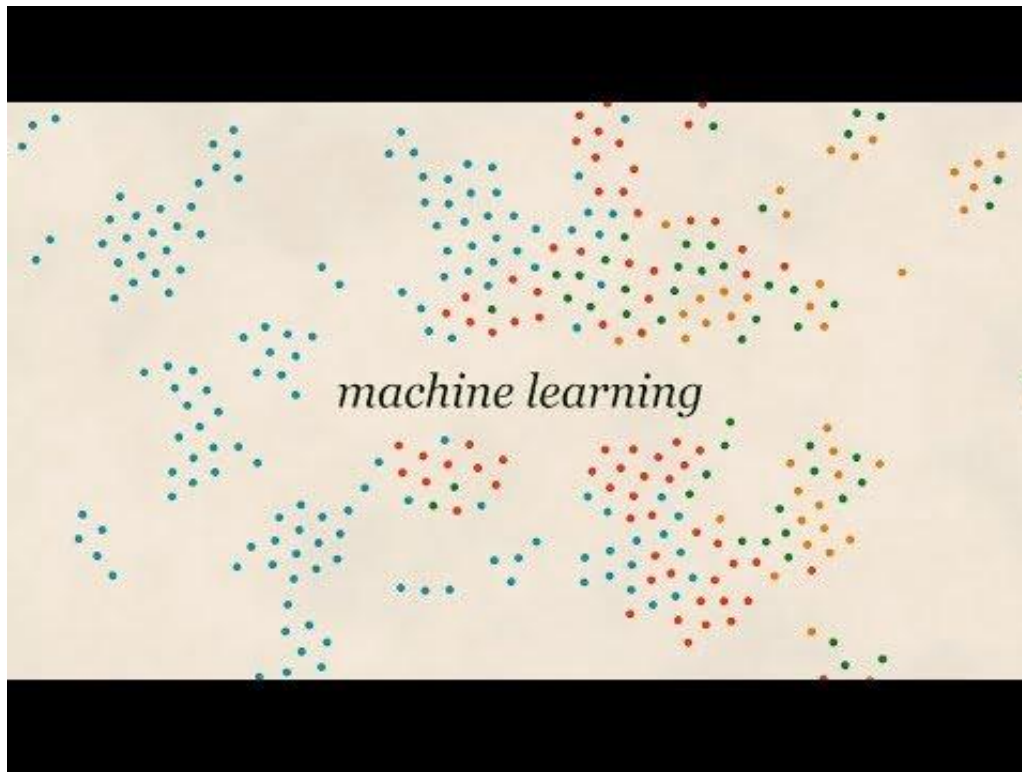
Machine Learning Allows You to Solve a Problem Without Codifying the Solution

Google Cloud AI

✓ Recognizes patterns in data

✓ Predictive analytics at scale

✓ Builds ML models seamlessly

✓ Fully managed service

✓ Deep Learning capabilities

Google Cloud End-to End AI Platform

Accelerate Business Outcomes with Enterprise-Ready Machine Learning Pipeline
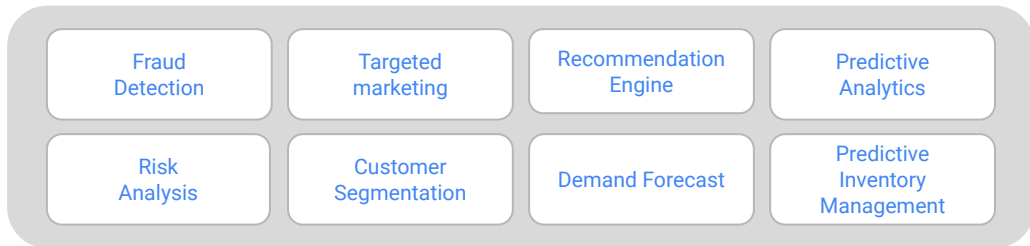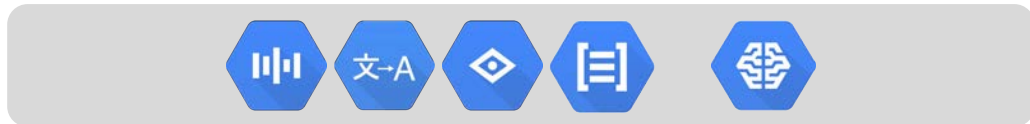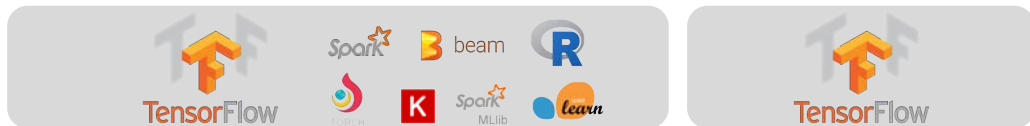
Google Cloud TPU

End to End: Google Cloud AI Spectrum

TensorFlow

Cloud MLE

ML Perception services
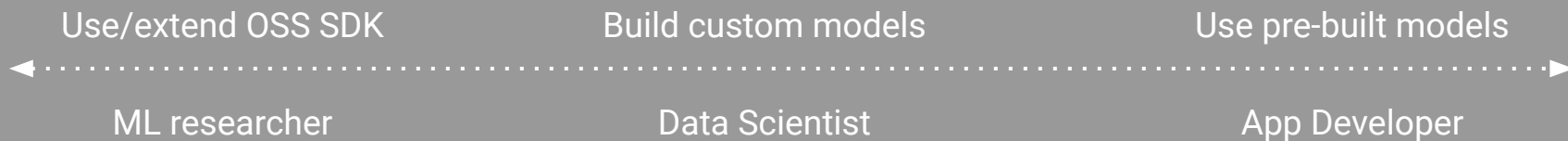
Use/extend OSS SDK   Build custom models   Use pre-built models

ML researcher   Data Scientist   App Developer

Google Cloud

Flow to build a custom ML model
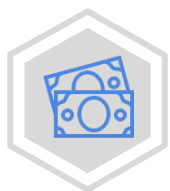
Identify business problem

Develop hypothesis
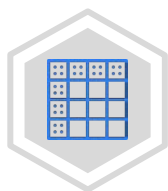
Acquire + explore data

Build a model

Train the model

Apply and scale

1   2   3   4   5   6

ML Perception Services

Cloud Natural Language API

Cloud Speech API

Cloud Translation API

Cloud Video Intelligence

Cloud Vision API

Google is a **world leader** in applying Machine Learning to real-world situations, inside and outside of Google.

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 usahe

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

Google Cloud

Cloud Machine Learning Use Cases

Manufacturing

- Predictive maintenance or condition monitoring
- Warranty reserve estimation
- Propensity to buy
- Demand forecasting
- Process optimization
- Telematics

Retail

- Predictive inventory planning
- Recommendation engines
- Upsell and cross-channel marketing
- Market segmentation and targeting
- Customer ROI and lifetime value

Healthcare and Life Sciences

- Alerts and diagnostics from real-time patient data
- Disease identification and risk satisfaction
- Patient triage optimization
- Proactive health management
- Healthcare provider sentiment analysis

Travel and Hospitality

- Aircraft scheduling
- Dynamic pricing
- Social media – consumer feedback and interaction analysis
- Customer complaint resolution
- Traffic patterns and congestion management

Financial Services

- Risk analytics and regulation
- Customer Segmentation
- Cross-selling and up-selling
- Sales and marketing campaign management
- Credit worthiness evaluation

Energy, Feedstock and Utilities

- Power usage analytics
- Seismic data processing
- Carbon emissions and trading
- Customer-specific pricing
- Smart grid management
- Energy demand and supply optimization

Google Cloud

Machine Learning Advanced Solutions Lab

Solving the biggest machine learning challenges, alongside our customers

Better customer value

- Intensive ML training led by our best experts
- Long-term engagement with Google ML engineers
- World-class, customer facilities on Google campuses

- Faster time to value
- Acquisition of best practices
- Competitive advantage through bleeding edge solution from Google

Next Steps for Gaining Competitive Advantage With Machine Learning

Get your arms around Big Data

Invest time in understanding Machine Learning

Work with us. Best practices, partners to help you

Google Cloud

How to get start with GCP

Google Cloud

https://cloud.google.com/

Take the next step

Get $300 in free credits to learn and build on Google Cloud for up to 12 months.

Get started for free

Google Cloud

https://cloud.google.com/

https://cloud.google.com/

https://cloud.google.com/

Google

Create your Google Account

to continue to Google Cloud Platform

First name
Daniel

Last name
Liu

Your email address
danieltliu@google.com

You'll need to confirm that this email belongs to you.

Create a new Gmail address instead

Password

Confirm

Use 8 or more characters with a mix of letters, numbers & symbols

One account. All of Google working for you.

Sign in instead

Next

English (United States)

Help   Privacy   Terms

Google Cloud

GCP Console (Demo)

DASHBOARD   ACTIVITY   RECOMMENDATIONS   ✏ CUSTOMIZE

ℹ   How Google Cloud is helping during COVID-19   Learn more   DISMISS

Project info ⋮

Project name
DanielDataProject

Project ID
danieldataproject

Project number
317928653106

ADD PEOPLE TO THIS PROJECT

→   Go to project settings

Resources ⋮

BigQuery
1 dataset

Trace ⋮

No trace data from the past 7 days

→   Get started with Stackdriver Trace

Getting Started ⋮

API   Explore and enable APIs

Deploy a prebuilt solution

APIs ⋮

Requests (requests/sec)

0.106
0.104
0.102
0.100
0.098
0.096
0.094

1:15   1:30   1:45   2 PM

→   Go to APIs overview

Google Cloud Platform status ⋮

All services normal

→   Go to Cloud status dashboard

Error Reporting ⋮

No sign of any errors. Have you set up Error Reporting?

→   Learn how to set up Error Reporting

News ⋮

How the Google Maps Platform community is responding to COVID-19: Visualizing helpful info
58 minutes ago

Admin Essentials: Managing Chrome Browser in a controlled environment
2 hours ago

How to add "buy online, pick-up in store" capabilities to your website with Google Maps Platform
23 hours ago

→   Read all news

Documentation ⋮

Learn about Compute Engine

Public Dataset

Filter by

147 results

TYPE

Datasets ✕

CATEGORY

Advertising (10)
Analytics (14)
Big data (9)
Climate (26)
Databases (1)
Datasets for COVID-19... (16)
Developer tools (2)
Economics (33)
Education (2)
Encyclopedic (33)
Financial services (45)
Genomics (3)
Healthcare (33)
Machine learning (3)
Maps (4)
Public safety (16)
Science & research (52)
Social (3)
Transportation (4)
Other (11)

About COVID-19 Public Datasets
BigQuery Public Datasets Program

Getting started with COVID-19 Public Datasets

American Community Survey (ACS)
United States Census Bureau

Detailed US demographic data at various geographic resolutions

Argentina Real Estate Listings
Properati

Monthly property listing data for Argentina since 2016

Austin Crime Data
City of Austin

City of Austin crime data for 2014 and 2015

Births Data Summary
Centers for Disease Control

Natality Data from CDC Births

Bitcoin Cash Cryptocurrency Dataset
Bitcoin Cash

The Bitcoin Cash blockchain loaded to BigQuery and updated daily

Bitcoin Cryptocurrency
Bitcoin

Bitcoin blockchain transactions and blocks, loaded in BigQuery

Brazil Real Estate Listings
Properati

Monthly property listing data for Brazil since 2016

Broad References
Broad Institute of MIT and Harvard

Human genomics reference files used for sequencing analytics

Catalonia cell coverage
Government of Catalonia

Cell coverage data collected by the GenCat crowdsourcing app

Census Bureau US Boundaries
United States Census Bureau

Datasets that define US political and statistical boundaries

Center for Medicare and Medicaid Services - Dual...
U.S. Department of Health & Hum...

Dual enrollment of Medicare and Medicaid by county

CFPB Consumer Complaint Database
Consumer Financial Protection Bu...

Complaints about consumer financial products and services

ChEMBL Data
Google Patents Public Datasets

Chicago Crime Data
City of Chicago

Chicago Police Department crime data from 2001 to present

Chicago Taxi Trips
City of Chicago

Chile Real Estate Listings
Properati

Clemson DICE Lab - TrafficVision Tracklets 2019

ClimaCell - CBAM India Weather Forecasts

Cloud-to-Ground Lightning Strikes

That's a wrap.

HACC Kick Off
Saturday, October 24 at the Hawaii State Capitol

Google Cloud