



# Advanced Big Data:

## Course Intro

**Dan Zaratsian**

Tech Lead, AI/ML Architect  
Gaming Solutions @ Google

[d.zaratsian@gmail.com](mailto:d.zaratsian@gmail.com)

<https://github.com/zaratsian>

# TOPICS

- **Session 1: Course Intro, Trends, and Approach to AI/ML**
- **Session 2: SQL and NoSQL**
- **Session 3: Cloud Machine Learning Services**
- **Session 4: Distributed ML with Spark and Tensorflow**
- **Session 5: Cloud Generative AI Services and Architectures**
- **Session 6: Serverless ML, Architectures, and Deploying ML**

# What I hope you get out of this course:

- Learn the core components of a **scalable machine learning solution**
- Develop an understanding of **distributed databases** (SQL and NoSQL)
- Train and deploy a **distributed ML model**
- Gain additional experience with **Generative AI** models and APIs
- Understand **Cloud Architectures for ML** (primarily focused on Google Cloud)
- Helpful as part of your **interview** process
- Ability to **recommend & use** this technology at your new company

# Class Structure and Expectations:

- Takeaways & the level of interest will vary by student
- Class content & structure will be flexible
- Teaching approach will be Applied (vs theoretical)
- Hands-on & Troubleshooting

 [iaa-2024](#) (Public)

[Pin](#) [Unwatch 1](#) [Fork 0](#) [Star 0](#)

[main](#) [Branch 1](#) [Tags 0](#)  Go to file [Add file](#) [Code](#)

z aratsian Added README	cf67b93 · now	2 Commits
LICENSE	Initial commit	last month
README.md	Added README	now

[Readme](#) [MIT license](#)

## Institute for Advanced Analytics

Cloud AI & Machine Learning - Dan Zaratsian, March 2024

### IAA Module - Session 1 - Course Intro

Slides

- Intro and Module Agenda
- Class Poll
- Trends in AI/ML
- ML Architectures
- [Google Colab Notebook Environment](#)
- [Google BigQuery Sandbox](#)

### IAA Module - Session 2 - SQL and NoSQL Services

Slides

**About**

Institute for Advanced Analytics, 2024

- [Readme](#)
- [MIT license](#)
- [Activity](#)
- [0 stars](#)
- [1 watching](#)
- [0 forks](#)

**Releases**

No releases published [Create a new release](#)

**Packages**

No packages published [Publish your first package](#)

Google coinbase sas



## Dan Zaratsian, Tech Lead, AI/ML Architect, Gaming Solutions @ Google

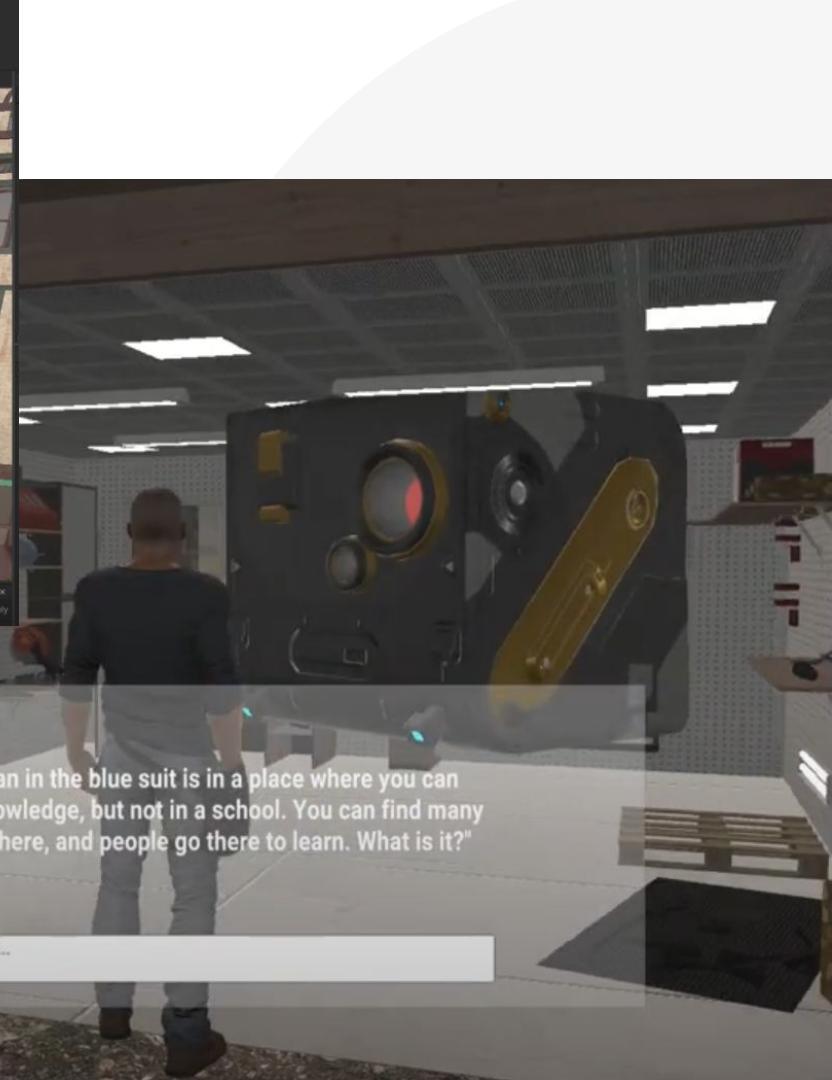
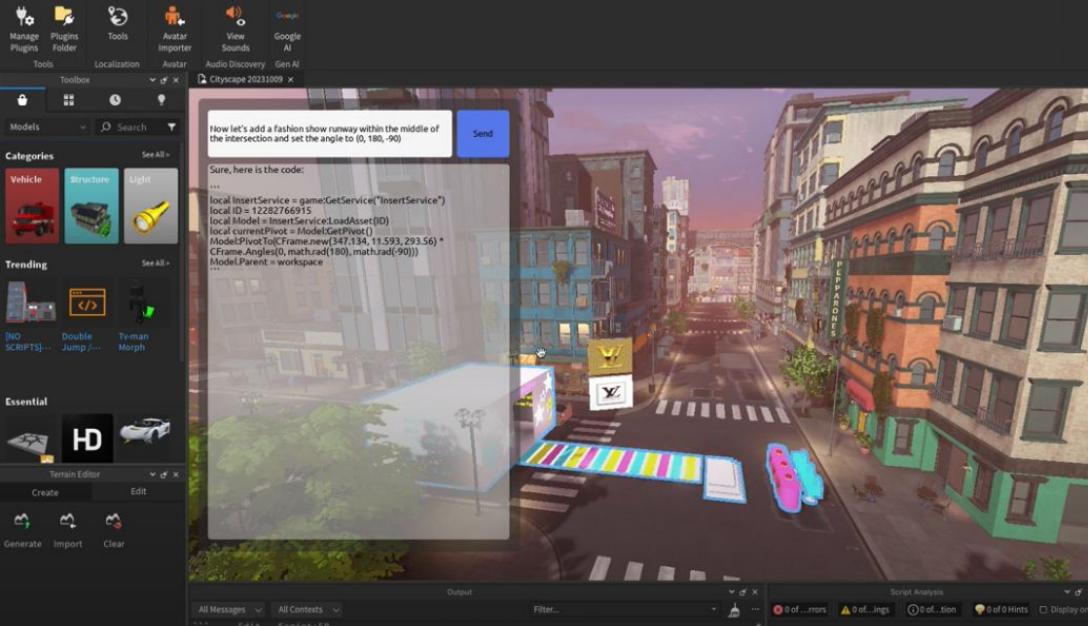
Architect, develop, and deploy  
gaming solutions on Google Cloud.

University of Akron  
B.S. Electrical Engineering, 2007



Institute for Advanced Analytics  
M.S. Advanced Analytics, 2011





[Menti](#)

Part 1

# Tech area that you'd like to learn/improve?

Waiting for responses ...

[Menti](#)

Part 2

# Generative AI: What are you most excited about?

Waiting for responses ...

# Current Trends in AI & Machine Learning

1. Dynamic AI/ML Skill & Role Requirements
2. Generative AI
3. AutoML and Low-code ML
4. Specialization

# Dynamic AI/ML Skill Requirements

## Proprietary Software



## Open Source Frameworks



## Cloud Services



Google Cloud



## Languages

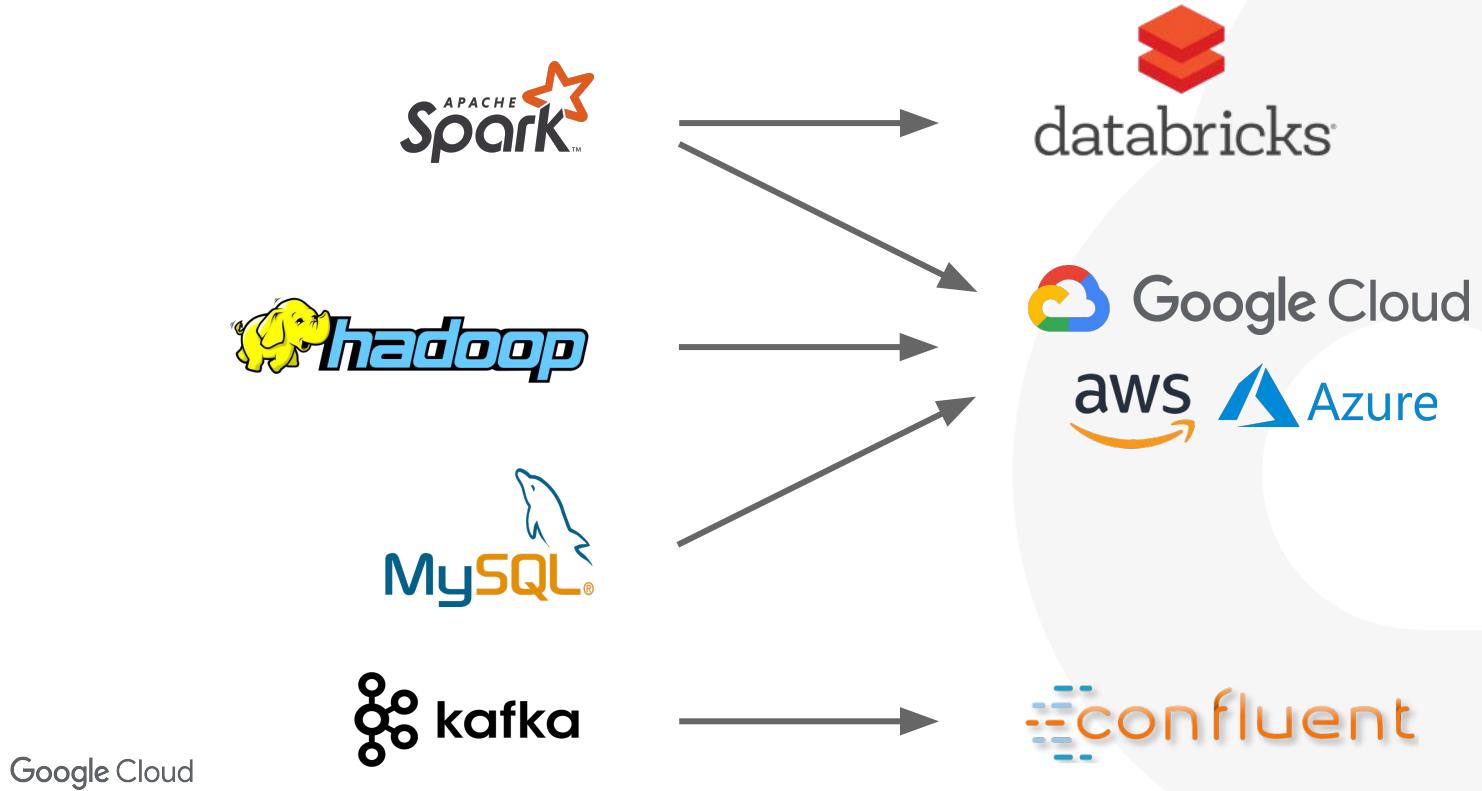


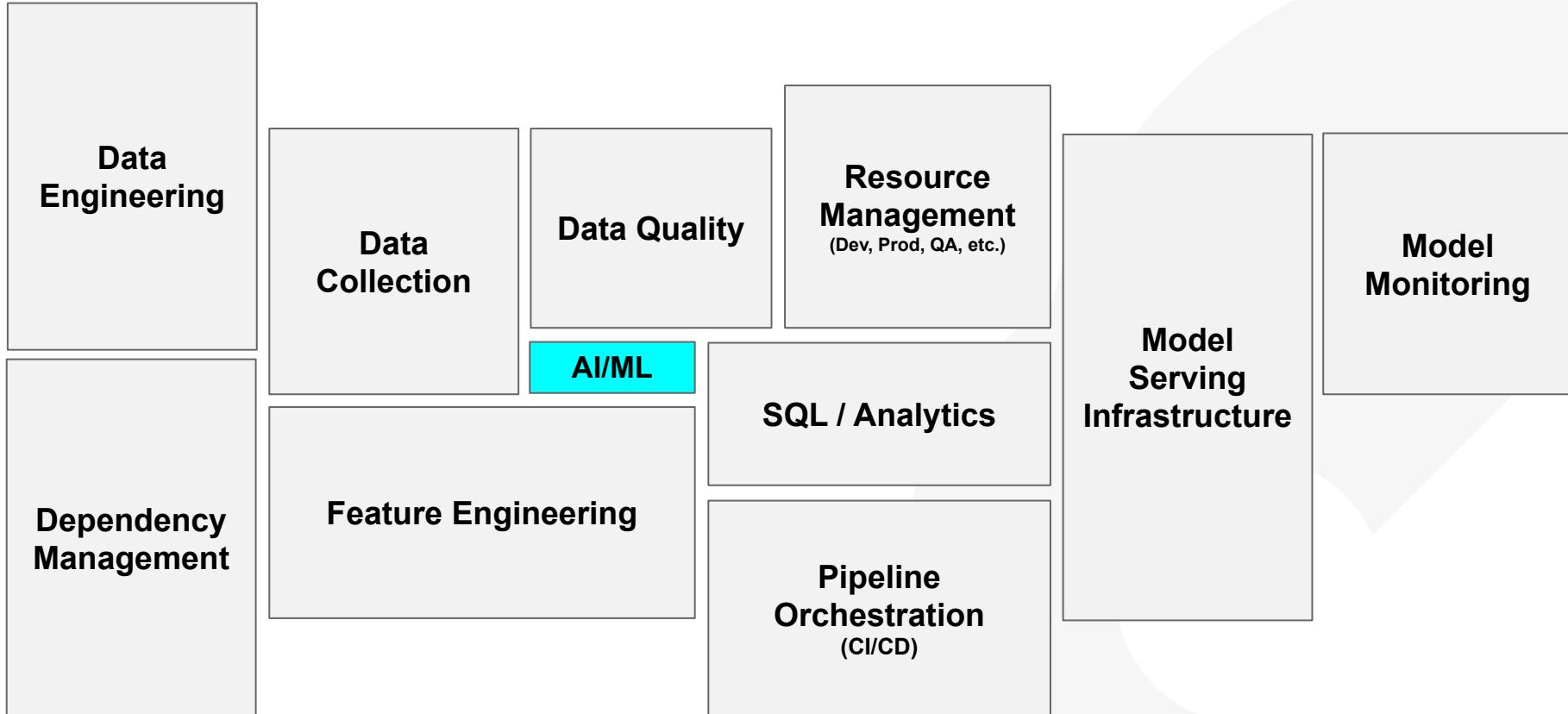
...many others...

Google Cloud

# Open Source to Managed Service

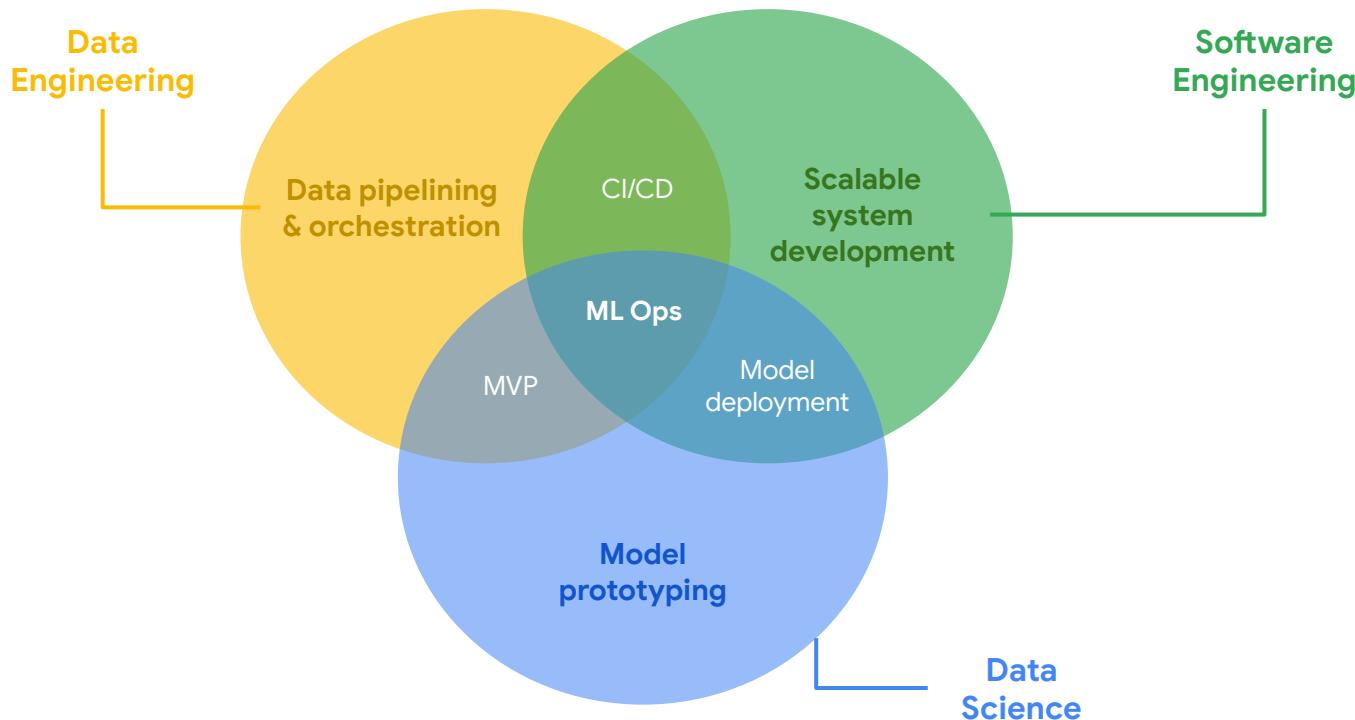
Evolution of deeply technical services turned into managed services and wrapped in software to make them easier to use.





**Secure • Scalable • Repeatable • Explainable • Flexible / Parameterized**

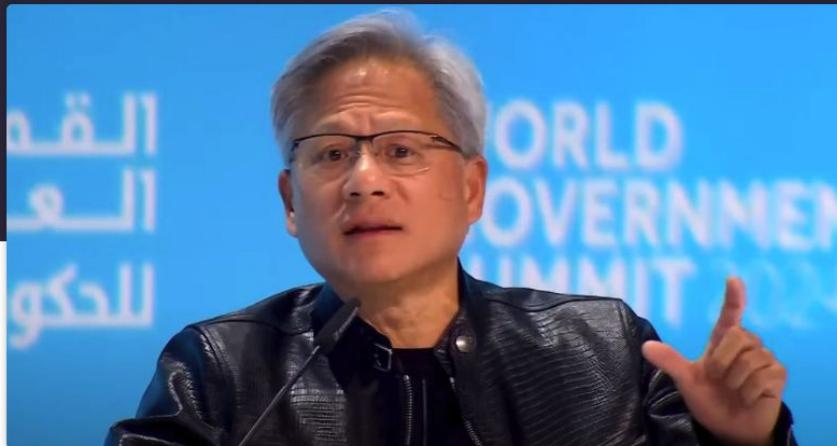
# Growing important of understanding MLOps



# Don't learn to code: Nvidia's founder Jensen Huang advises a different career path



Michael Petraeus · February 23, 2024



WATCH LIVE

In this article

- You're already a programmer
- Specialise
- Pick your street

4.2K people like this. Sign Up to see what your friends like.

*Disclaimer: Unless otherwise stated any opinions expressed below belong solely to the author.*

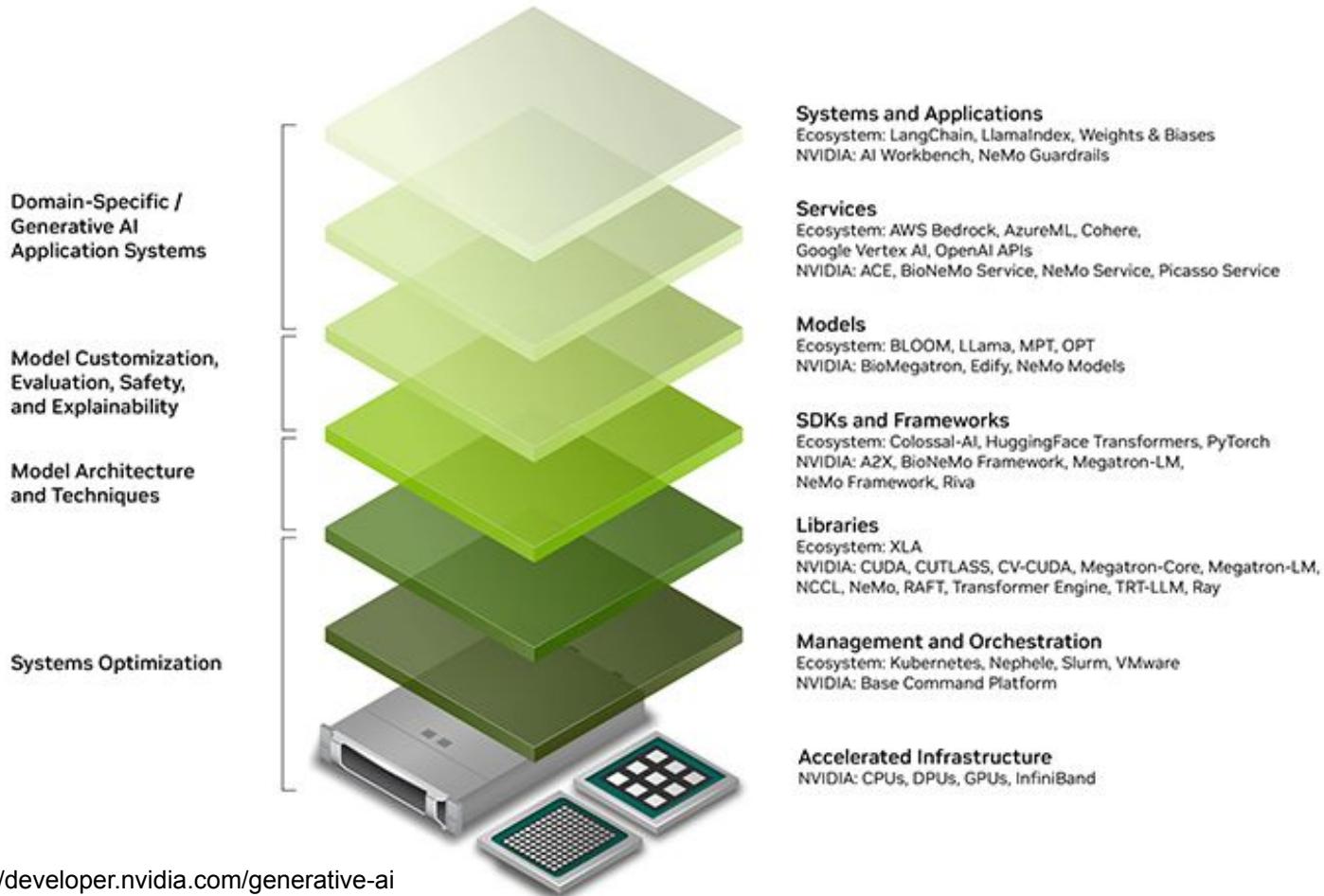
Nvidia, the long-time graphics card company turned AI giant, joined the US\$1 trillion valuation club last year and is now just US\$50 million shy of US\$2 trillion, following the surge in demand for its accelerator cards that large language models like ChatGPT are trained on.

Its multibillionaire founder and CEO, Jensen Huang, is now one of the richest people in the world, with a net worth of close to US\$70 billion, and one that is most sought after to speak about the future that his company is helping to build.

He was one of the headline guests at the World Government Summit in Dubai, which concluded a few days ago. At the event, he spoke at length about the future of the world in this new AI reality, including the impact it's going to have on our careers going forward.

# Generative AI

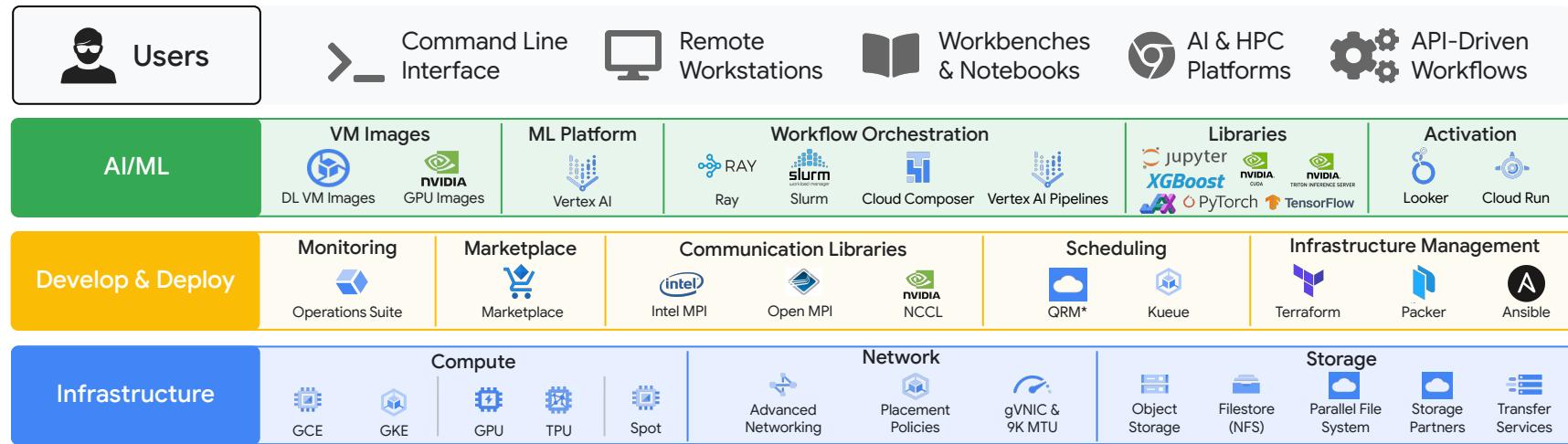
- Use it for Dev (with review)
- Use it as an API (integrated into ML workflows and apps)
- Experiment with frameworks, models, embeddings, etc.
- Monitor emerging use cases



Source: <https://developer.nvidia.com/generative-ai>

NOTE: It's a little hard to read, but this was too good not to add

# Google Cloud's AI Infrastructure Platform



# Needs across the Gen AI spectrum

## Foundation Model Builders

"I train and serve foundation models."

### I care about...

- "Speed and Cost-Efficiency"
- "Control over AI infrastructure"
- "Familiar & flexible SW options"

## Foundation Model Tuners

"I tune and serve foundation and OSS models with my data."

### I care about...

- "Adapt models with my data"
- "Data and AI integration."
- "Choice of AI accelerators"

## Gen AI App Developers

"I quickly build and deploy AI models into my apps."

### I care about...

- "off-the-shelf AI models"
- "easy-to-use AI platforms"
- "high-level of abstraction"



Lower Abstraction

(Optimized for Flexibility  
& "closer" to HW)

Higher Abstraction

(Optimized for Ease-of-Use  
& "farther" from HW)

# AutoML and Low-Code ML Services



# Democratizing Machine Learning

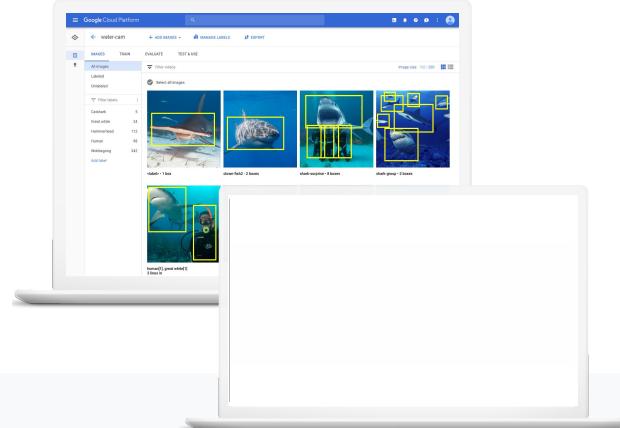
Out of the box

DIY



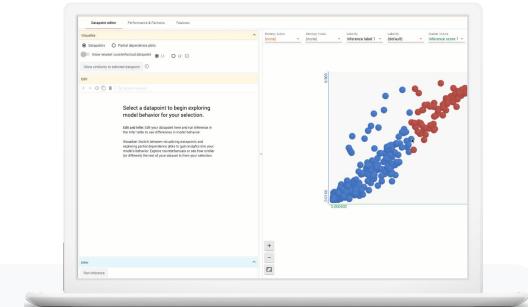
## Pre-trained APIs & AI Solutions

No training data needed, get started right away



## Custom AI with AutoML or BigQuery ML

Easily create custom models  
(A no-code / low-code approach)



## End-to-end AI with core tools

Help data scientists and ML engineers build and deploy AI

Your Data + Pretrained Model

Your Data + Custom Model

## Machine Learning APIs



Serverless

## AutoML



\* Offline / Edge Compatible

Serverless

## SQL-Based Machine Learning



Serverless

## Vertex AI Platform



\* Offline / Edge Compatible

Serverless Deployment

Developer-focused

Analyst

Data Scientist

Your **Data** + Pretrained **Model**

Your **Data** + Custom **Model**

**REST API**

**UI-Driven**

**SQL**

**Code**  
**(TF,Sklearn,Spark)**

**Developer-focused**

**Analyst**

**Data Scientist**





# Vertex AI Pre-Trained Models



## Vision

-  Vision
-  AutoML Vision
-  Video Intelligence
-  AutoML Video Intelligence



## Language

-  Translation
-  AutoML Translation
-  Natural Language
-  AutoML Natural Language



## Conversation

-  Dialogflow
-  Speech-to-Text
-  Text-to-Speech
-  Speaker ID

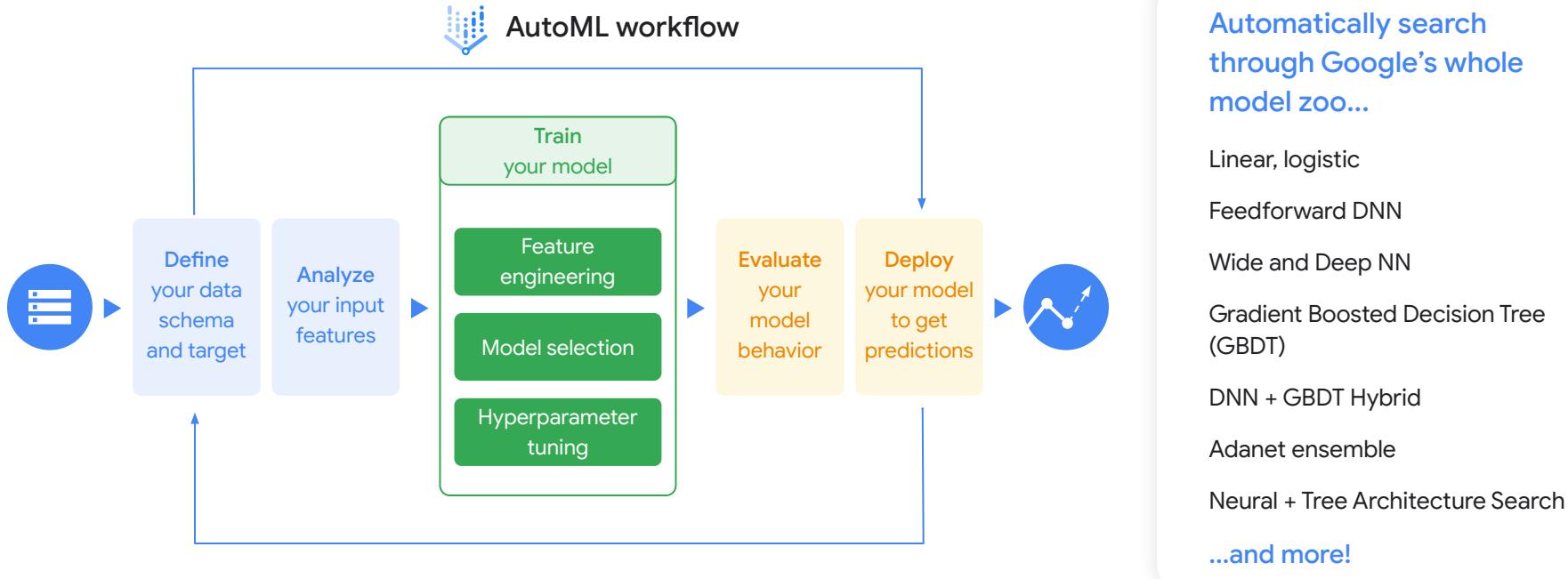


## Structured data

- AutoML Tables
- Tabnet
- Time Series Insights API
- Fleet Routing API
- Vertex AI Forecast

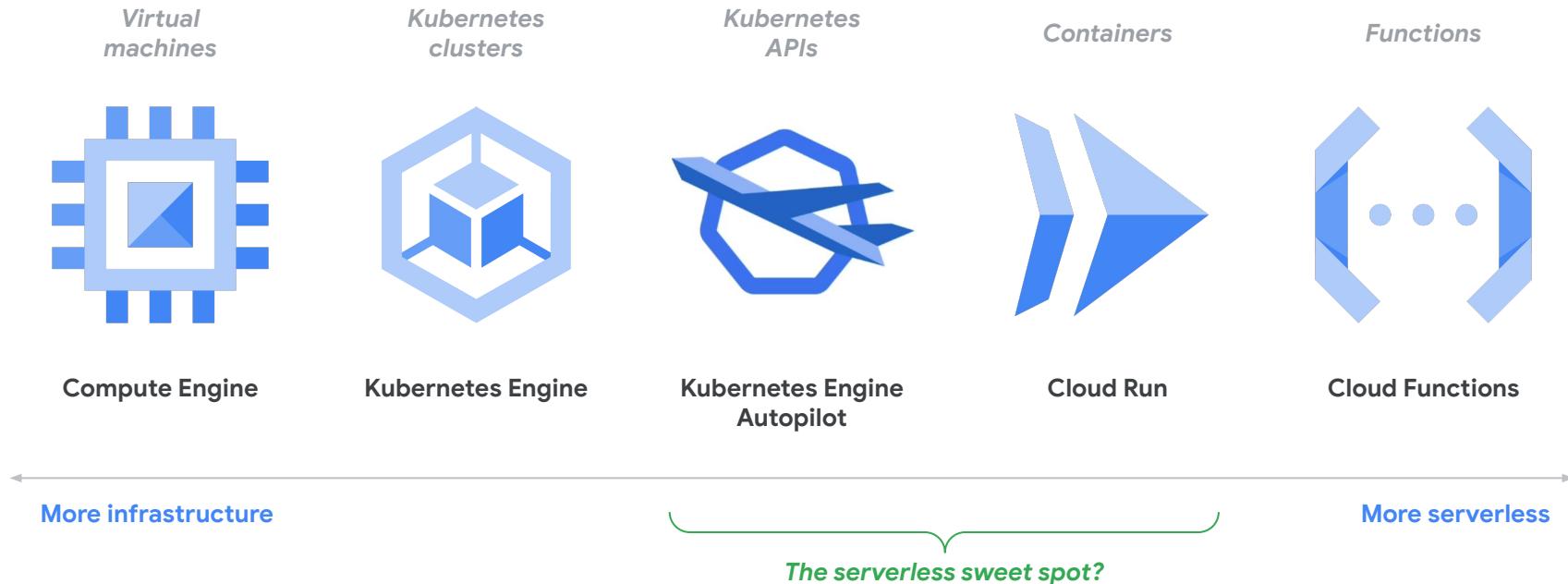
Low/No code

# Point and click to build custom, high-quality models using the AutoML



# Compute continuum towards Serverless

Compute has evolved into managed, serverless services,  
and AI/ML is doing the same.



# Specialization

# Specialization: Considerations for continual learning

- Data Scientist
- Data Analyst
- Business Analyst
- Industry Expert (ie. AI for Healthcare)
- Domain Expert (ie. Specialist in Computer Vision)
  
- ML Engineer / ML Ops
- AI Infrastructure
- Cloud ML Engineer
  
- XAI Researcher/Developer
- GenAI App Developer
- Deep Learning or LLM Engineer/Researcher

[Menti](#)

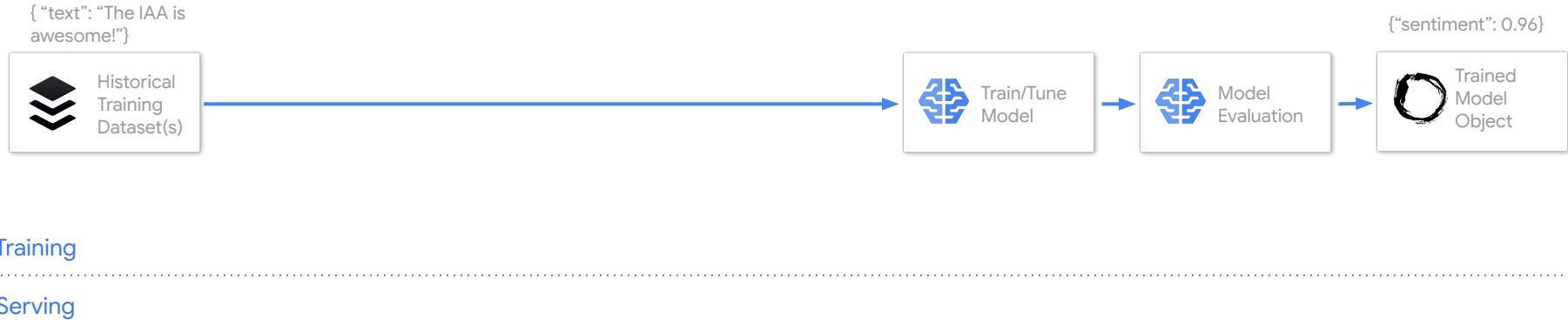
Part 3

# What is the most challenging part of being a Data Scientist?

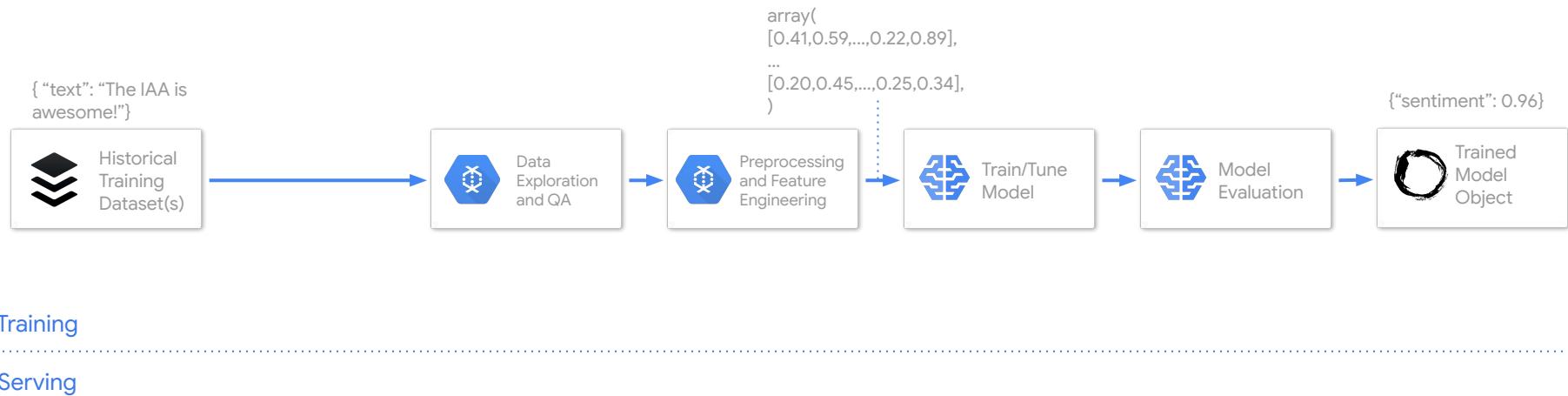
Waiting for responses ...

# Architectures

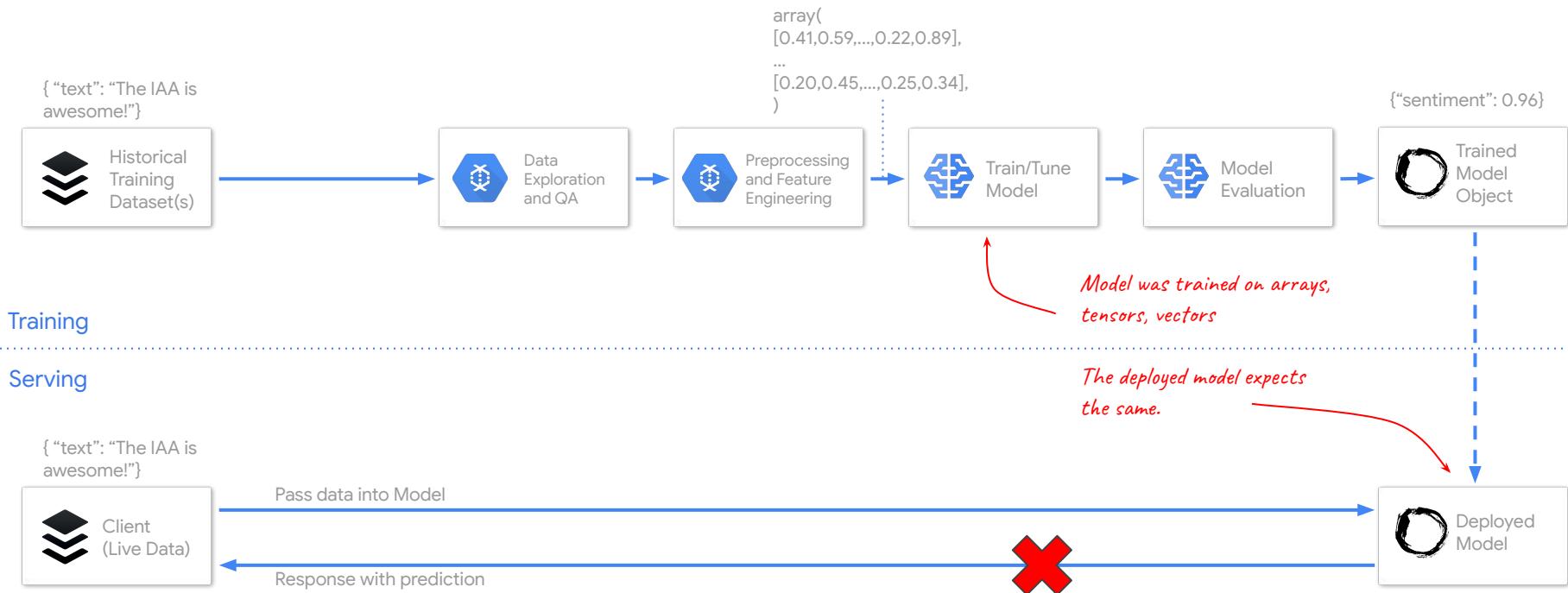
# Let's start with a simple ML example



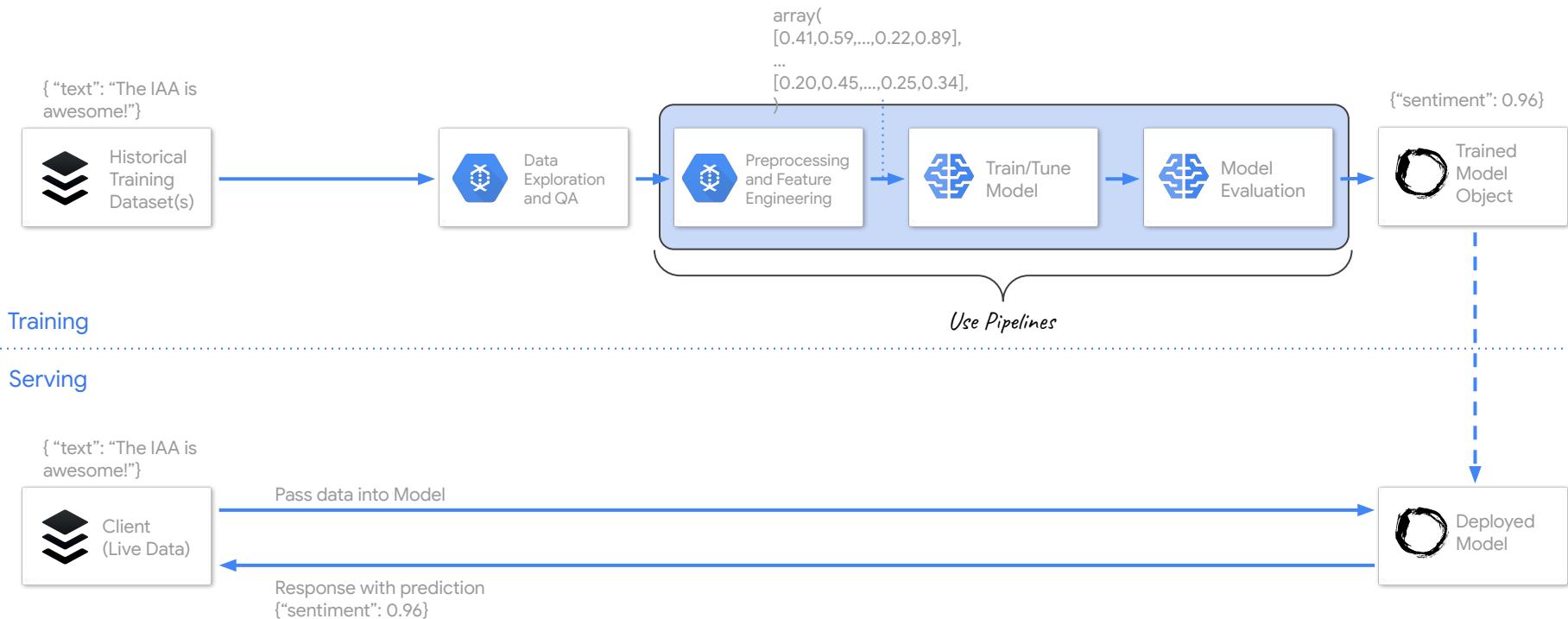
# Machine Learning Pipeline



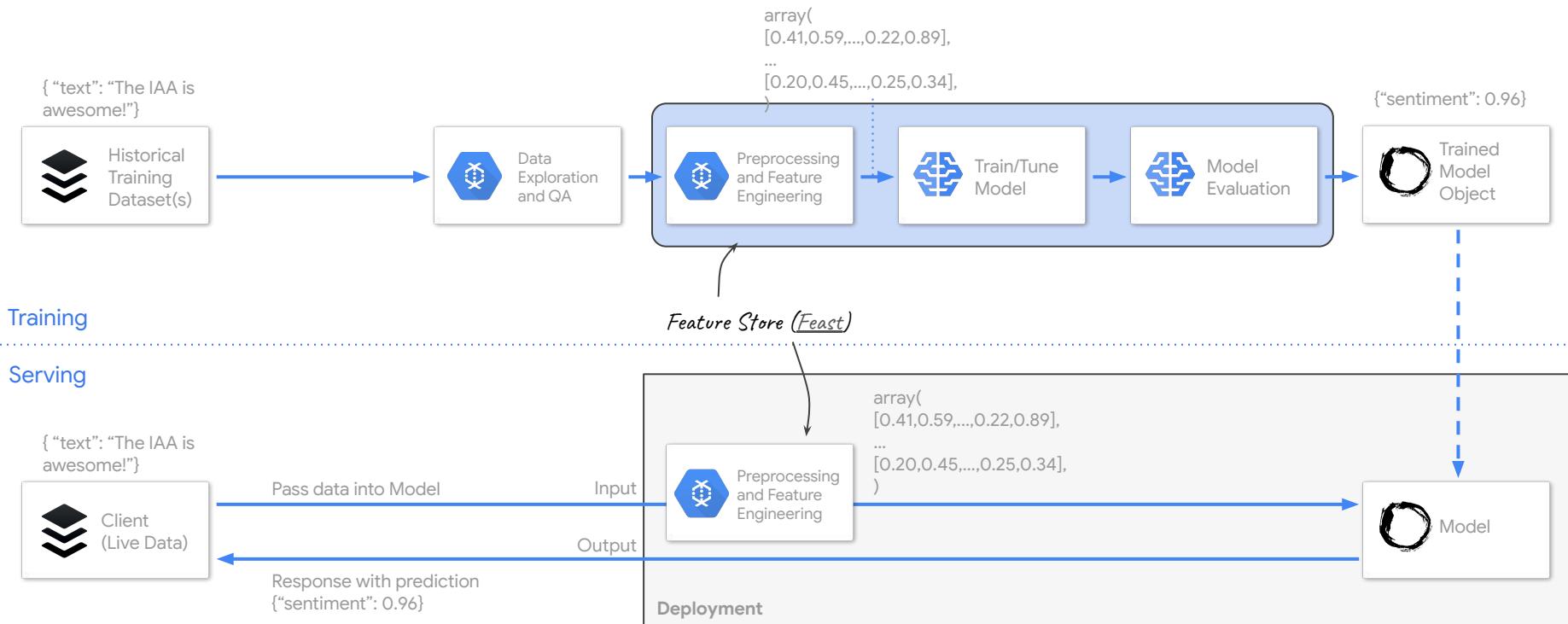
# Machine Learning Pipeline



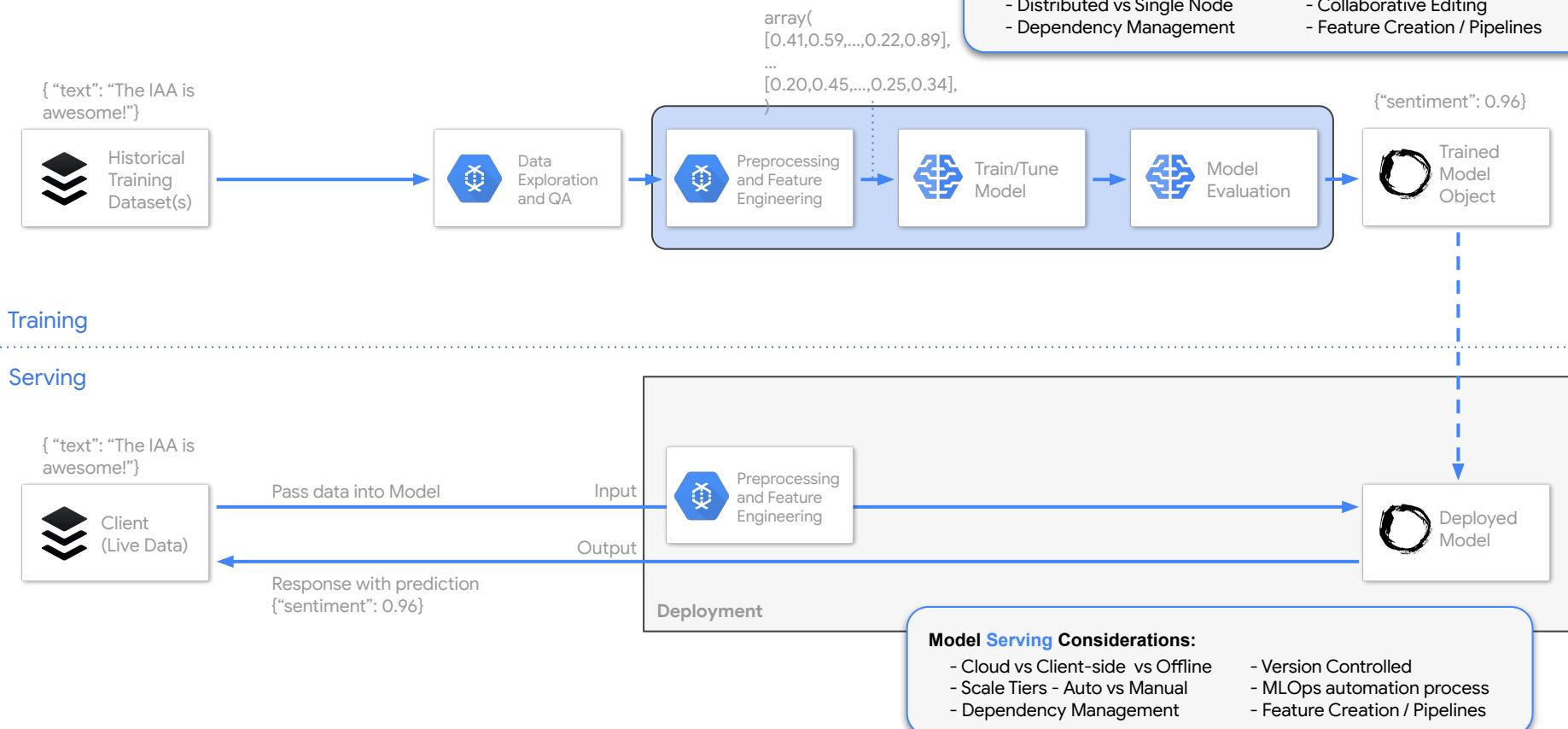
# Machine Learning Pipeline



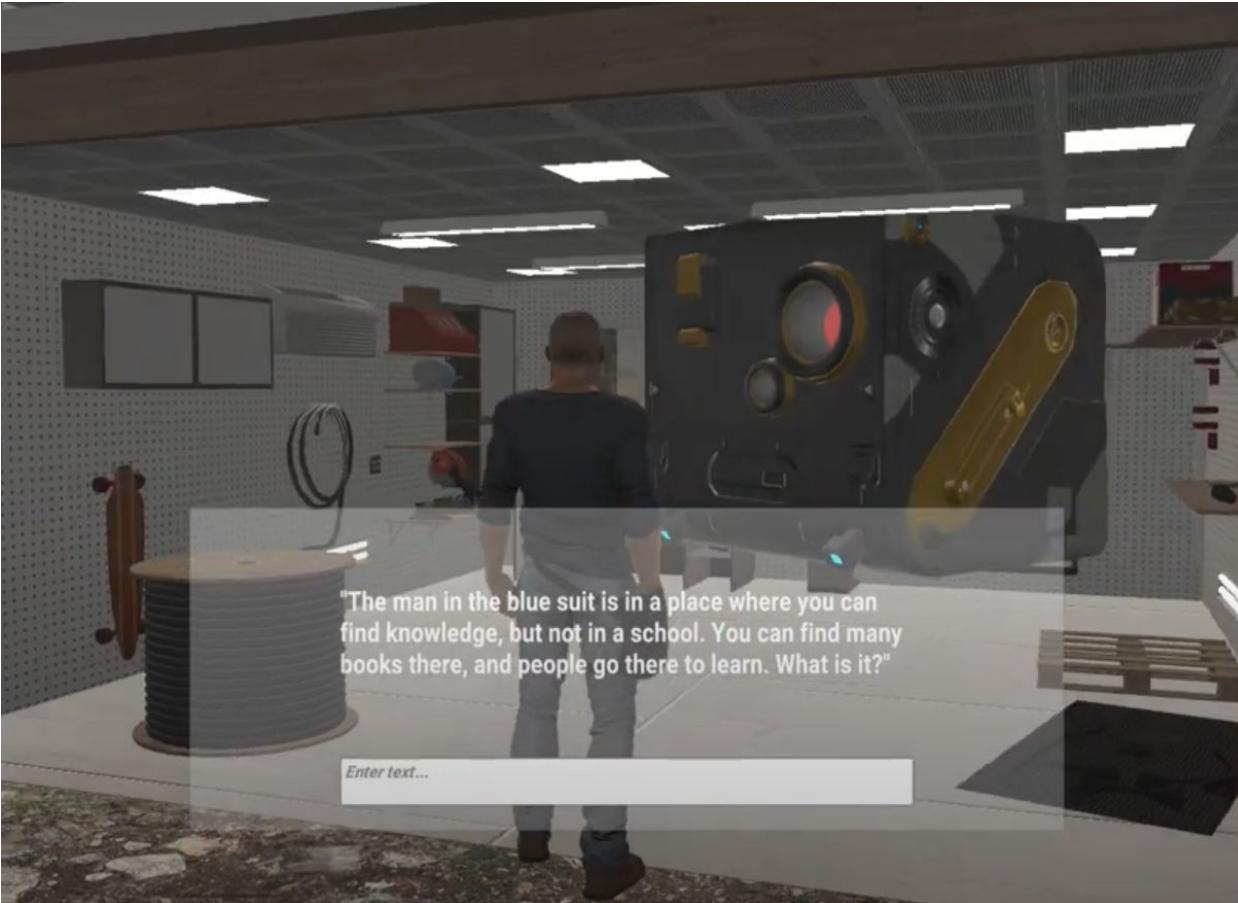
# Machine Learning Pipeline



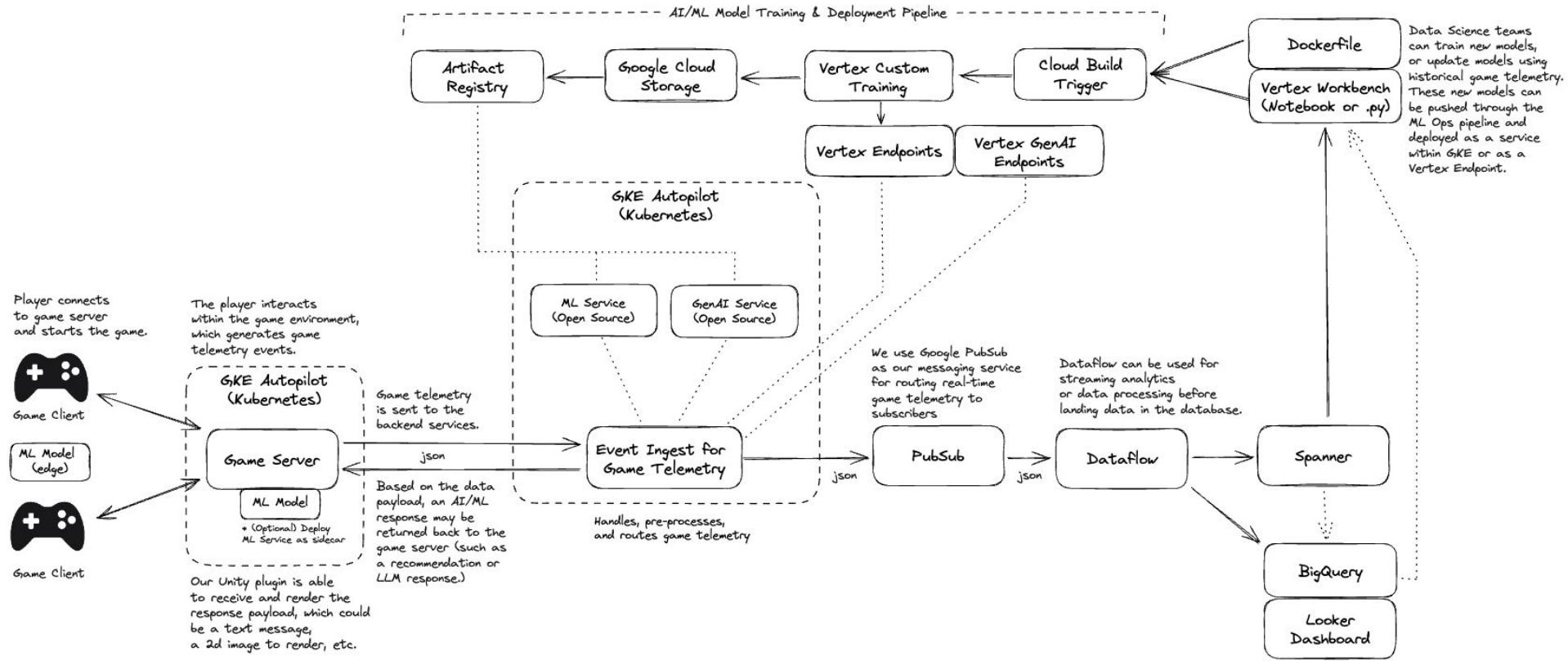
# Machine Learning Pipeline



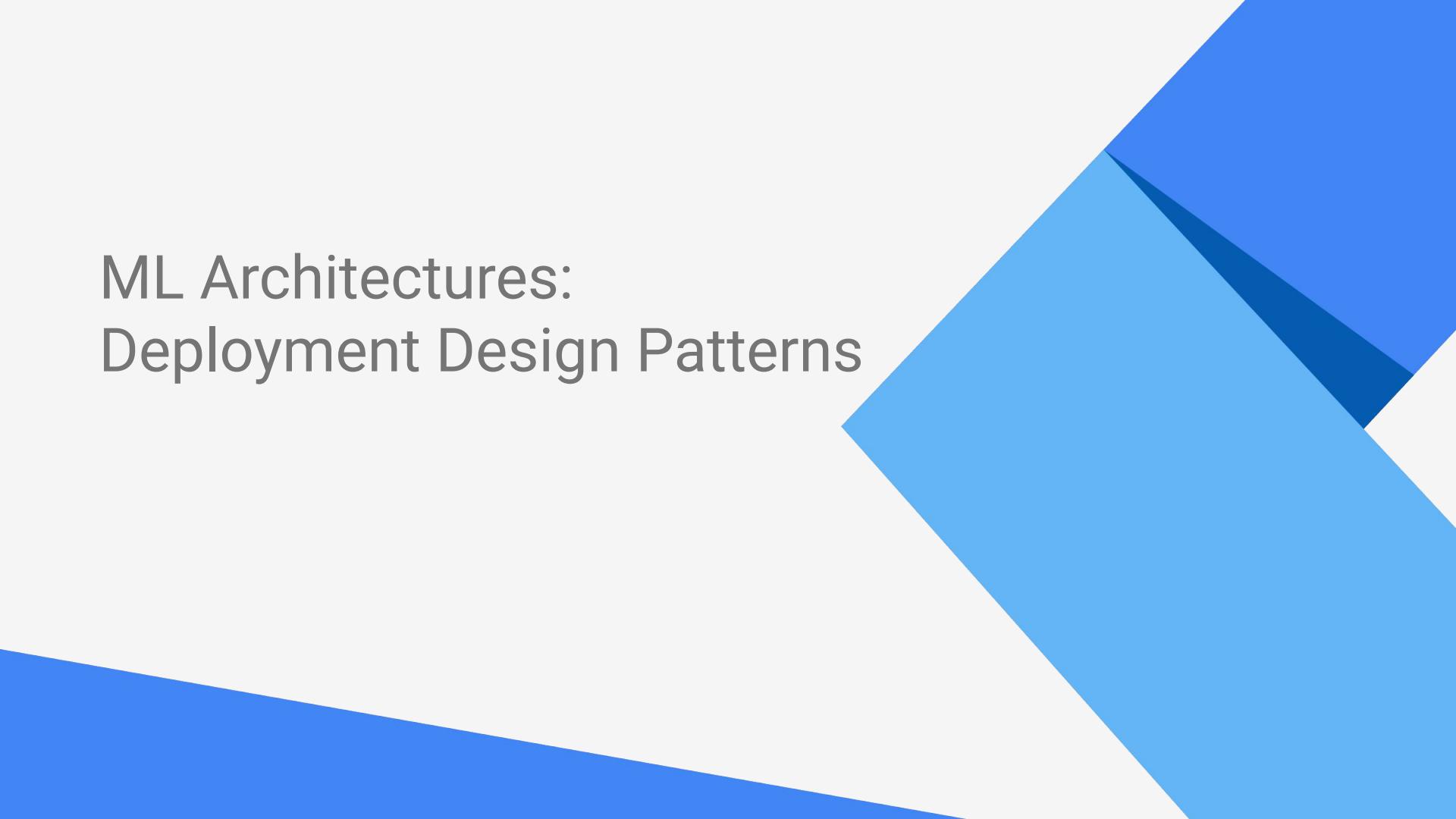
# Gaming Reference Architecture (Example Use Case)



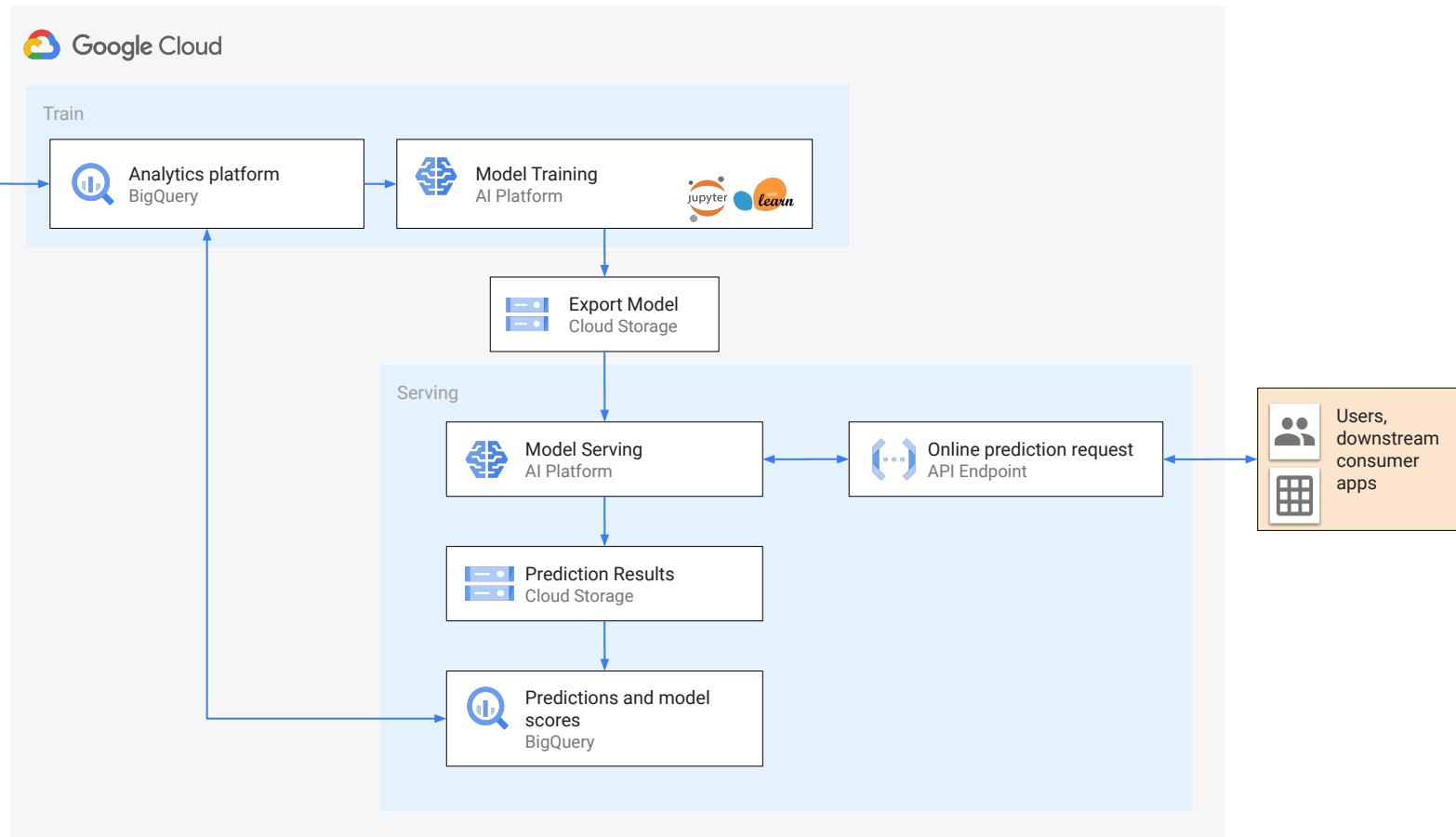
# Gaming Reference Architecture (Example Use Case)



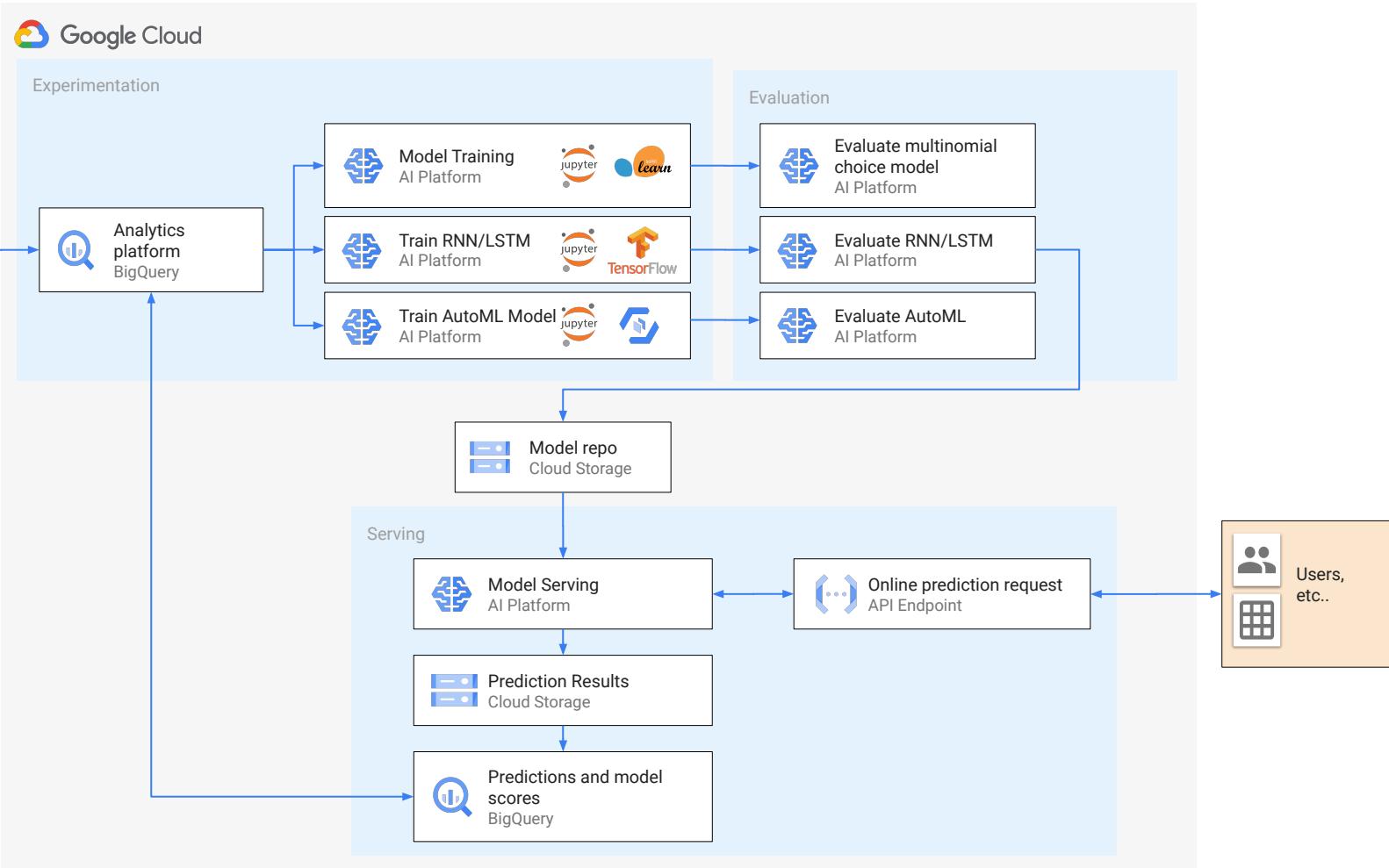
# ML Architectures: Deployment Design Patterns



# Crawl...



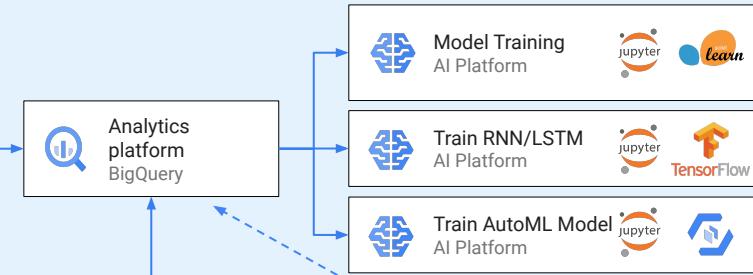
# Walk...



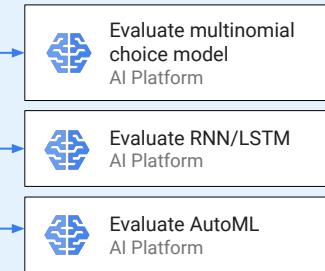
# Run...



## Experimentation

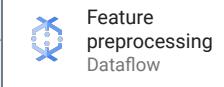
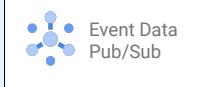


## Evaluation



Batch inputs

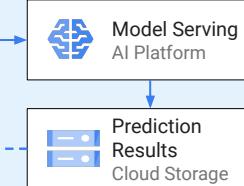
Streaming inputs



## Model Retraining



Serving





@PVERGADIA  
THECLOUDGIRL.DEV  
8.13.2020

# How to build a **scalable** DATA ANALYTICS PIPELINE



## CAPTURE

Data ingestion at any scale



CLOUD PUB/SUB  
Scaled messaging platform



DATA TRANSFER SERVICE  
Fast data migration from saas apps



STORAGE TRANSFER SERVICE  
Data migration from other cloud or on-prem



CLOUD IoT CORE  
Stream events from IoT devices

## PROCESS

Reliable streaming data pipeline



CLOUD DATAFLOW  
Stream and batch processing



Hadoop + Spark  
CLOUD DATAPROC  
Managed Hadoop & Spark platform



CLOUD DATAPREP  
Data prep using visual tool



CLOUD DATA FUSION

Fully managed, code-free data integration service to manage ETL/ELT pipelines and also track lineage of that data.

## STORE

Data lake and data warehousing



CLOUD STORAGE  
Use as your data lake for structured and unstructured data



BIGQUERY STORAGE  
Cloud-native, highly scalable serverless data warehouse



BIGQUERY  
Analysis engine



Unified view of all datasets

## ANALYZE

Data warehousing

## USE

Advanced analytics



CLOUD AI PLATFORM  
For machine learning



TENSORFLOW  
For machine learning



LOOKER  
For your analysis

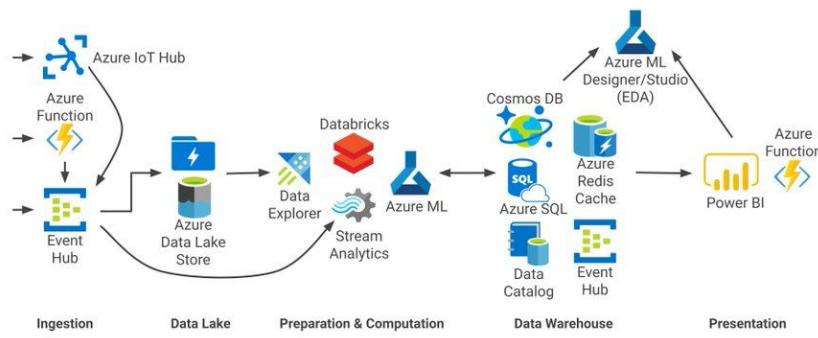
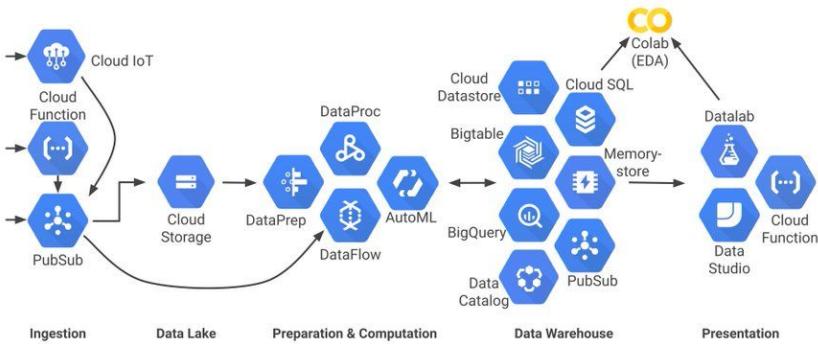
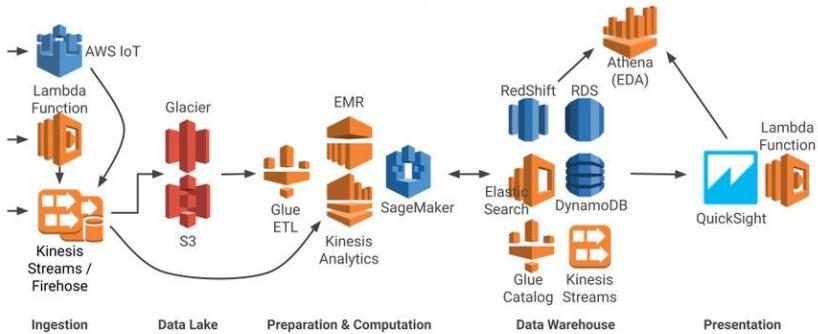


SPREADSHEETS  
For your analysis



CLOUD COMPOSER

Fully managed, workflow orchestration built on Apache Airflow



Google Cloud

Next Up:  
SQL and NoSQL

Thanks!

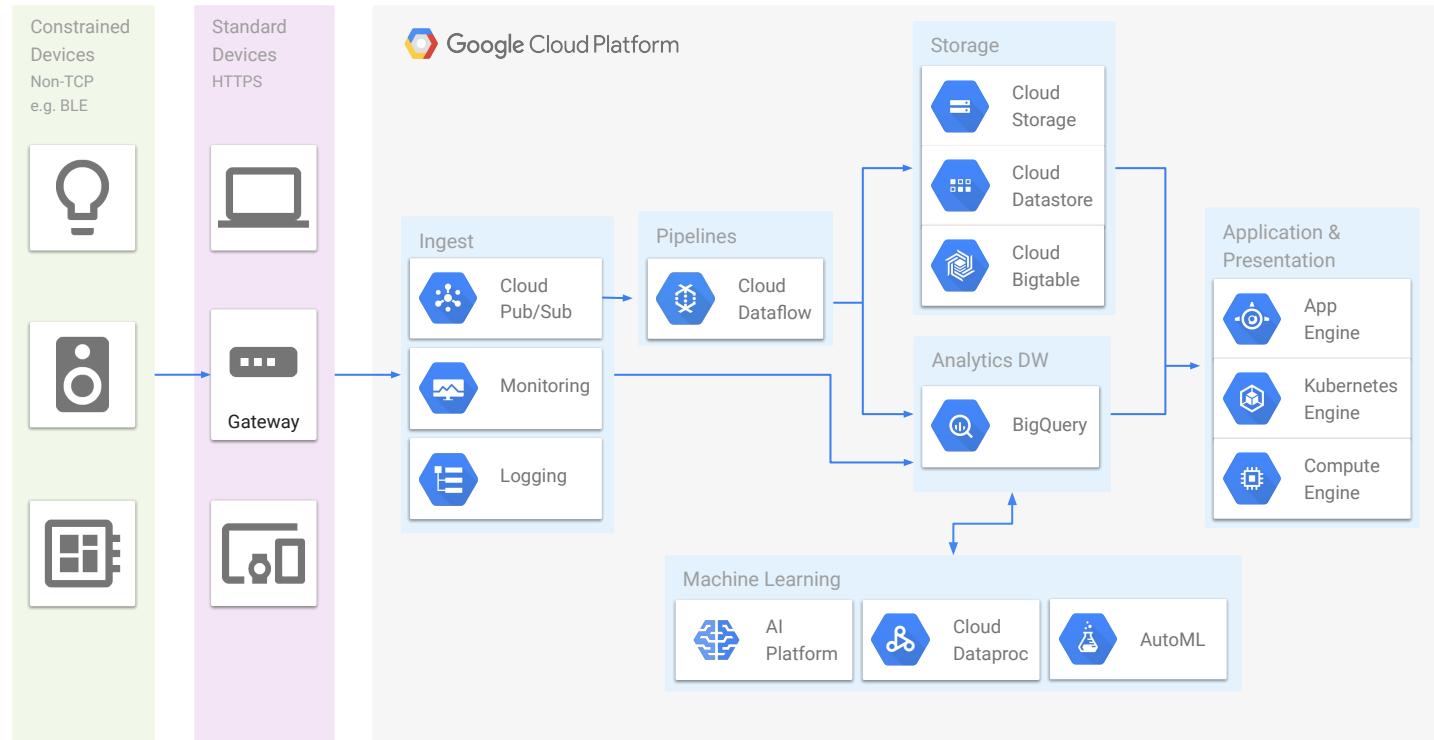
# Advanced Big Data: Distributed Machine Learning

**Dan Zaratsian**

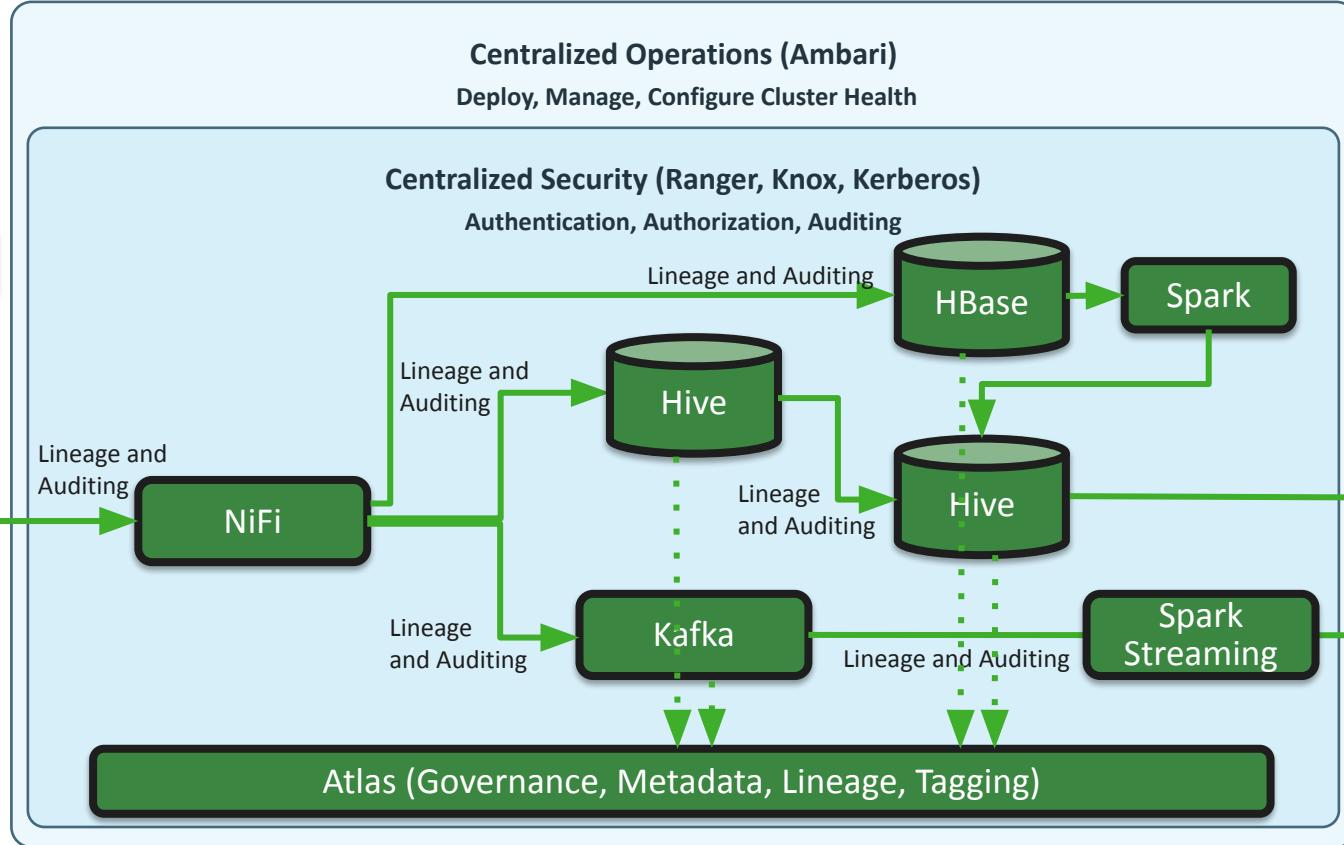
AI/ML Architect, Gaming Solutions @ Google

# Real Time Stream Processing IoT

## Architecture: General > Real Time Stream Processing - Internet of Things



# Reference Architecture



Google Cloud

Cloud Deployment (Google, ...)