



---

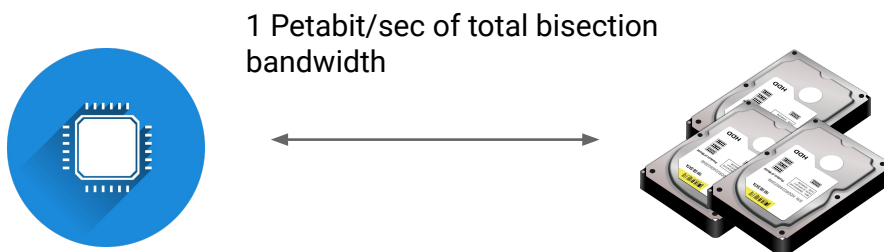
## Summary

**Carolyn Ujcic**

Cloud Machine Learning Engineer

Congratulations! You've made it to the end. We hope you had fun learning new and innovative approaches to solving some of the toughest big data and machine learning challenges out there. Let's do a brief review of what you have learned in this course. At the end I'll also recommend the next courses and labs you should take on your big data journey with GCP.

## Google's data center network speed enables the separation of compute and storage



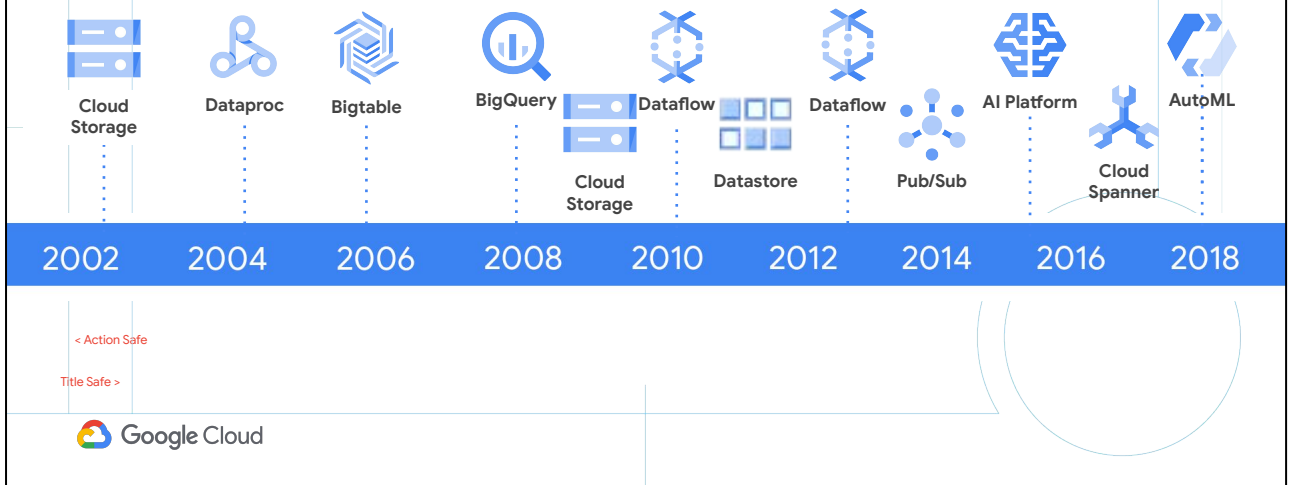
We first learned that Google technologies are based on really fast networking speeds at the petabit per second level.

You no longer need to do everything on a single machine or single cluster of machines with their own dedicated storage.

If every machine can talk to every other machine at 10 Gbps, racks don't matter for data analytics and ML training.

<https://cloud.google.com/blog/products/gcp/bigquery-under-the-hood>

Google Cloud opens up that innovation and infrastructure to you



You then saw Google technologies grow and mature as the internet as grown. This suite of Google solutions is available to you as part of Google Cloud Platform.

## There are many roles in a data-driven organization

### Data Analyst

Analyst

### Data Engineer

Data Engineer

### Applied ML Engineer

Data Scientist

Ethicist

Statistician

Social Scientist

Applied ML  
Engineer

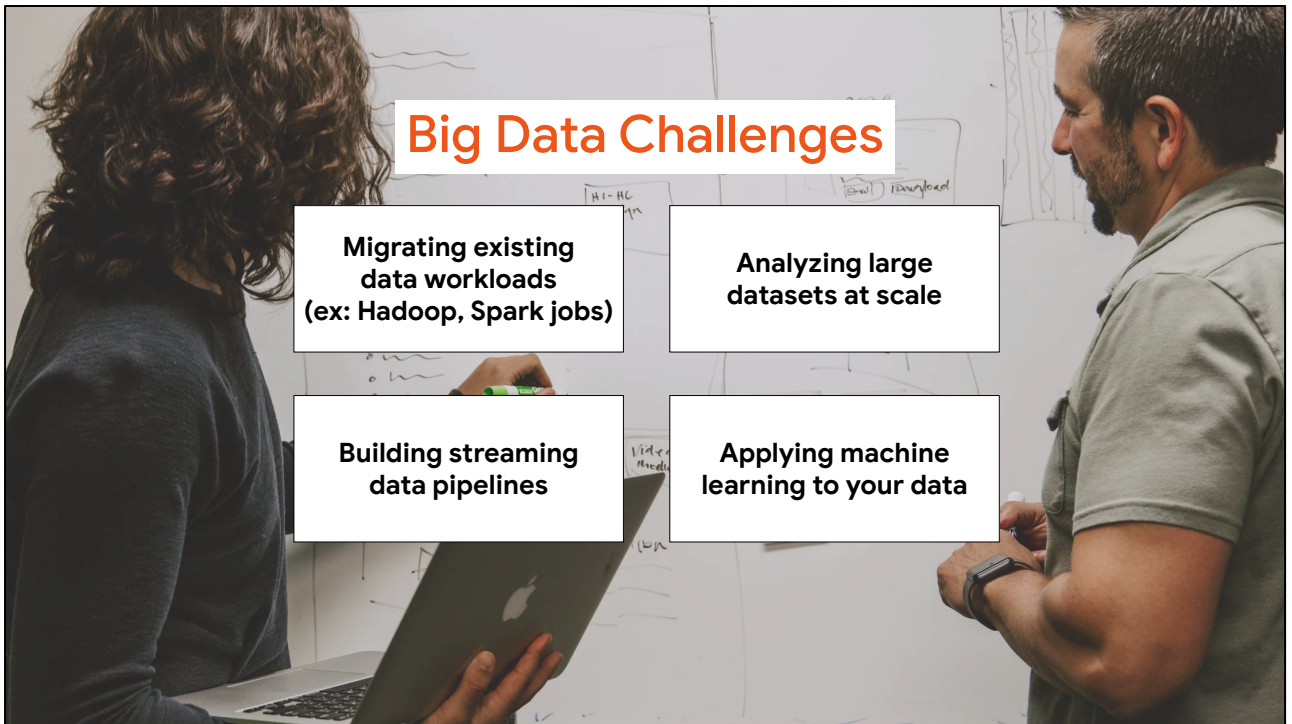
Researcher

### Tech Lead

Analytics Manager

Decision Maker

You explored which tools are used by which data roles in your organization from analysts, data engineers, ML engineers, and tech leads.



## Big Data Challenges

**Migrating existing  
data workloads  
(ex: Hadoop, Spark jobs)**

**Analyzing large  
datasets at scale**

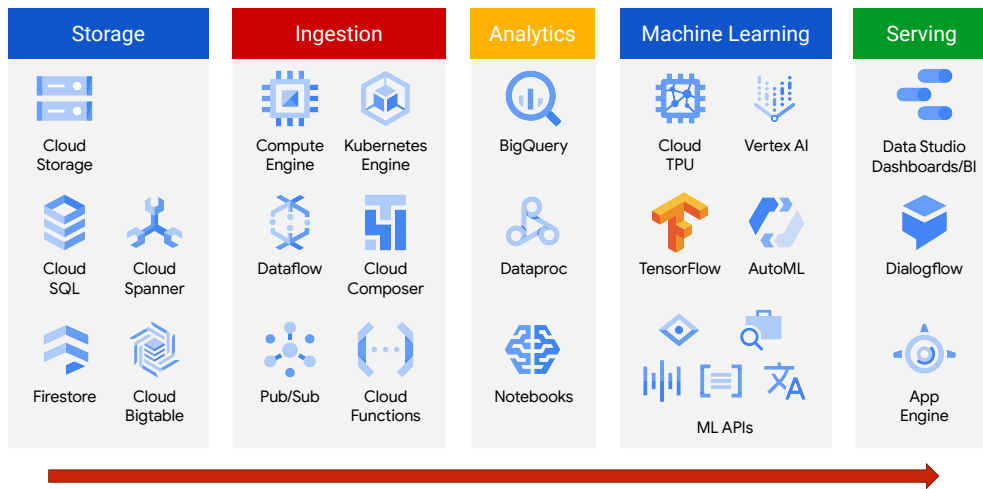
**Building streaming  
data pipelines**

**Applying machine  
learning to your data**

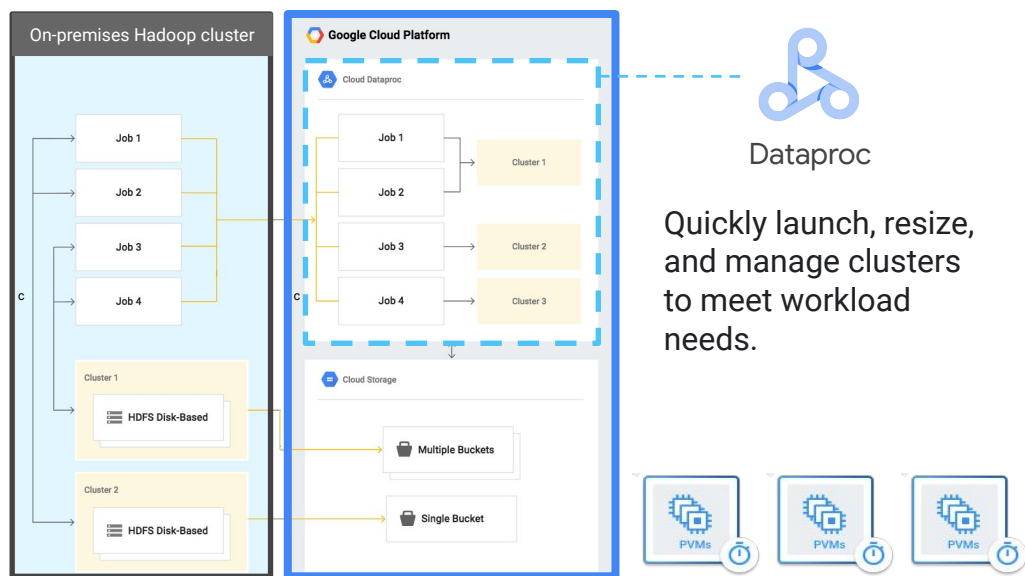
And which of the core big data challenges each role was likely to face.

- You practiced in your hands on labs how to:
- Migrate existing workloads to the Cloud with your Spark ML job
- Analyzing large datasets at scale with BigQuery
- Building streaming data pipelines with Cloud Dataflow
- And Applying machine learning throughout the course on both structured and unstructured datasets.

## The suite of big data products on Google Cloud

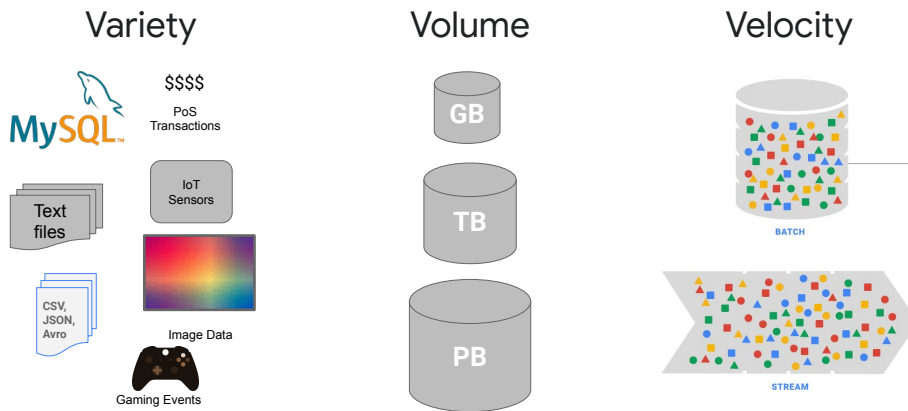


You learned where and when to use the fundamental big data products and services to build streaming data pipelines into your data warehouse for machine learning.



You learned that Cloud Dataproc can provide flexibility and scalability for your Hadoop workloads without the need for cluster management. You saw the ease in which you can re-size your clusters as you need to meet workload demands.

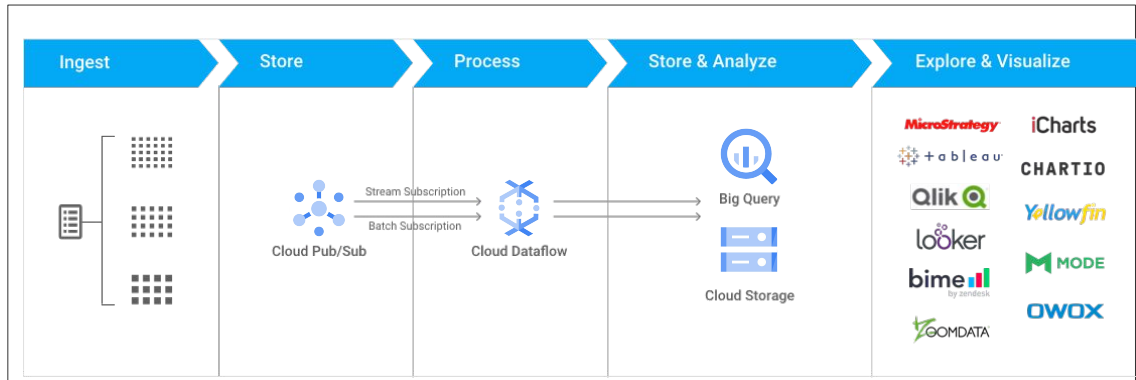
## Modern big data pipelines face many challenges



Then we introduced the challenges of streaming data pipelines which come from a variety of different data sources and at large volumes. You also had to solve for streaming data and building an architecture around receiving and publishing distributed IoT device messages into your pipeline.

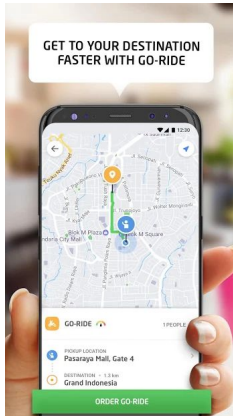


# Google Cloud's serverless big-data pipeline



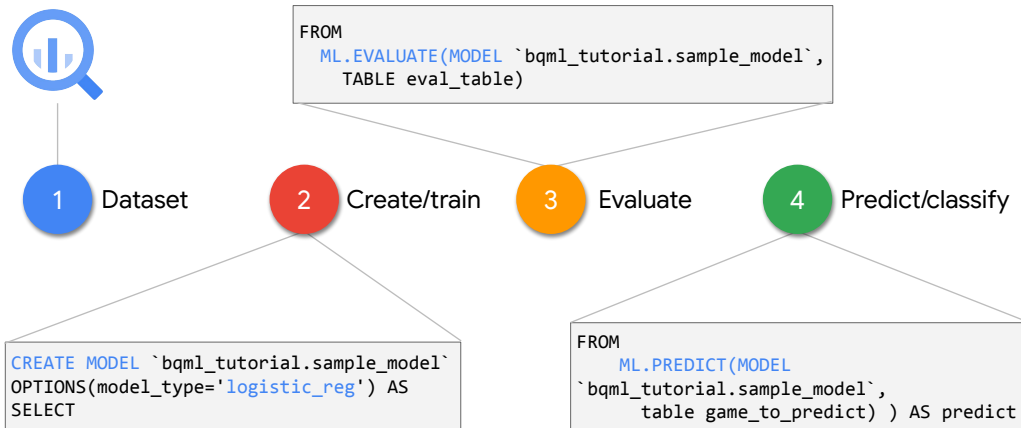
You saw and built an end-to-end architecture for solving those streaming pipeline changes. You learned that Cloud Pub/Sub receives those messages in a pub/sub topic and is then able to publish them to subscribers -- one of which is Cloud Dataflow for ingestion and transformation into BigQuery for analysis and downstream visualization.

GO-JEK brings goods and services to over  
2 million families in 50 cities in Indonesia



You explored real customer use cases like Go-JEK who use GCP to bring goods and services to over 2 million families in Indonesia.

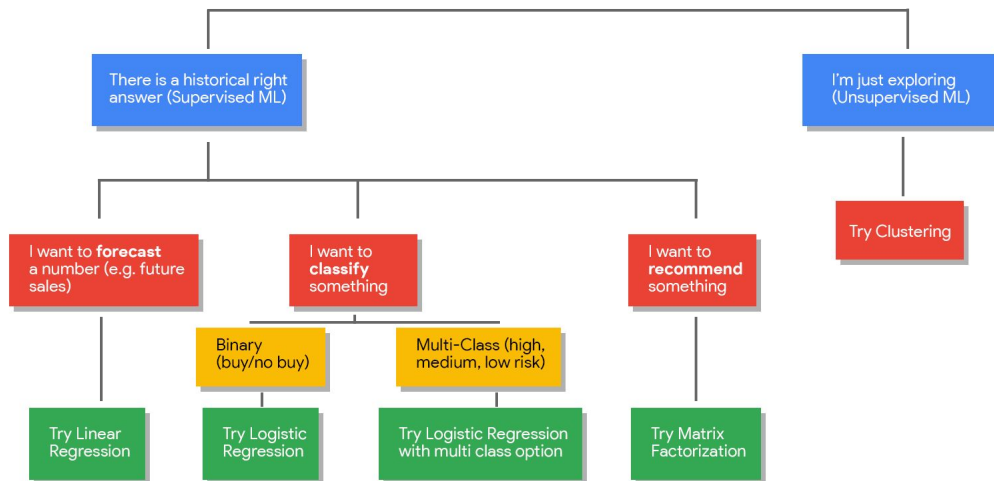
## Recap: BigQuery ML



You build custom machine learning models on your structured datasets with BigQuery ML. Recall the four key steps:

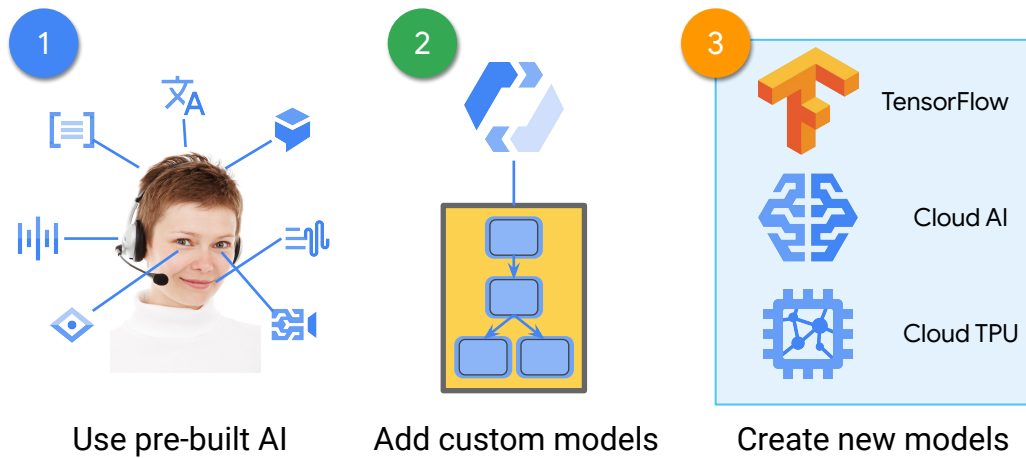
- 1) Collect, clean, and transform the data
- 2) Create your model and selecting a model type
- 3) Evaluate model performance
- 4) And finally, predict with the model on unknown data

## Choose the right model type for your structured data use case



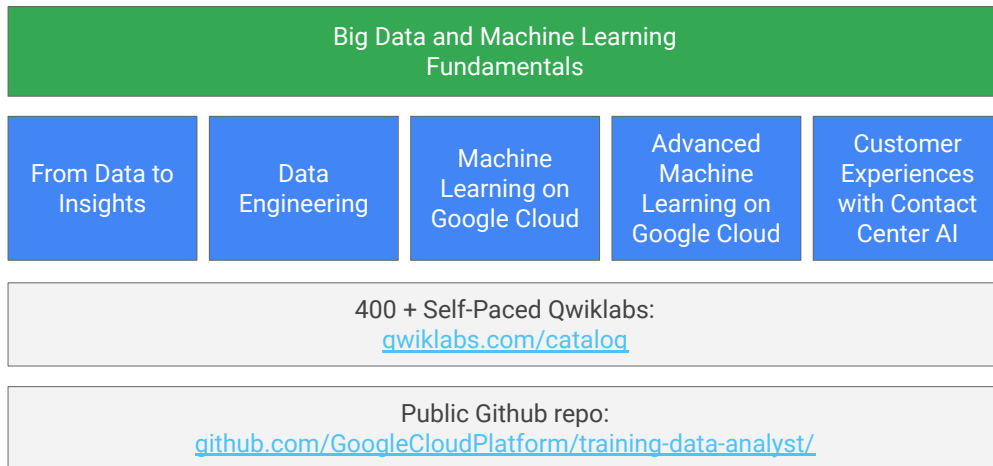
We reviewed how to choose an appropriate model type for BigQuery ML based on your use case. We focused primarily on forecasting and classification where you used linear regression and logistic regression models.

## Artificial Intelligence application strategy



Lastly we reviewed the three main approaches to an AI strategy for your organization and projects. You can leverage the AI building blocks like Cloud Vision and Natural Language and you saw how we could extend them with our own data by using AutoML for a custom model with no code required.

This course is the start to your big data journey



We hope after taking this course you have gotten a taste for some of the real-world big data applications you can tackle for your next project. We'll provide resources at the end of the course on each of the topics covered and what the subsequent courses are that we recommend you take to dive deeper in your knowledge of Google Cloud.

This fundamentals course is just the start to your big data journey. If you want to deep dive into BigQuery for data analysis check out our specialization called From Data to Insights with Google Cloud. If you want to continue building on your knowledge of pipelines try Data Engineering with Google Cloud. If you are ready for custom ML model building with TensorFlow try our ML on Google Cloud and Advanced ML on Google Cloud specializations. If you want to learn how to design, develop, and deploy customer conversational solutions using Contact Center Artificial Intelligence (CCAI), then the Customer Experiences with Contacts Center AI specialization is designed for you.

Lastly, there are over 200 self-paced labs like the ones you took in this course for additional practice.



From all of us here in Google Cloud, thank you for exploring the possibilities with big data and machine learning as part of this course and we hope to see you again.