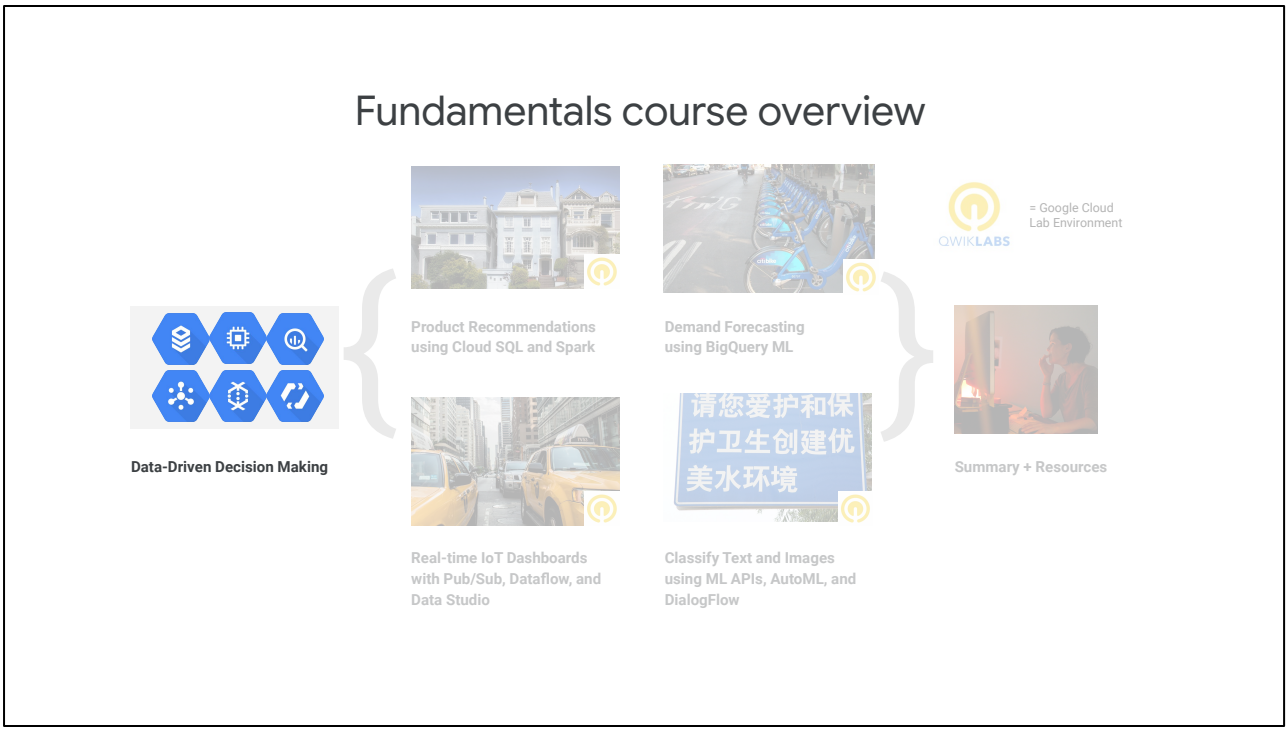
Google Cloud Big Data and Machine Learning Fundamentals

Lak Lakshmanan
Tech Lead, Big Data + ML

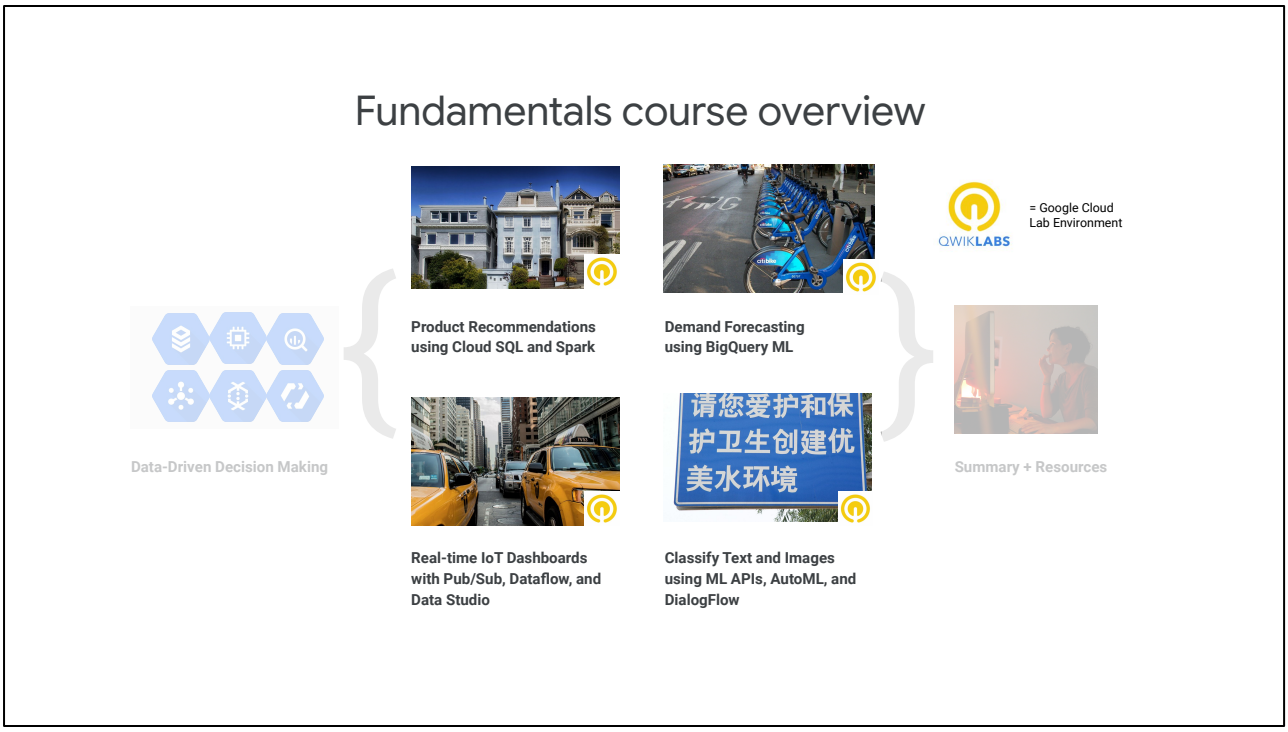
Hi, I'm Lak, and I lead a team that helps customers of Google Cloud successfully build applications that use our big data and machine learning products. Among the things we do is creating big data and machine learning training courses and labs.

This course, Google Cloud Big Data and Machine Learning Fundamentals, was designed to showcase real world data and ML challenges and give you practical hands-on expertise in solving those challenges using Google Cloud in our labs. It's a critical course to master because it covers the most common use cases you and your team will encounter on your big data journey.

This course is divided into six content modules.

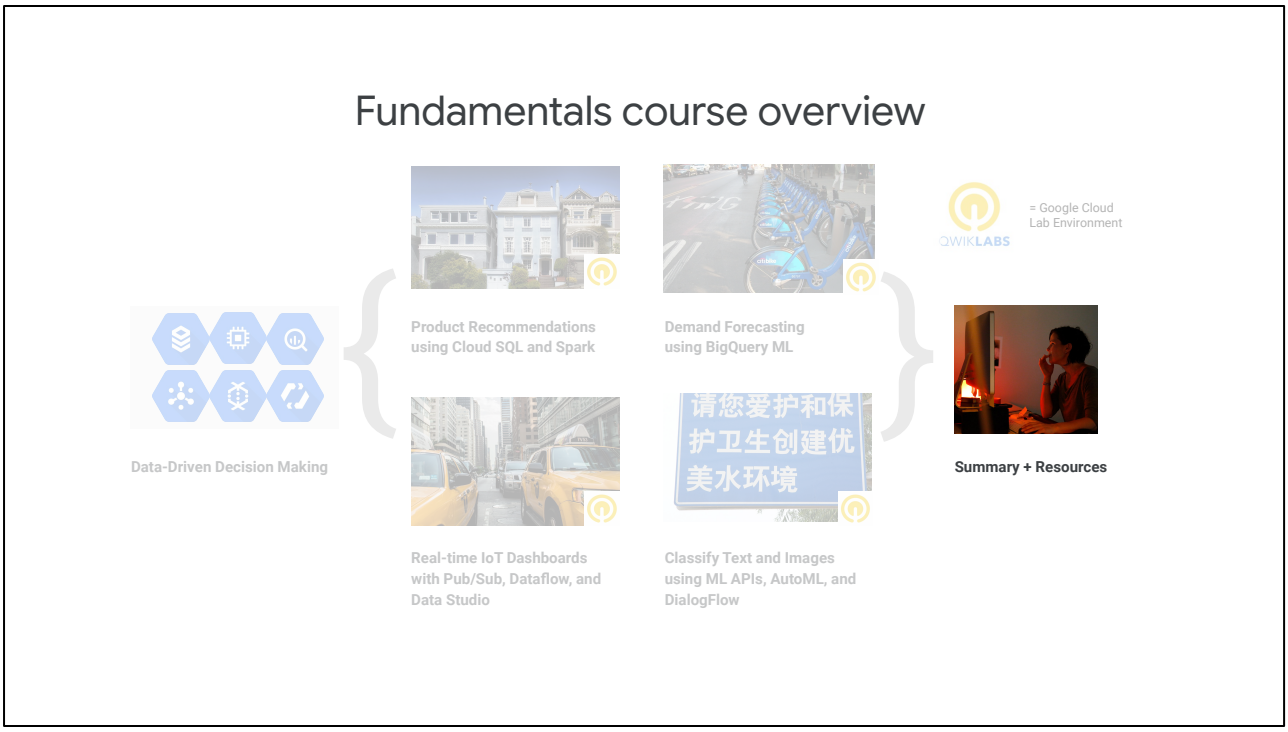


In the first module on Data-Driven Decision Making, you'll learn all about the data and ML tools available on Google Cloud Platform from a high level organizational perspective.



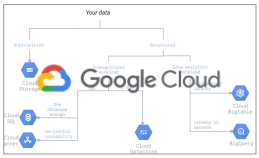
In the next four modules, you will be introduced to Google Cloud products in context as they are employed to solve real-world problems.

In the four modules, in addition to expanding your knowledge about our products and platform, you will also get to practice with hands-on Qwiklabs.

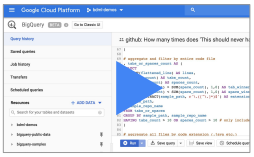


Finally, in the summary module, we'll do a recap of everything you learned in this course and provide you with additional resources on the topics you've practiced.

Welcome to Big Data and Machine Learning Fundamentals with Google Cloud Platform



Learn Google Cloud Platform



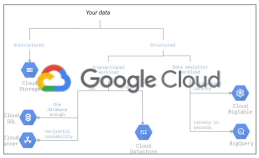
See it in action with demos

Practice with Qwiklabs

Analyze Customer Use Cases

In each module, this is the typical order we'll follow:

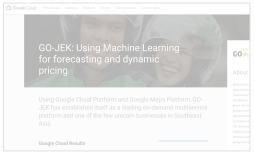
Welcome to Big Data and Machine Learning Fundamentals with Google Cloud Platform



Learn Google Cloud Platform

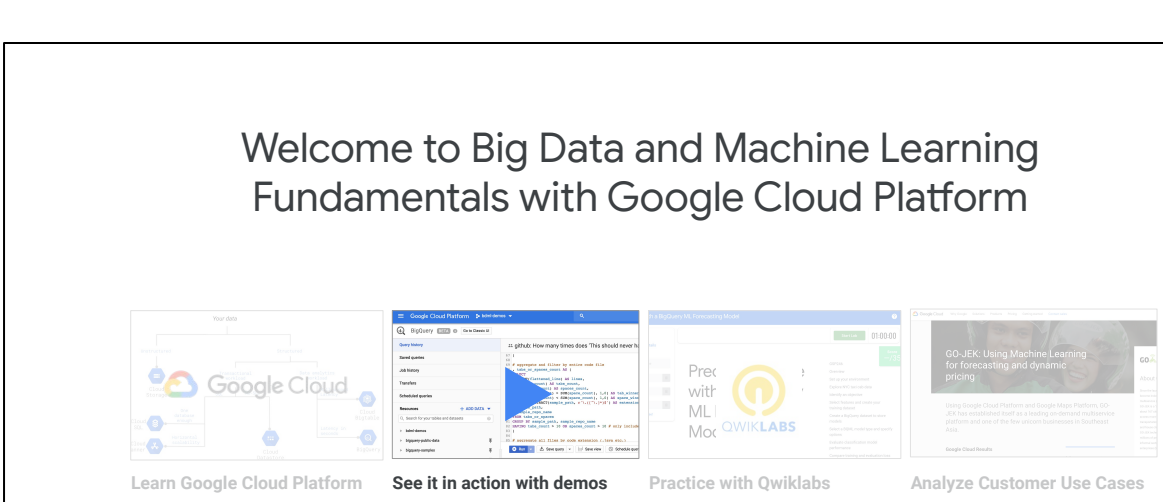
See it in action with demos

Practice with Qwiklabs

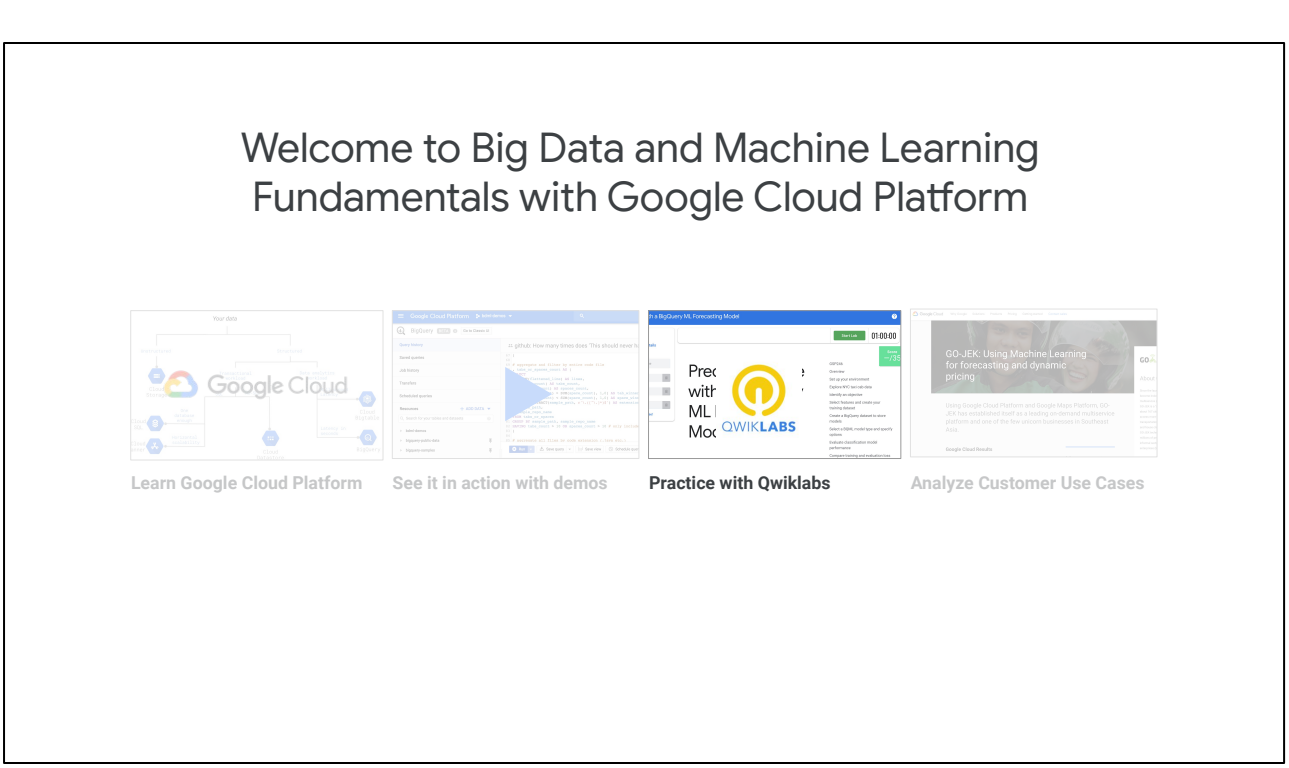


Analyze Customer Use Cases

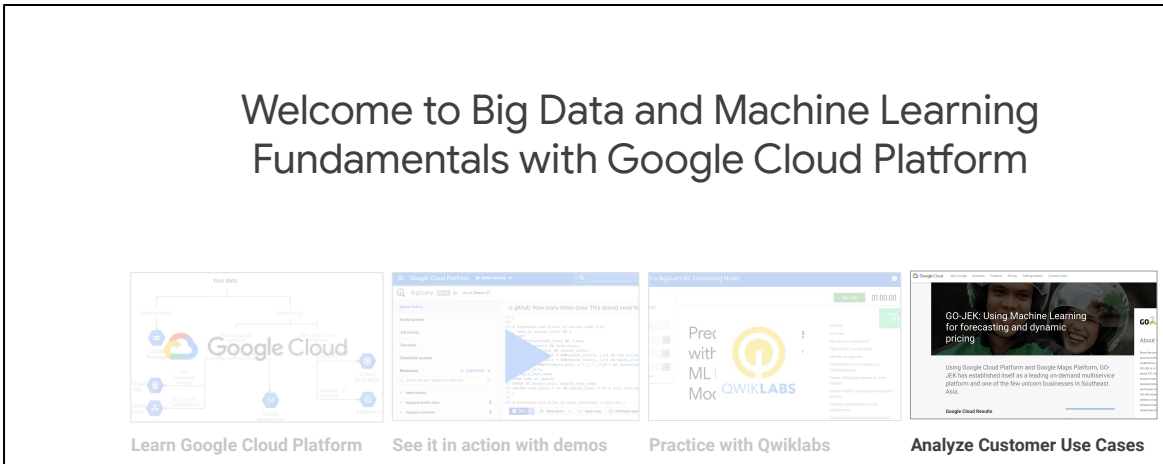
First we'll have lectures from subject matter experts where we will introduce the big data scenario and the challenges, and how to address them with cloud technologies.



Next, you'll see a demo of the solution in action, which will highlight key features that you will learn and practice in your labs.



After you understand the scenarios and have watched the demos, it's time for you to practice with Qwiklabs in a real Google Cloud Platform account.



Finally, you'll explore real customer use cases and architectures so you can familiarize yourself with best practices and get inspired for your own solutions.

In other words, we describe a common class of big data and ml problems and hone in on a specific problem. Then, show you a demo of a solution. Then, talk about why we built the solution that way, using that opportunity to cover how and when to use the various products used in the solution. Finally, we widen the lens and show you real world applications that are variants of the principles covered in the chapter.

[pause]

An explosion of data

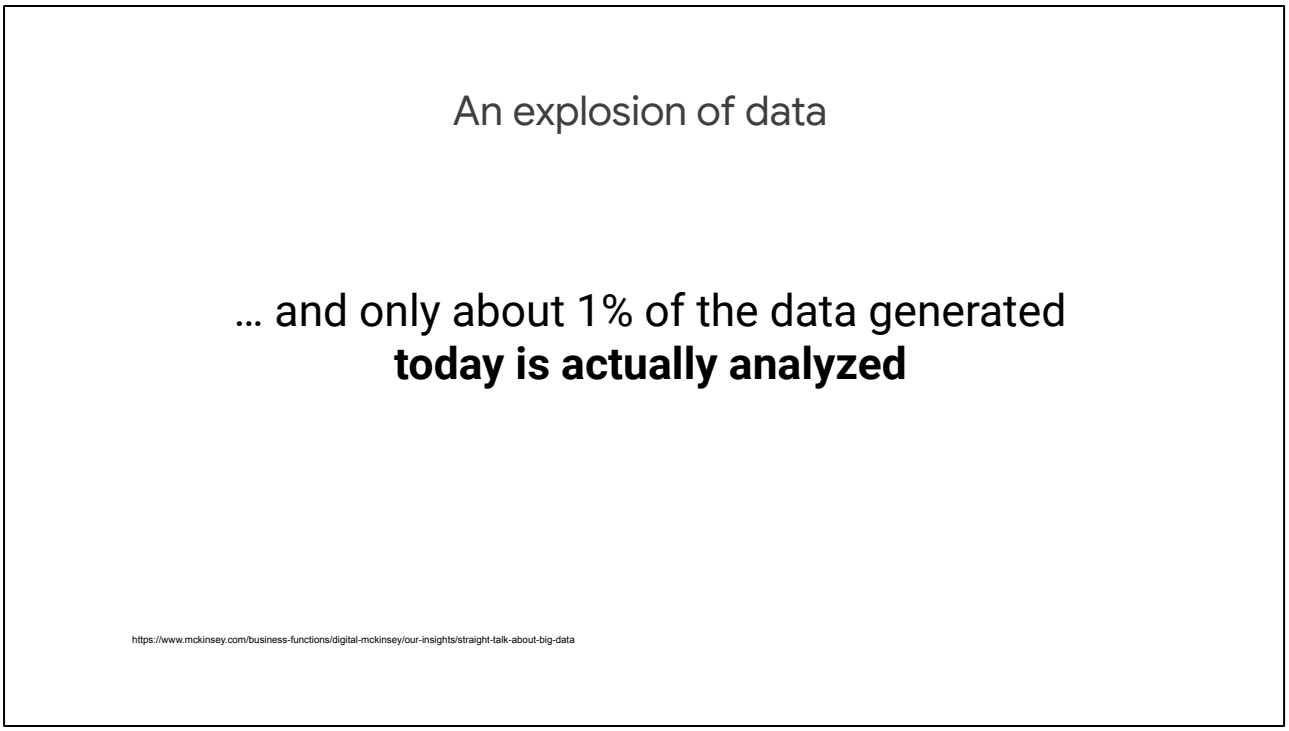
"By 2020, some 50 billion smart devices will be connected, along with additional billions of smart sensors, ensuring that the **global supply of data will continue to more than double every two years**"

https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/straight-talk-about-big-data

You're already taking this course, which means that you recognize the importance of big data processing. But why is this skill set in such high demand?

According to McKinsey research, by 2020 we'll have 50 billion devices connected in the Internet of Things.
These devices will cause the supply of data to double every two years.

https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/straight-talk-about-big-data

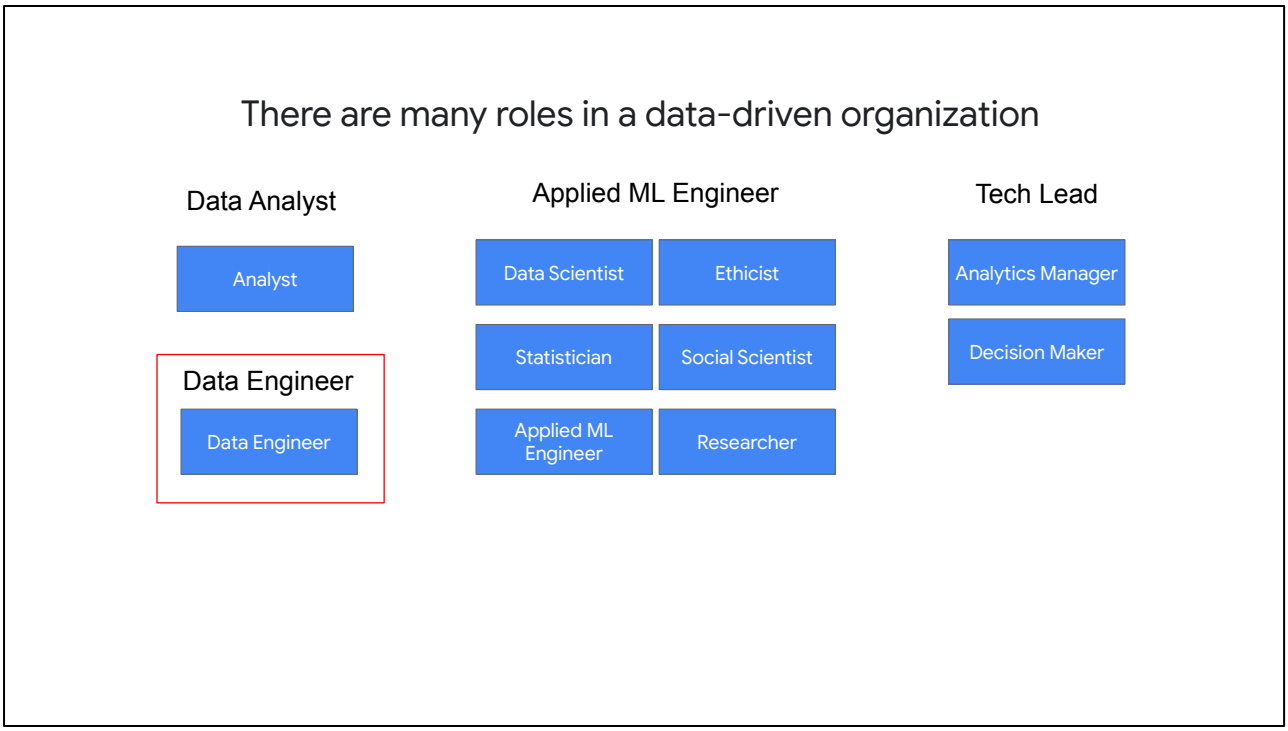


Unfortunately, though, only about 1% of the data generated today is actually analyzed, according to McKinsey.

This state of affairs provides a wide-open opportunity because there is a lot of value in data.

I believe that the ability to build applications that handle large amounts of data and derive insights from that data in an automated manner is a skill that will be well-rewarded in the marketplace. Individuals who have this skill will have many opportunities open to them, and companies that develop this skill will become more successful.

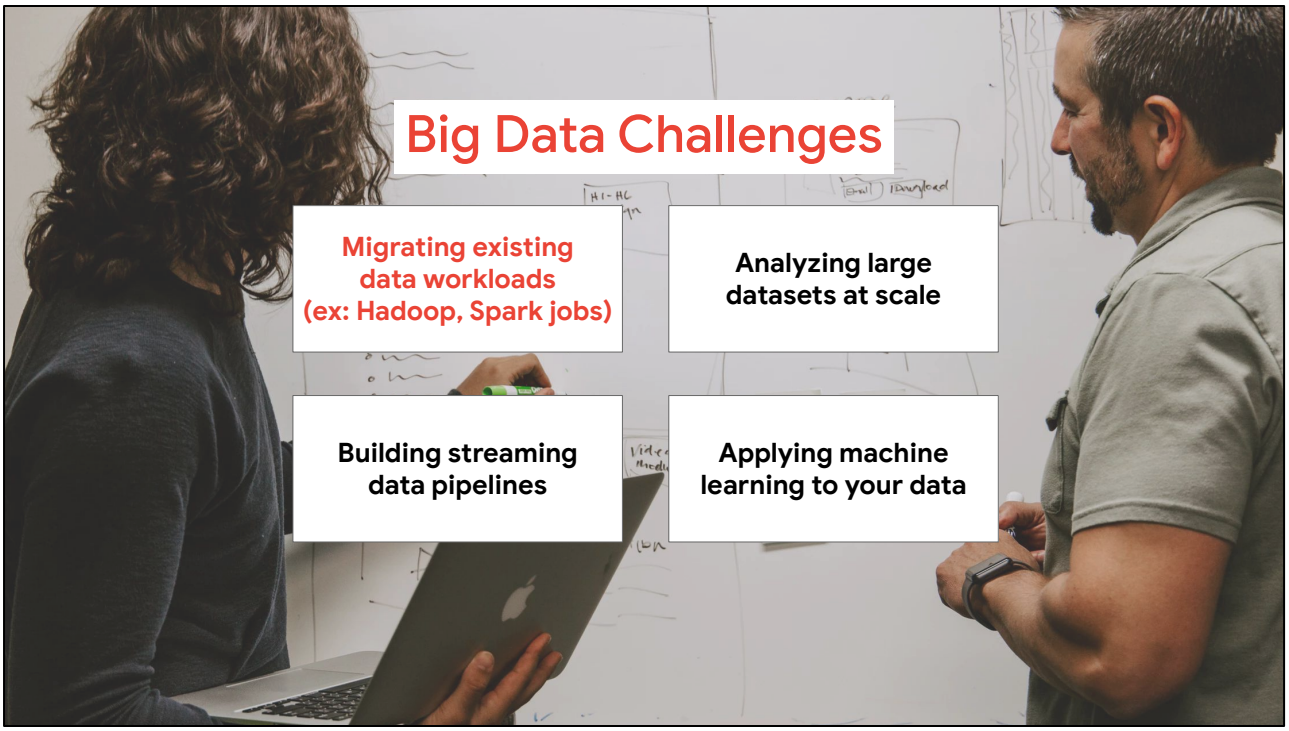
https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/straight-talk-about-big-data



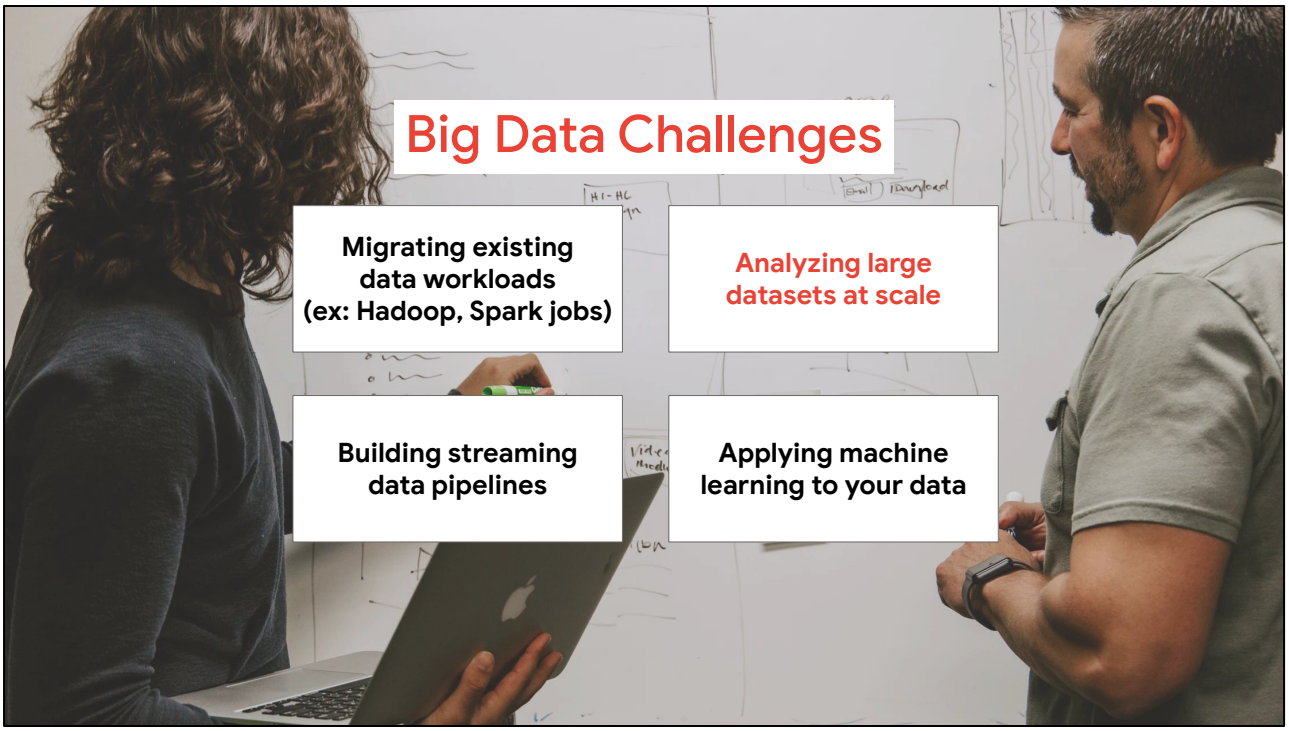
So, the opportunity for data analysts, data scientists, and data engineers—we'll talk about what these roles are, and what the differences are—the opportunity for all three of these roles is clear.

At its core, this course is primarily geared towards Data Engineers. That said, if you're an analyst, ML engineer, or tech lead for your team it's a valuable skill to know how all of the big data and ML products interact to solve some of the most common challenges that data engineers face.

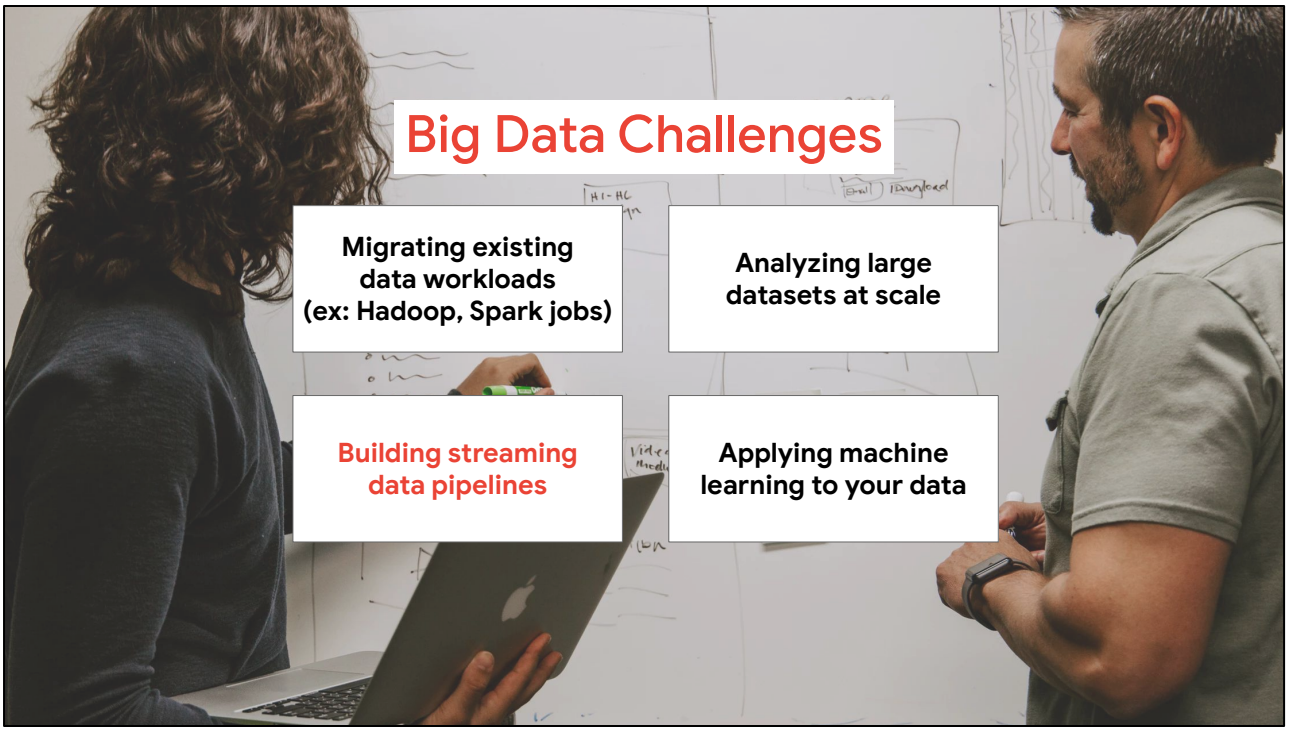
And those challenges are...



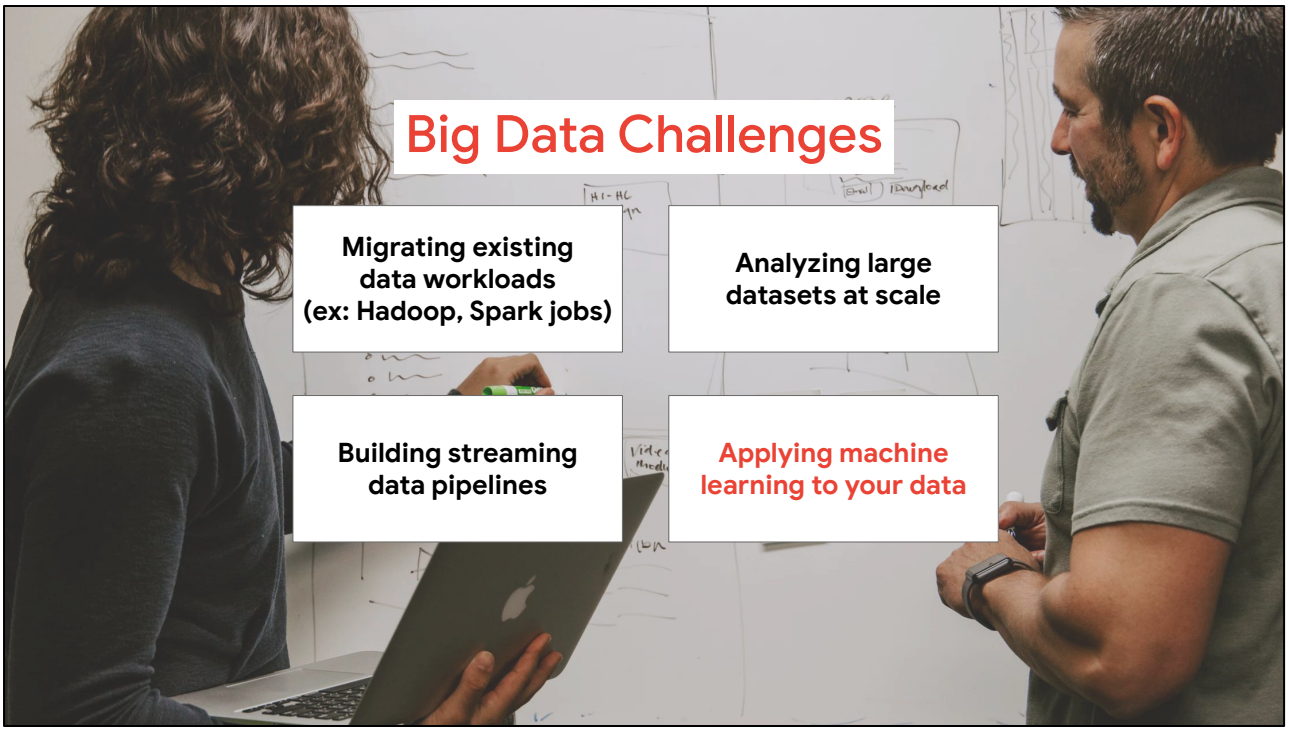
Migrating your existing big data workloads to an environment where you can effectively analyze your data.



Interactively analyzing large (and by that I mean terabytes to petabytes) datasets of historical data.



Building scalable pipelines that can handle streaming data, so that your business can make data-driven decisions more quickly.



And finally, building ML models so that you are not just reacting to data, you are able to make predictive, forward-looking actions using your data.