**Modernizing Data Lakes and Data Warehouses with Google Cloud.**

**Course series introduction.**

Damon: Hello, and welcome to the Data Engineering on Google Cloud course series.

I'm Damon, and I'm a technical curriculum developer at Google.

Together with my fellow instructors, we look forward to showing you how to design data processing systems, build end to end data pipelines, analyze data, and implement machine learning.

In addition to video lectures, you will also complete a series of hands-on labs.

As part of the data engineering learning path, we will first discuss the differences between data lakes and data warehouses, the two key components of any data pipeline.

This course highlights use cases for each type of storage, and dives into the available data lake and data warehouse solutions on Google Cloud in technical detail.

Also, this course describes the role of a data engineer, the benefits of a successful data pipeline to business operations, and examines why data engineering should be done in a Cloud environment.

Data pipelines typically fall under one of the extract load, extract load transform, or extract transform load paradigms.

So the next course, building batch data pipelines, describes which paradigm should be used and when for batch data.

Furthermore, it covers several technologies on Google Cloud for data transformation, including BigQuery, executing spark on data proc, pipeline graphs and data fusion, and serverless data processing with Dataflow.

Processing streaming data is becoming increasingly popular as streaming enables organizations to get real time metrics on operations.

So the third course covers how to build streaming data pipelines on Google Cloud.

Pub Sub is the primary product for handling incoming streaming data.

The course also covers how to apply aggregations and transformations to streaming data using Dataflow, and how to store or process records in BigQuery, or Big Table for analysis.

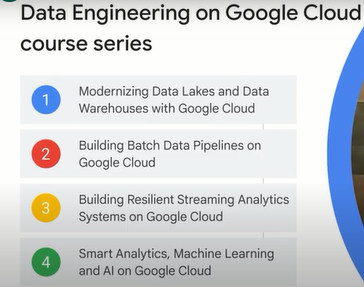
Incorporating machine learning into data pipelines increases the ability of organizations to extract insights from their data.

The final course covers several ways for machine learning to be included in data pipelines on Google Cloud depending on the level of customization required.

For little to no customization, the course covers Auto ML.

For more tailored machine learning capabilities, the course introduces notebooks and BigQuery machine learning.

Also, the final course covers how to productionize machine learning solutions using Kubeflow.



**Course introduction.**

Damon: Welcome to modernizing data lakes and data warehouses with Google Cloud, the first course of the data engineering learning path. We'll start off by describing the role of a data engineer.

We'll talk about a data engineer's clients and what the benefits of a successful data pipeline are for your organization.

Also, we will explain why data engineering should be done in a Cloud environment.

We'll concentrate on data lakes and data warehouses in this course, these are the two key components of any data pipeline.

We'll describe the differences between data lakes and data warehouses, and highlight use cases for each type of storage.

Also, we'll go into the available data lake and data warehouse solutions on Google Cloud in some technical detail.

Finally, you'll get hands-on experience with data lakes and data warehouses by using quick labs.

