

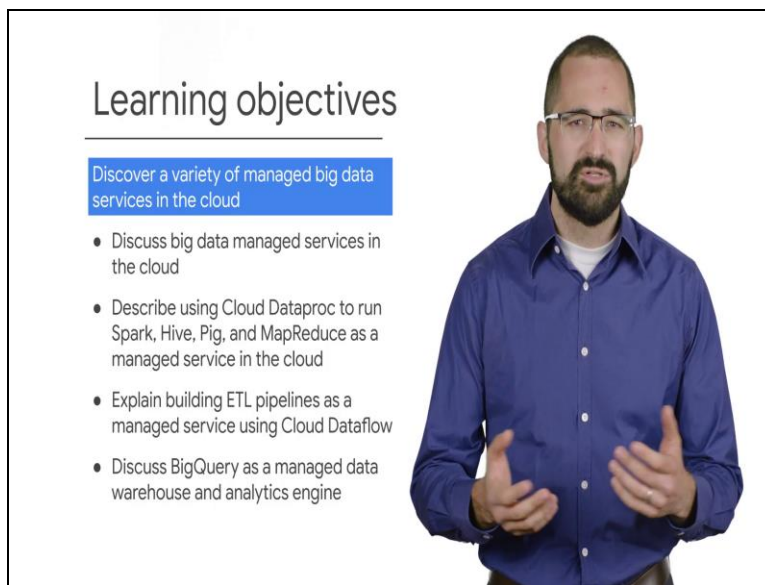
**Google Cloud Computing Foundation Course**  
**Evan Jones**  
**Technical Curriculum Developer**  
**Google Cloud**

**Lecture-69**  
**Module Introduction**

Hi welcome to; you have data but what are you going to do with it. I am Evan a technical curriculum developer here at Google and in this module you learn how you can gain insight through data using manage Big Data services. So, far this Google cloud computing foundation course you have discussed what cloud computing is the Google cloud platform and using GCP to build apps. You then explored storage options, the role of api's, cloud security, networking and the role GCP can play and automating the creation and management of your GCP resources.

In this module you look at some of the managed services that Google offers to process your Big Data.

**(Refer Slide Time: 00:39)**



**Learning objectives**

Discover a variety of managed big data services in the cloud

- Discuss big data managed services in the cloud
- Describe using Cloud Dataproc to run Spark, Hive, Pig, and MapReduce as a managed service in the cloud
- Explain building ETL pipelines as a managed service using Cloud Dataflow
- Discuss BigQuery as a managed data warehouse and analytics engine

The slide features a video thumbnail of Evan Jones, a man with a beard and glasses wearing a blue shirt, standing with his hands open as if speaking.

The objective of this module is for you to discover a variety of manage Big Data services in the cloud the specific learning objectives to achieve this include you being able to discuss Big Data managed services available in the cloud, describe the use of cloud data proc to run Spark, Hive Pig and MapReduce as a managed service. Explain the building of extract transform and load

pipelines using cloud dataflow. And discuss BigQuery as a managed data warehouse and analytics engine.

**(Refer Slide Time: 01:09)**

Agenda	
Introduction to Big Data Managed Services in the Cloud	Lab: Dataflow: Qwik Start - Templates
Leverage Big Data Operations with Cloud Dataproc	Lab: Dataflow: Qwik Start - Python
Lab: Dataproc: Qwik Start: Console	BigQuery, Google's Enterprise Data Warehouse
Lab: Dataproc: Qwik Start: Command Line	Quiz
Build Extract, Transform, and Load Pipelines using Cloud Dataflow	Summary

This agenda shows that topics that make up the module. The module starts with an introduction into Big Data managed services in the cloud before moving on to how big data operations can be leveraged through cloud data proc. You will then complete two labs will use the GCP console and then the G cloud command line tool to create a Cloud Data proc cluster and perform various tasks. After the labs you will explore the use of cloud dataflow to perform extract, transform and load operations.

The next two labs provide an opportunity to learn more about cloud dataflow. In the first you will create a streaming pipeline using a cloud dataflow template and in the second you will set up a Python development environment get the cloud dataflow SDK for Python and run an example pipeline using the GCP console. In the final topic you will learn about the role of BigQuery as a data warehouse. You will complete the module with the short quiz and a review of the main learning points of the module.