Scientific Data Processing

Advanced 6 Steps 6 hours 30 Credits

Big data, machine learning, and scientific data? It sounds like the perfect match. In this advanced-level quest, you will get hands-on practice with GCP services like Big Query, Dataproc, and Tensorflow by applying them to use cases that employ real-life, scientific data sets. By getting experience with tasks like earthquake data analysis and satellite image aggregation, Scientific Data Processing will expand your skill set in big data and machine learning so you can start tackling your own problems across a spectrum of scientific disciplines.

Infrastructure Data Business Transformation Machine Learning

Prerequisites:

This Quest requires hands-on experience with GCP data processing and machine learning services like Dataproc, Dataflow, and Cloud ML Engine. It is recommended that the student have at least earned a Badge by completing the hands-on labs in the Baseline: Data, ML, and AI Quest before beginning.

Quest Outline

Introduction to SQL for BigQuery and Cloud SQL

In this lab you will learn fundamental SQL clauses and will get hands on practice running structured queries on BigQuery and Cloud SQL.

1 hour | Introductory | 1 Credit

Rent-a-VM to Process Earthquake Data

In this lab you spin up a virtual machine, configure its security, access it remotely, and then carry out the steps of an ingest-transform-and-publish data pipeline manually. This lab is part of a series of labs on processing scientific data.

40 minutes | Introductory | 1 Credit

Weather Data in BigQuery

In this lab you analyze historical weather observations using BigQuery and use weather data in conjunction with other datasets. This lab is part of a series of labs on processing scientific data.

35 minutes | Fundamental | 5 Credits

Distributed Image Processing in Cloud Dataproc

In this lab, you will learn how to use Apache Spark on Cloud Dataproc to distribute a computationally intensive image processing task onto a cluster of machines.

1 hour | Advanced | 7 Credits

Analyzing Natality Data Using AI Platform and BigQuery

In this lab you analyze a large (137 million rows) natality dataset using Google BigQuery and Cloud Datalab. This lab is part of a series of labs on processing scientific data.

30 minutes | Advanced | 7 Credits

Predict Baby Weight with TensorFlow on AI Platform

In this lab you train, evaluate, and deploy a machine learning model to predict a baby's weight. You then send requests to the model to make online predictions. This lab is part of a series of labs on processing scientific data.

1 hour 30 minutes | Expert | 9 Credits

Quest Complete!

Congrats! You completed this quest and earned a badge. Become a cloud expert and start another.