Modernizing Data Lakes and Data Warehouses with Google Cloud

Introduction

Introduction to Data Engineering

- Module introduction
- The role of a data engineer
- Data engineering challenges
- Introduction to BigQuery
- Data lakes and data warehouses
- Transactional databases versus data warehouses
- Partner effectively with other data teams
- Manage data access and governance
- Demo: Finding PII in your dataset with the DLP API
- Build production-ready pipelines
- Google Cloud customer case study
- LAB : Using BigQuery to do Analysis

Building a Data Lake

- Module Introduction
- Introduction to data lakes
- Data storage and ETL options on Google Cloud
- Build a data lake using Cloud Storage
- Secure Cloud Storage
- Store all sorts of data types
- Cloud SQL as a relational data lake
- LAB : Loading Taxi Data into Google Cloud SQL

Building a Data Warehouse

- Module Introduction
- The modern data warehouse
- Introduction to BigQuery
- Demo: Querying TB of data in seconds
- Get started with BigQuery
- Load data into BigQuery
- LAB : Loading data into BigQuery
- Explore schemas
- Demo: Exploring Schemas
- Schema design
- Nested and repeated fields
- Demo: Nested and repeated fields
- Design the optimal schema for BigQuery
- LAB : Working with JSON and Array data in BigQuery
- Optimize with partitioning and clustering