Google Cloud

Managing a Data Mesh with Dataplex

Dataplex is an intelligent data fabric that enables organizations to centrally discover, manage, monitor, and govern their data across data lakes, data warehouses, and data marts. You can use Dataplex to build a data mesh architecture to decentralize data ownership among domain data owners.

In this course, you will learn how to discover, manage, monitor, and govern your data across data lakes, data warehouses, and data marts through guided lectures and independent exercises using sample data.

DURATION
2 days

LEVEL
Intermediate

FORMAT
Instructor led

What you'll learn

- Identify the importance of a modern data platform
- Configure and set up Dataplex
- Secure data lakes, zones, and assets
- Implement tagging for resources and use tags to search for assets
- Process data using Dataplex tasks
- Design, execute and report on data quality processes

Overview	7 modules · 6 labs
Who this course is for	Customers
Products	• Dataplex • Cloud Storage • BigQuery • Dataflow
Prerequisite	Completion of the "Modernizing Data Lakes and Data Warehouses with Google Cloud" and "Building Batch Data Pipelines on Google Cloud" courses in the "Data Engineer" learning path or equivalent experience using Google Cloud.
Not covered	This course does not cover the interaction of Dataplex with Dataproc Metastore nor does it do a deep dive into BigLake concepts.

Module 01   Introduction to Dataplex

Topics
- Modern Data Platforms and Data-Oriented Design
- Pillars of Data Governance
- What is Dataplex?
- Dataplex Capabilities
- Dataplex compared with other products on Google Cloud

Objectives
- Identify the importance of a modern data platform
- Explain the role of Dataplex on Google Cloud

Module 02   Creating a Data Mesh on Dataplex

Topics
- What is a data mesh?
- Dataplex concepts
- Creating data lakes and zones
- Assets in Dataplex

Objectives
- Define key Dataplex concepts
- Configure and set up Dataplex

Activities
Lab: Provision a Data Mesh using Dataplex

Module 03   **Processing Data on Dataplex**

Topics
- Processing data on Dataplex
- Data preparation tasks
- Ingestion jobs
- Dataflow and Spark tasks

Objectives
- Understand different data processing options in Dataplex
- Configure and run data preparation tasks on Dataplex

Activities   Lab: Standardize Data using Dataplex Tasks

Module 04   **Managing Data Security through Dataplex**

Topics
- IAM permissions and roles
- Securing your data lake
- Policy management
- Metadata security

Objectives   Secure data lakes, zones, and assets in Dataplex

Activities   Lab: Manage Data Security using Dataplex

Module 05   **Data Tagging and Data Catalog**

Topics
- Introduction to Data Catalog
- Technical metadata vs. business metadata
- Tags and tag templates
- Entries and entry groups
- Data lineage

Objectives   Implement tagging for resources and use tags to search for assets

Activities   Lab: Data Catalog and Data Lineage

Module 06   **Data Quality and Profiling**

Topics
- Data quality tasks and AutoDQ
- Reporting on data quality
- Data profiling

Objectives	Design, execute and report on data quality processes
Activities	Lab: Data Quality and Profiling your Data in BigQuery

---

Module 07   Dataplex Best Practices

Topics	• Best practices
	• End-to-end demo
Objectives	Implement best practices for Dataplex
Activities	Challenge Lab: Managing a Data Mesh with Dataplex