

Managing a Data Mesh with Knowledge Catalog

Knowledge Catalog 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 Knowledge Catalog 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**

200 - Intermediate

**FORMAT**

Instructor-led

What you'll learn

- Secure data lakes, zones, and assets
- Implement tagging for resources and use tags to search for assets
- Process data using Knowledge Catalog tasks
- Design, execute and report on data quality processes

Overview	7 modules · 6 labs
Who this course is for	• Customers
Products	• Knowledge Catalog • BigQuery Studio • Dataflow Prime • Storage Insights • Lakehouse • Managed Service for Apache Spark
Prerequisites	• 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	• The interaction of Knowledge Catalog with Dataproc Metastore • Deep dive into BigLake concepts

Module 01

Introduction to Knowledge Catalog

Topics	• Modern Data Platforms and Data-Oriented Design • Pillars of Data Governance • What is Knowledge Catalog? • Knowledge Catalog Capabilities • Knowledge Catalog compared with other products on Google Cloud
Objectives	• Identify the importance of a modern data platform • Explain the role of Knowledge Catalog on Google Cloud

Module 02

Creating a Data Mesh on Knowledge Catalog

Topics	• What is a data mesh? • Knowledge Catalog concepts • Creating data lakes and zones • Assets in Knowledge Catalog
Objectives	• Define key Knowledge Catalog concepts • Configure and set up Knowledge Catalog
Activities	• Lab: Provision a Data Mesh using Knowledge Catalog

Module 03

Processing Data on Knowledge Catalog

Topics	• Processing data on Knowledge Catalog • Data preparation tasks • Ingestion jobs • Dataflow and Spark tasks
Objectives	• Understand different data processing options in Knowledge Catalog • Configure and run data preparation tasks on Knowledge Catalog

Module 04

Managing Data Security through Knowledge Catalog

Topics	• IAM permissions and roles • Securing your data lake • Policy management • Metadata security
Objectives	• Secure data lakes, zones, and assets in Knowledge Catalog

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

Knowledge Catalog Best Practices

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