
Build a Data Mesh with Dataplex: Challenge Lab

Task 1: Create Dataplex Lake with Zones and Assets

◆ **Step 1: Create Lake**

1. Go to **Navigation menu > Dataplex > Lakes**
2. Click **Create Lake**
 - a. **Lake ID:** sales-lake
 - b. **Region:** us-central1
 - c. Click **Create**

◆ **Step 2: Create Zones**

- After creating the lake, go to **sales-lake > Zones > Create Zone**

Zone 1

- ID: raw-customer-zone
- Type: Raw
- Region: us-central1
- Click **Create**

Zone 2

- ID: curated-customer-zone
- Type: Curated
- Region: us-central1
- Click **Create**

◆ **Step 3: Add Assets**

Asset for Raw Zone:

- Go to **sales-lake > raw-customer-zone > Assets > Create Asset**
 - **Asset ID:** customer-engagements
 - **Type:** Cloud Storage

- **Storage path:** gs://qwiklabs-gcp-02-e257d9be8762-customer-online-sessions
- Click **Create**

Asset for Curated Zone:

- Go to sales-lake > curated-customer-zone > Assets > Create Asset
 - **Asset ID:** customer-orders
 - **Type:** BigQuery
 - **BigQuery path:** qwiklabs-gcp-02-e257d9be8762.customer_orders
 - Click **Create**

The screenshot shows two main sections of the Google Cloud Platform interface.

Top Section (Sales Lake Overview):

- Lakes** tab selected.
- CREATE** button.
- LEARN** button.
- Filter** button and text input: "Filter instances".
- Table Headers:** Display name ↑, Assets, Status, Resources requiring action, Region, Last modified, Labels.
- Table Data:**| SALES LAKE | 2 | Active | 0 | us-central1 (Iowa) | June 23, 2025 | None | ⋮ |

Bottom Section (Curated Customer Zone Details):

- Sales Lake** page title.
- EDIT** and **DELETE** buttons.
- LEARN** button.
- Zones requiring action:** 0.
- Assets requiring action:** 0.
- Details:**

Lake ID	sales-lake
Display name	Sales Lake
Status	Lake active
- Actions:**
 - ZONES** (selected), **DETAILS**, **PERMISSIONS**, **ACTIONS**.
 - + ADD ZONE** and **DELETE ZONE** buttons.
- Filter** button and text input: "Filter instances".
- Table Headers:** □, Display name ↑, Type, Status, Assets requiring action, Assets, Data locations, Last modified, Labels.
- Table Data:**| □ | CURATED CUSTOMER ZONE | Curated Zone | Active | 1 | Regional (us-central1) | June 23, 2025 | None | ⋮ |
| □ | RAW CUSTOMER ZONE | Raw Zone | Active | 1 | Regional (us-central1) | June 23, 2025 | None | ⋮ |

Task 2: Create & Apply Tag Template

◆ Step 1: Create Tag Template

1. Go to **Navigation > Data Catalog > Tag Templates**
2. Click **Create Tag Template**
 - a. **ID:** protected-customer-data-template
 - b. Check **Make public**
 - c. Add 2 fields (Enum):
 - i. Field 1: raw-data-flag, values: Yes, No
 - ii. Field 2: protected-contact-information-flag, values: Yes, No
 - d. Click **Create**
- ◆ **Step 2: Apply Tag Template**
 1. Go to **Dataplex > Lakes > sales-lake > raw-customer-zone**
 2. Click **Edit Tags**
 3. Choose **Tag Template:** protected-customer-data-template
 4. Set both fields to Yes
 5. Click **Save**

Task 3: Assign IAM Role to Another User

- ◆ **Step 1: IAM Permission**
 1. Go to **Navigation > IAM & Admin > IAM**
 2. Click **Grant Access**
 - a. **New Principal:** student-02-7d1fd88eacd3@qwiklabs.net
 - b. **Role:** Dataplex Metadata Writer (least privilege to upload files to asset)
 - c. Click **Save**

Attach Tags

⚠ Tags are deprecated. Consider using Aspects to take advantage of the latest functionalities.

1 Choose the tag templates

Selected tag templates Protected Customer Data Template

2 Fill in tag values

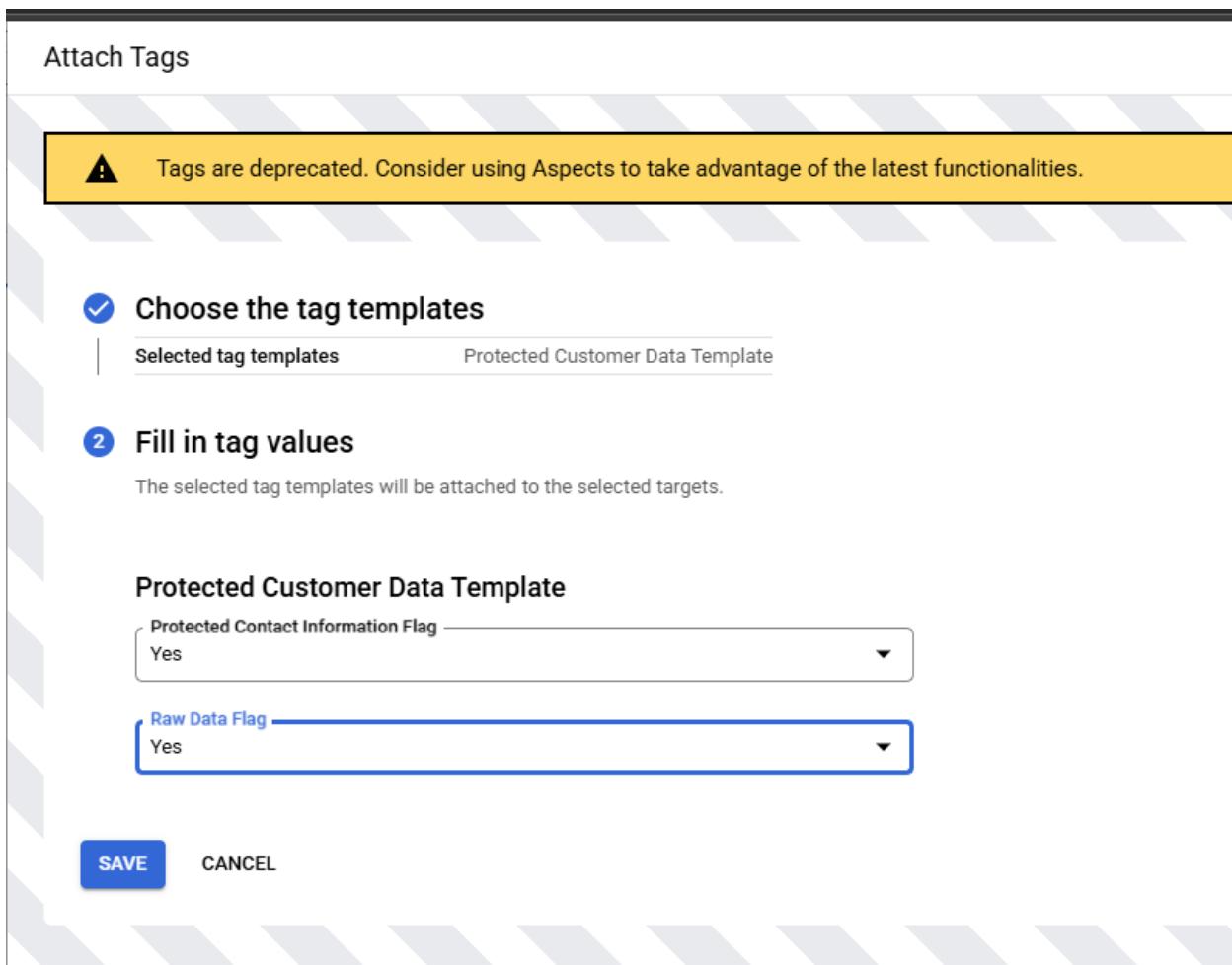
The selected tag templates will be attached to the selected targets.

Protected Customer Data Template

Protected Contact Information Flag — Yes

Raw Data Flag — Yes

SAVE CANCEL



Task 4. Create and upload a data quality specification file to Cloud Storage

The Cloud Storage bucket for step 2 has been pre-created in this lab.

1. Create a data quality specification file named **dq-customer-orders.yaml** with the following specifications:
 - **NOT NULL** rule applied to the **user_id** column of the **customer_orders.ordered_items** table
 - **NOT NULL** rule applied to the **order_id** column of the **customer_orders.ordered_items** table
2. Upload the file to the Cloud Storage bucket named **Project ID-dq-config**.

Task 5. Define and run a data quality job in Dataplex

The BigQuery dataset for step 1 has been pre-created in this lab.

- Define a data quality job using the **dq-customer-orders.yaml** file with the following specifications:

Property	Value
Data Quality Job Name	Customer Orders Data Quality Job
BigQuery destination table for the results	_____.orders_dq_dataset.results
User service account	Compute Engine default service account

- Run the data quality job immediately.

The screenshot shows the Google Cloud Data Catalog interface. On the left, there's a sidebar with navigation links like Discover, Manage Metadata, Manage Lakes, Govern, Process, and Settings. The main area shows a 'Raw Customer' dataset under 'Discover'. A message at the top states 'Data Catalog is deprecated' with a note for administrators about transitioning to the Universal Catalog. Below this, a 'Dataplex zone details' section shows a zone named 'Raw Customer Zone' with location 'us-central1' and resource URL 'raw-customer-zone'. A success message 'Tag successfully attached' is displayed over the zone details. Further down, there's a 'Create task' section with options like 'Convert to Curated Formats', 'Tier from BQ to GCS Assets', and 'Check Data Quality [Legacy offering]'. At the bottom, there are sections for 'Define an ingestion job' with options 'Ingest to Dataplex' and 'Ingest JDBC to Dataplex', and a 'LEARN MORE' button.

Dataplex Universal Catalog / Create data quality task

Create task

- Define task

Provide parameters for this task
- Set schedule (optional)

Run this task once or on a schedule

Check data quality

Dataplex lake * sales-lake

Display name Customer Orders Data Quality Job

ID * customer-orders-data-quality-job

Description

Optional 0 / 1024

Task configuration

Data quality specification

Specify a Cloud Storage URI to a Data Quality specification file e.g. gs://bucket-name/my/path/to/config.yaml. Supported paths include .yaml, .yml, .zip.

CREATE CANCEL

Dataplex Universal Catalog / Create data quality task

Create task

Dataplex now offers a new, improved Data Quality feature with rich UI, API, rule recommendations and more. It is highly recommended for users to start using Dataplex AutoDQ [Learn more](#).

Before creating a Dataplex Data Quality task, please make sure to follow the "Before you begin" instructions [here](#).

Define task

Provide parameters for this task

Set schedule

Run Once Repeat

Start Immediately

Jobs scheduled to run once at 5:13 PM.

CREATE

← 4f3e0f89-f63d...

CLONE CANCEL DELETE VIEW LOGS REFRESH VIEW SPARK HISTORY SERVER

Subminor runtime version specified. It is recommended to not select a specific subminor runtime version in production, as old subminor runtime versions are deprecated after 365 days of their release and become not available.

Batch ID	4f3e0f89-f63d-4898-b5ad-82689874a162-0
Batch UUID	a71a731e-60d0-4c08-a14f-ad0ceb94cd64
Resource type	Batch
Status	✓ Succeeded