Day 1 – Agenda

- Oracle Cloud Infrastructure Data Integration – Introduction
- OCI Data Integration product features
- Use cases
- Pricing
- Modern Data Challenges
- Why Oracle Cloud?
- Key Capabilities
- Data Integration Users
- Getting Started with Data Integration
- Service Flow
- Data Integration Concepts
- OCI-DI vs ODI-MP
- Workspace
- Connecting to Data Sources
- Projects
- Folders

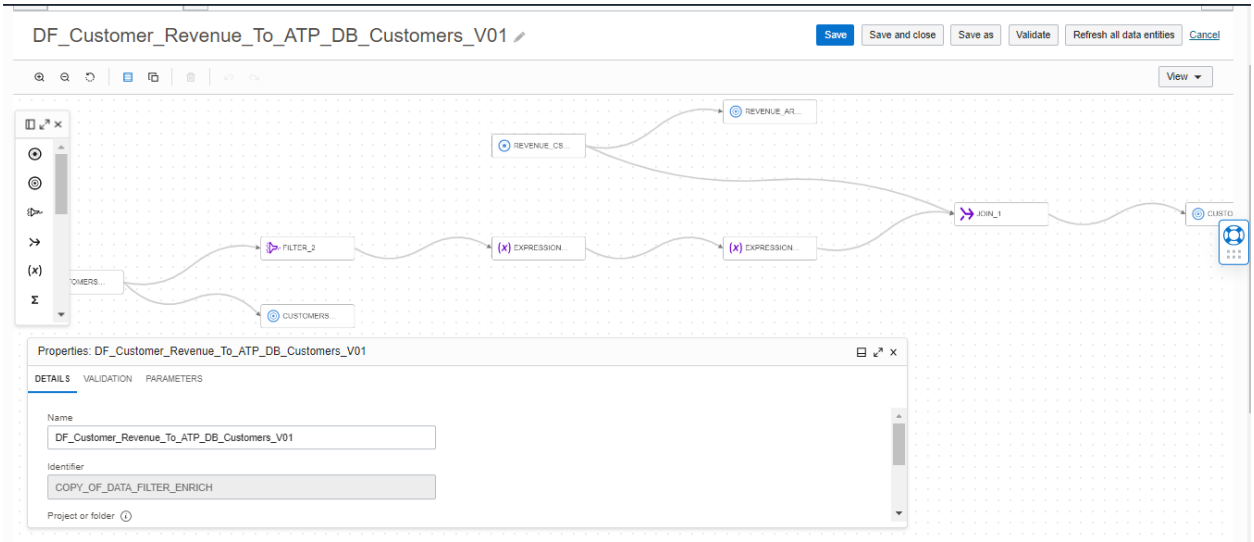
Oracle Cloud Infrastructure
Data Integration- Introduction

Easily extract, transform, and load (ETL) data for data science and analytics. Design code-free data flows into data lakes and data marts. Part of Oracle’s comprehensive portfolio of integration solutions.

OCI Data Integration product features

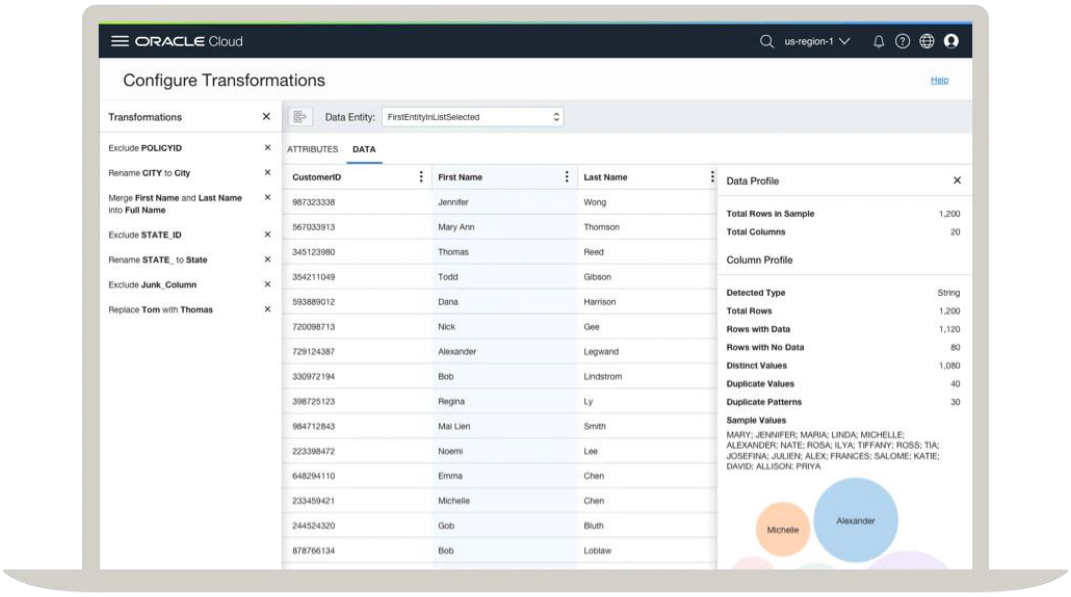
No-code data flow design

- Simplified processes
- Powerful transformations
- Flexible data integration



Data-immersive experience

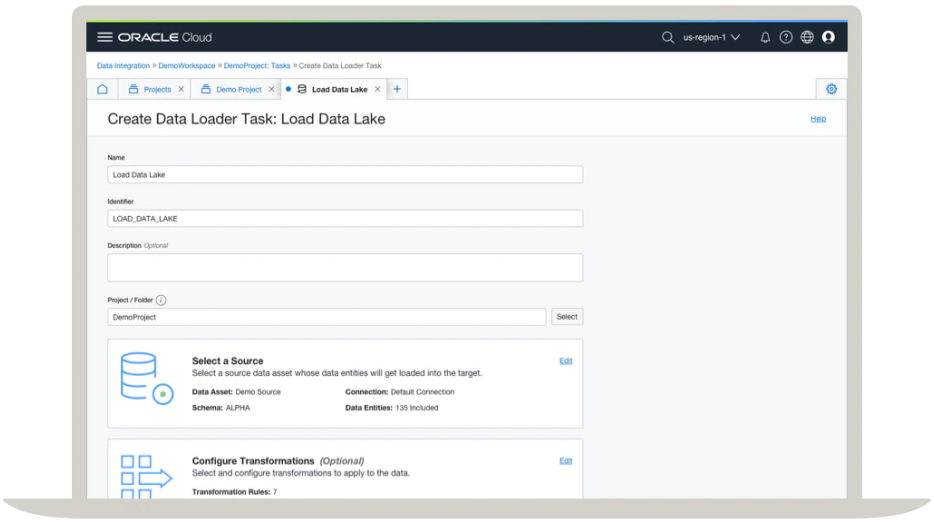
- Visually preview data
- Modify integrations



Automated schema drift protection

- Rule-based integration flow

- Pattern-based data mapping



Hybrid data execution

- Spark ETL
- E-LT processing
- Built-in optimizer

Optimized for Oracle Cloud

- Consistent performance
- Scalable

Pay As You Go pricing

- Reduce your capital expenditure
- Save energy and money

Use cases

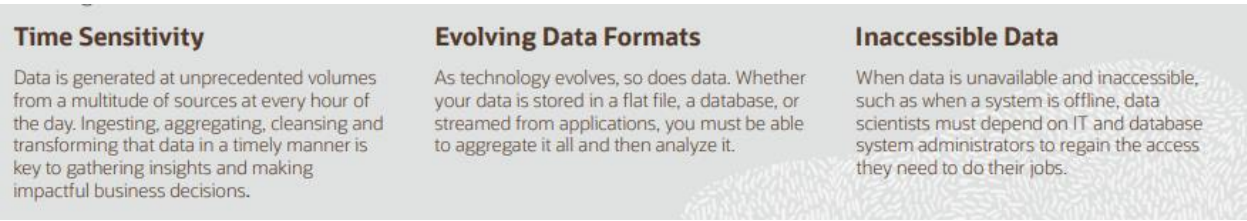
- Data integration for big data, data lakes, and data science.
- Ingest data faster and more reliably into data lakes for data science and analytics.
- Create high-quality models more quickly.
- Data integration for data marts and data warehousing
- Load and transform transactional data at scale. Create an organized view from large data volumes.

Pricing

Product	Unit Price	Metric
Oracle Cloud Infrastructure Data Integration: Workspace	$0.16	Workspace Usage Per Hour—UCM
Oracle Cloud Infrastructure Data Integration	$0.04	Gigabyte of Data Processed Per Hour—UCM
Pipeline Operator Execution for the Pipeline (orchestration) execution	$0.30	Annual Flex/Pay As You Go: Priced at $0.30 per Hour. The first 30 hours of Execution Hour per tenant per month is free.

Modern Data Challenges

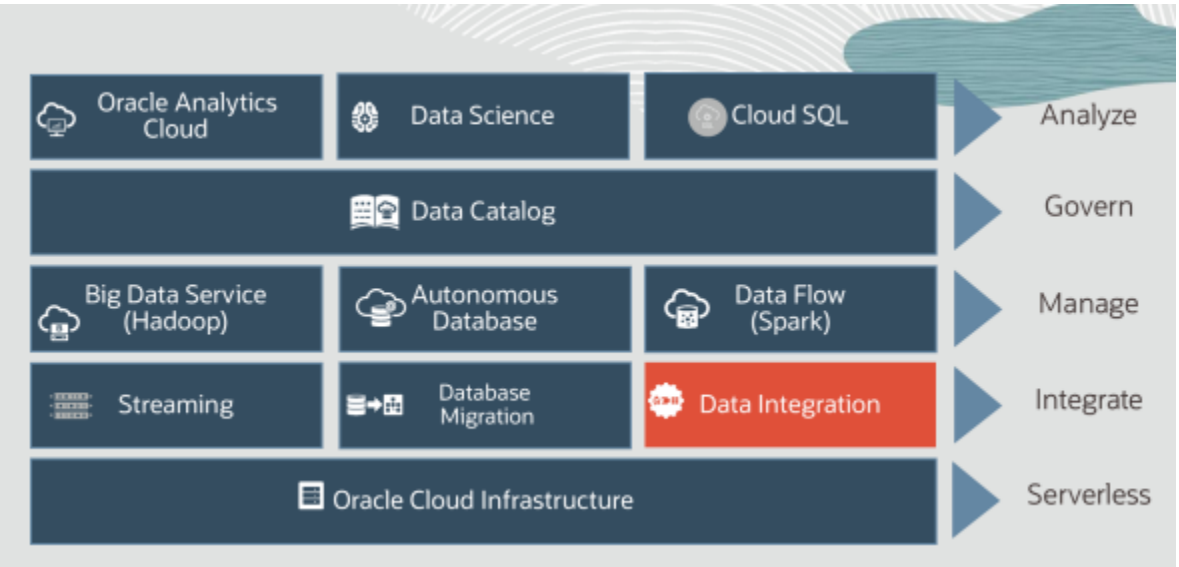
Can your data-driven business keep up with the vast amount of data generated on a daily basis? In order to maintain a competitive edge, you must be able to consolidate and rationalize data to take advantage of modern analytics and data science, and overcome the following challenges:



Evolve and adapt with Oracle Cloud. Oracle Cloud Infrastructure Data Integration is a fully-managed, serverless cloud service built on Oracle's 2nd generation cloud infrastructure, offering speed, reliability, and security. The convergence of Cloud, Big Data, Analytics, and Data Science in Oracle Cloud Infrastructure creates new and exciting opportunities for you, such as elastic compute resources and infinite, modern storage options – all designed to withstand the demands of today's enterprise operations.

Why Oracle Cloud?

A modern data platform for all your needs. Oracle Cloud Infrastructure enables you to take advantage of Oracle's fully-managed services, including ability to deploy in all regions, easily scale compute instances in and out, and automatically upgrade and patch your services to meet your organization's immediate needs. Spend more time managing your data and less time managing your systems.



Key Capabilities:

Broad Connectivity

Oracle's modern data storage solutions offer scalability, flexibility, and security. Your options include Autonomous Databases, such as Autonomous Transaction Processing and Autonomous Data Warehouse, Oracle Database, Oracle Object Storage on Oracle Cloud Infrastructure.

Powerful Data Flow

Design your custom data flows using an interactive, easy-to-use visual designer with auto-mapping operators, such as Filter, Join, Aggregate, Expression, and Sequence, to name a few. With Parameterization, you can create flexible, reusable Data Flows so you can spend less time designing similar data flows and more time analyzing and learning from your data.

Data Lake Loader

Oracle Cloud Infrastructure Data Integration is capable of ingesting a vast amount of data into a Data Lake, while performing data preparation and transformation operations on that data. Creating and running a Data Loader Task is as easy as 1, 2, 3:

1. Select your source

2. Configure transformations (optional)

3. Select your target Spend more time extracting value out of your data and less time preparing and transforming it.

Hybrid Execution at Scale

Oracle Cloud Infrastructure Data Integration uses a serverless Spark-based engine to process transformations at scale and can push-down transformations to sources and/or targets depending on the workload. It can intelligently determine which engine to use at run-time to perform transformations in the most efficient manner.

Data Integration Users

Administrators, data engineers, ETL developers, and operators are among the different types of data professionals who use Oracle Cloud Infrastructure Data Integration.

Administrators:

Manage, and monitor lifecycle management. Apply security policies for accessibility.

Data engineers and ETL developers:

Develop, build, and test data integration solutions.

Operators:

Manage, monitor, and diagnose data integration executions.

Getting Started with Data Integration

Before you create a Data Integration workspace, review the prerequisites and list of tasks.

You must have the following resources and minimum policies in your tenancy. If you don't have the proper rights, have your administrator create them for you.

Before you start setting up the Data Integration service for use, you must have:

- An Oracle Cloud Infrastructure account with administrator privileges
- Access to the Data Integration service

List of Tasks

Create Oracle Cloud Infrastructure resources for your Data Integration activities

In Oracle Cloud Infrastructure Identity and Access Management (IAM), create your compartments, users, and groups of users.

To create resources for your Data Integration activities:

- Create a compartment in your tenancy for Data Integration activities.
- If your data sources will be in a private network, create a VCN with at least one subnet in the compartment.
- Create a group for users in charge of workspaces, and then add users to the group.
- Take note of the group name. You then create policies for the group.

Configure networking components for your data sources

In Oracle Cloud Infrastructure Networking, you can set up virtual cloud networks (VCNs) and subnets. Only regional subnets are supported, and DNS hostnames must be used in the subnets. Depending on the location of your data sources, you might have to create other network objects such as service gateways, network security groups, and Network Address Translation (NAT) gateways.

Create policies to access and use Data Integration

In Oracle Cloud Infrastructure Identity and Access Management (IAM), create the required policies that give groups of user’s proper access to Data Integration resources.

Data Integration must also have permission to manage the virtual networks and subnets that you set up for integration.

To control non-administrator user access to Data Integration resources and functions, you create groups in Oracle Cloud Infrastructure Identity and Access Management (IAM). Then you write IAM policies that give the groups proper access.

You can use Data Integration policy templates in the IAM Policy Builder to create a policy, or you can manually enter the policy statements in the manual editor.

You can create most of the Data Integration policies at the tenancy level or at the compartment level.

Policies

Workspace Policy for Creation

This policy gives permission to a group to create Data Integration workspaces.

allow group <*group-name*> to manage dis-workspaces in compartment <*compartment-name*>

Service Flow

- The administrator applies the necessary policies and satisfies the necessary connectivity requirements. After this set-up Data Integration service can establish a connection to the data sources.

- Workspaces are then created by the administrator and provides access to them. Multiple workspaces can be created and organized based on requirements.

- Developers then create data assets to connect source and target data sources.

- To create an integration task, start with a data flow. The designer in Data Integration is an easy-to-use graphical user interface where you can select from different operators and visually build the data flow. It includes validation and debug features to help you identify and correct potential issues before running the task.

- When you create a data loader task, you specify your source data asset, and then configure transformations to cleanse and process the data as it is loaded into the target data asset.

- To execute a specific set of processes in a sequence, you create a pipeline. Designing a pipeline is similar to building a data flow, where you use operators to add the tasks and activities you want. After building a pipeline, you create a pipeline task that uses the pipeline.

- After you create tasks, you publish them to the Default Application or to your own Application. From the Application, you can run tasks and then monitor their progress and status. You can also schedule tasks for automated runs.

Data Integration Concepts

Workspace
: The container for all Data Integration resources, such as projects, folders, data assets, tasks, data flows, pipelines, applications, and schedules, associated with a data integration solution.

Project
: A container for design-time resources, such as tasks or data flows and pipelines.

Folder
: A container within a project or another folder to organize your design-time resources.

Data Asset
: Represents a data source such as a database, an object store, a file or document store containing the data source's metadata and connection details.

Connection
: Includes the necessary details to establish a connection to a data source. A connection is always associated to one data asset. A data asset can have more than one connection.

Data Entity
: A collection of data, such as a database table or view, or a single logical file, with many attributes that describe its data.

Schema
: A collection of data entities within a data asset.

Data Flow
: A design-time resource that defines the flow of data and any operations on the data between the source and target systems. To run a data flow, you add the data flow to an integration task.

Pipeline
: A design-time resource for orchestrating tasks and activities in a sequence or in parallel to facilitate a process from start to finish. To run a pipeline, you add the pipeline to a pipeline task.

Operator
: An operator represents an input source or output target, or a transformation in a data flow. In a pipeline, an operator represents a published task or an activity such as merge or end.

Parameter
: A type of variable you can assign to an operator's details so that you can reuse the data flow or pipeline design with different resources and values. When you use parameters and set default values during design time, you can then change the values later, either in tasks that wrap the data flow or pipeline, or when you run the tasks.

Task
: A design-time resource that specifies a set of actions to perform on data. You can create data loader tasks, integration tasks for data flows, and pipeline tasks for pipelines. You can also create SQL tasks and OCI Data Flow tasks. To run a task, you publish the task into an Application to test it or roll it out to production.

Application

A container for runtime artifacts, such as tasks that have been published along with their dependencies. You use Applications for testing and eventually roll them out into production.

Patch

An update to an Application. When you publish a single task or a group of tasks, or when you unpublish a task, these activities are logged as patches in an Application.

Run

A runtime artifact that represents the execution of a task.

Schedule

A runtime resource that defines when and how often any published tasks should run automatically.

Task Schedule

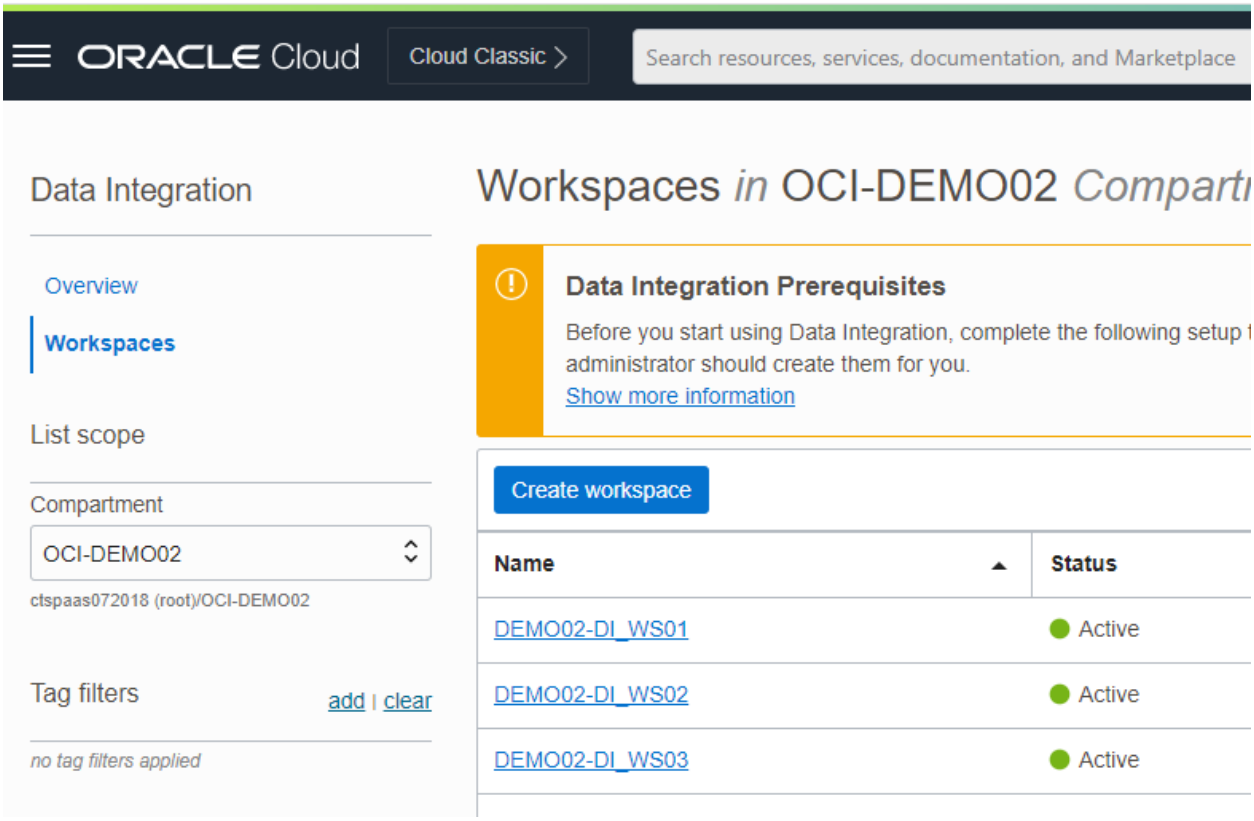
A runtime resource that is associated with a specific published task and an existing schedule to define when and how often the task should run automatically.

OCI-DI vs ODI-MP

OCI-DI	ODI-MP
ETL/ELT	ELT
Fully cloud based serverless architecture	OCI Compute Instance
Workspace	Repository
Data Asset	Data Server
Project, Folders	Project, Folders
Data Flow, Integration Task	Mapping
Publish	Regenerate/Generate
Pipeline	Package/Loadplan
Monitor Task Runs	Operator

Workspace

1. Login to cloud console. Go to Analytics & AI.
2. Select Data Integration.
3. In the next page click Workpaces.
4. Select your compartment, then select the workspace.



Workspace Home Page

The workspace home page consists of three tiles:

Recents:

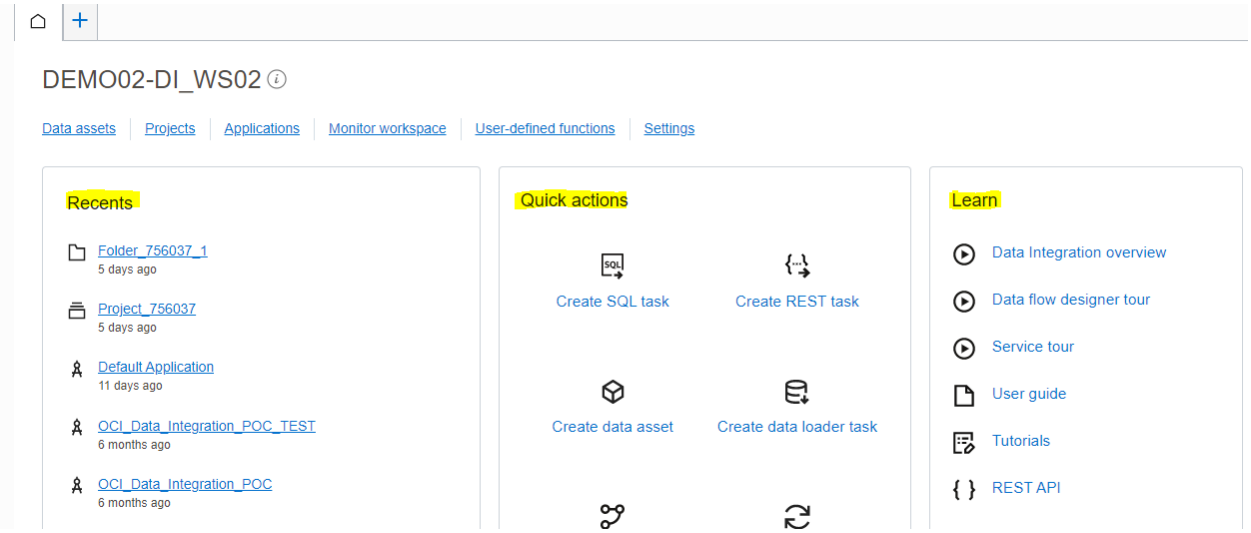
Displays the 10 most recent Data Integration resources you updated or created in the workspace. Click a resource name to quickly access recent objects such as data assets, data flows, pipelines, design tasks, applications, projects, and folders. A resource with a solid star icon means you have marked the resource as a favorite.

Quick Actions:

Displays links to different resources you can create to start planning and designing your data integration solution.

Learn:

Displays a list of resources that you might find helpful while you work with Data Integration



Accessing Your Data Integration Resources

Use the Open tab (plus icon) in the tab bar to access data assets, projects, user defined functions, and Applications in your workspace. You can also monitor metrics in the workspace. Each resource page opens as a new tab in the tab bar.

You can also use the navigation links below the tab bar to access your workspace resources.

Managing Tabs

During your Data Integration session, you might find yourself with several tabs open.
Click Manage tabs (gear icon) to search open tabs, navigate to other open tabs, or close tabs.

View Workspace Settings

The Workspace Settings page displays general workspace details, and network resource information, if available.

To view your workspace settings:

- From the workspace home page, select Settings on the Quick Actions tile.
- Alternatively, select Settings on the navigation links below the tab bar.

Mark Resources as Favorite

Select the resource and use the star icon in the favorite column.

The below resources can be set as favorites:

- Applications
- Data assets
- Data Flows
- Folders
- Projects
- Pipelines
- Tasks

Service Limits and Quotas

Service Limits

Five workspaces per region.

Compartment Quotas

You can limit the number of workspace resources in a compartment by creating a quota limit

Connecting to Data Sources

In Oracle Cloud Infrastructure Data Integration, data assets represent data sources that you can use as sources and targets for your integration tasks. Data assets can be databases, flat files, and so on.

You create and access data assets from the Data Assets page. From the workspace home page, you can access the Data Assets page by clicking Data Assets or by clicking the Open tab (plus icon) in the tab bar and then selecting Data Assets.

Creating a Data Asset

When you create a data asset, you select the data source type and then provide the data source details and the corresponding default connection details to that source. You can use the Console or the APIs to create data assets.

For most data assets, instead of providing the actual password for a connection, you can choose to use a secret in Oracle Cloud Infrastructure Vault that has the password for the connection user. Ensure that you have set up the required policies for using OCI vault and secrets.

Creating an OCI Object Storage Data Asset

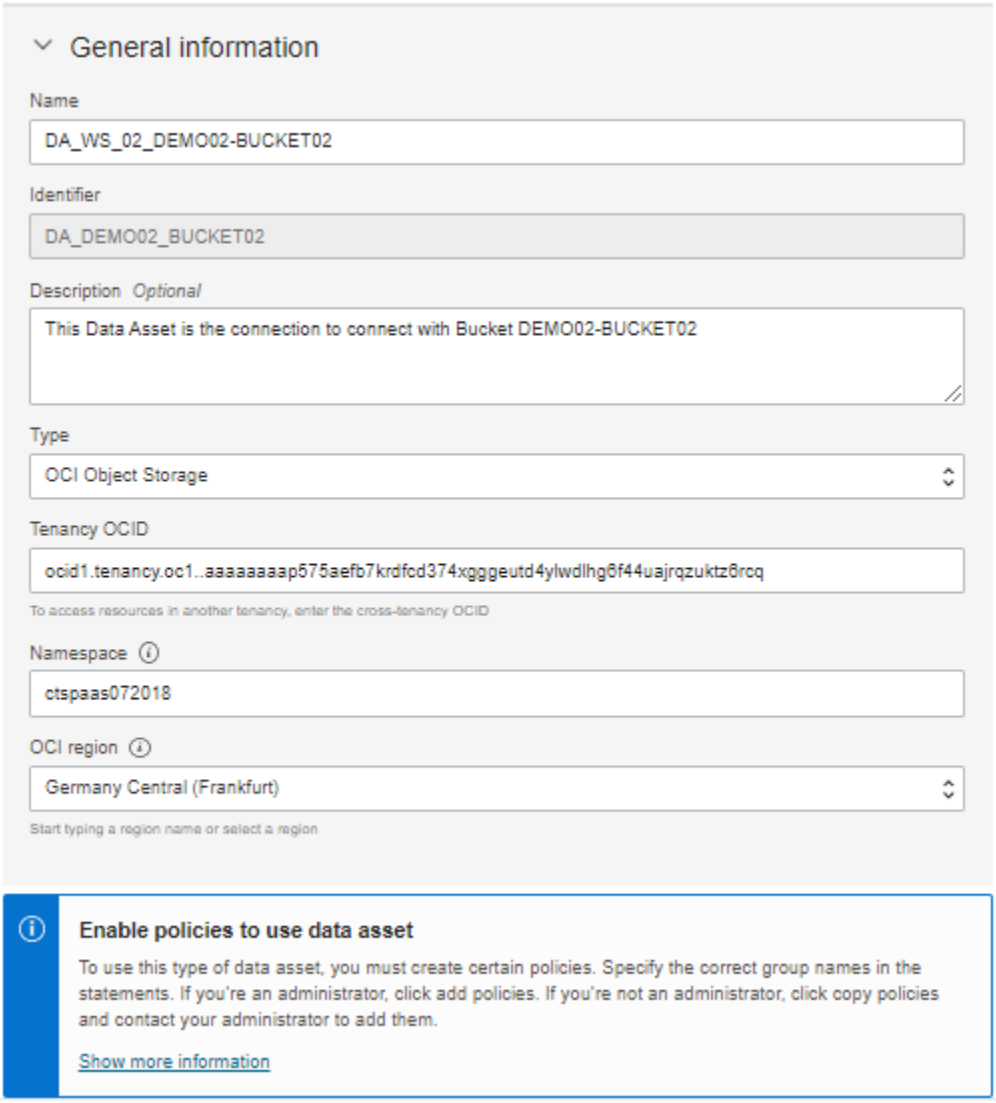
1) On the workspace's home page, select Create Data Asset
2) On the Create data asset page, enter a name and description (optional).
3) From the Type menu, select OCI Object Storage.
4) In Tenancy OCID, the Oracle Cloud ID of your current tenancy is auto-filled. You can enter a cross-tenancy OCID to access resources in another tenancy.
5) The Namespace field is auto-filled for you, based on the Tenancy OCID value.
6) For Namespace, enter the namespace for the Object Storage bucket, if the value is not auto-populated after completing the preceding step.
7) For OCI region, enter the region identifier for your Object Storage resource where your buckets are created.
8) Under Default connection information, enter a name and description.
9) After you complete all the required fields, you can click Test connection to ensure you have entered the data asset details correctly and click create.

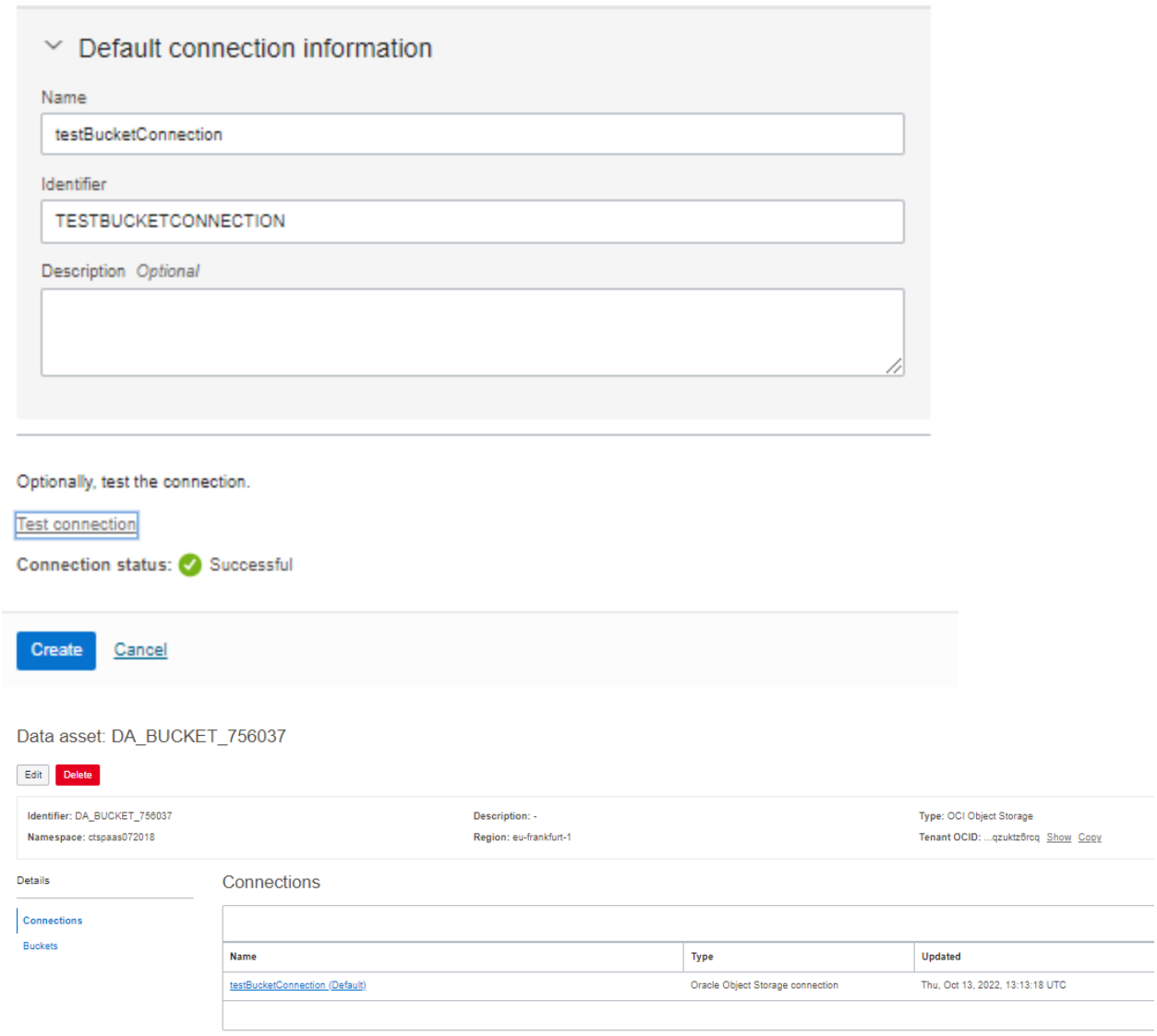
Data Asset -> Object Storage

allow group <*group-name*> to use object-family in compartment <*compartment-name*>

allow any-user to use buckets in compartment <*compartment-name*> where ALL {request.principal.type = 'disworkspace', request.principal.id = '<*workspace-ocid*>'}

allow any-user to manage objects in compartment <*compartment-name*> where ALL {request.principal.type='disworkspace',request.principal.id='<*workspace_ocid*>'}





Creating an Autonomous Data Warehouse or Autonomous Transaction Processing Data Asset

Using Autonomous Data Warehouse or Autonomous Transaction Processing as source or target data assets in Data Integration require database privileges and additional IAM policies.

Before you create an Autonomous Data Warehouse or Autonomous Transaction Processing data asset, ensure that you understand the permissions you might need to set up, and the use of a default staging location.

When you use Autonomous Data Warehouse or Autonomous Transaction Processing as a target, Data Integration uses an Object Storage bucket to stage the data before loading it into the target.

Policies and permissions to use Object Storage, and database privileges must be in place. Use the following checklist to ensure that you have the required setup you need.

Database privileges

Your database administrator must grant the following privileges to the Autonomous Data Warehouse or Autonomous Transaction Processing user in Data Integration. For example:

GRANT EXECUTE ON DBMS_CLOUD TO <*USER_NAME*> GRANT DWROLE TO <*USER_NAME*>

Data Asset Policies -> Autonomous Databases

Create this policy if you use an autonomous database as a target. Autonomous databases use Object Storage for staging data and need pre-authentication to complete operations.

allow any-user to manage buckets in compartment <*compartment-name*> where ALL {request.principal.type = 'disworkspace', request.principal.id = '<*workspace-ocid*>', request.permission = 'PAR_MANAGE'}

allow group <*group-name*> to read autonomous-database-family in compartment <*compartment-name*>

allow group <*group-name*> to manage dataflow-run in compartment <*compartment-name*>

Create data asset

A data asset represents a data source. Enter the data asset details, and then add the default connection details to that data asset.

General information

Name

DA_ATP_756037

Identifier

DA_ATP_756037

Description *Optional*

Type

Oracle Autonomous Transaction Processing

- (•) Upload wallet
  Upload the client credentials file downloaded from the autonomous database.
- ( ) Select database
  Client credentials are automatically retrieved. You must have the permission to download the file.
- ( ) Use vault secret OCID
  Specify OCID of the secret created in OCI Vault for the client credentials file and file password.

Wallet file

Drop file Select file

Wallet_ATPDB02.zip ×

Wallet password *Optional*

Treat NUMBER columns without precision and scale as VARCHAR.

Enable policies to use data asset

To use this type of data asset, you must create certain policies. Specify the correct group names in the statements. If you're an administrator, click add policies. If you're not an administrator, click copy policies and contact your administrator to add them.

Show more information

Default connection information

Name

testATPConnection

Identifier

TESTATPCONNECTION

Description *Optional*

Username

ADMIN

Password

••••••••••••••

TNS alias

atpdb02_low

Default staging location

To load data into an autonomous database, you need an Object Storage bucket to stage the data. Specify a default staging location that should be used when a staging location is not explicitly specified. To use an Object Storage bucket for staging, you must have PAR_MANAGE and OBJECT_CREATE permissions enabled.

Object storage data asset

DA_BUCKET_756037 | View all

Connection

testBucketConnection

Compartment

OCI-DEMO02

Bucket

DEMO02-BUCKET02 | View all

Optionally, test the connection.

Test connection

Connection status: Successful

Create | Cancel

Data asset: DA_ATP_756037

Edit | Delete

Identifier: DA_ATP_756037 **Description:** - **Type:** Oracle Autonomous Transaction Processing

Details

Connections

Schemas

Connections

Add connection | Search by connection

Name	Type	Updated
testATPConnection (Default)	Oracle ATP connection	Thu, Oct 13, 2022, 13:58:29 UTC

Showing 1 item | Page 1

Projects

You can access the projects in a workspace by using the navigation link on the workspace home page. Alternatively, you can select Projects from the Open tab (plus icon) menu in the tab bar.

Existing projects are listed on the Projects page. You can filter the list by project name. In the search field, enter the complete name of the project to do a full text search. You can also use the Favorite filter menu to display only those projects that are currently marked as favorite.

On the Projects list, the Actions menu for a project has these options:

View Details: Displays the project details, and the folders, tasks, and data flows that the project contains.

Edit: Lets you edit the name and description for the project.

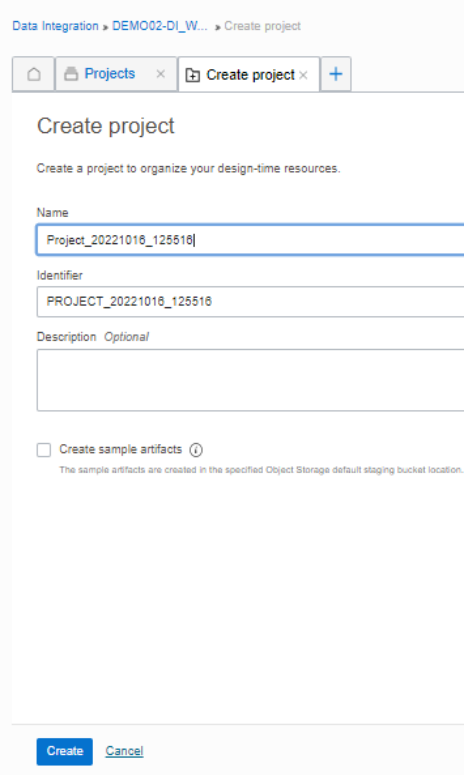
Copy Project Key: Copies to the clipboard the project key, which you can use with APIs.

Delete: Lets you delete the project.

Creating a Project

1) Click Create project on the Projects page.
2) On the Create project page, give your project a Name and optionally, a Description.
3) Click Create.

After you create the project, you're brought to the details page of your new project.



Folders

You can create folders in a project or in another folder, limited to two levels of folders. To access your folders, you must first navigate to the Projects page, select the parent project the folder belongs to, and then select Folders in the submenu on the project details page.

Existing folders are listed on the Folders page. You can filter the list by folder name. In the search field, enter the complete name of the folder to do a full text search. You can also use the Favorite filter menu to display only the folders that are currently marked as favorite.

On the Folders list, the Actions menu for a folder has these options:

View Details: Displays the folder details, and the tasks and data flows that the folder contains.

Edit: Lets you edit the name and description for the folder.

Copy Folder Key: Copies to the clipboard the folder key, which you can use with APIs.

Move: Lets you move the folder from its current location to another project or folder.

Delete: Lets you delete the folder.

Creating a Folder

You can create a folder in a project, or in another folder.

1) On the Projects page, select a project you want to create a folder in.
2) On the project details page, click Folders in the submenu.
3) On the Folders page, click Create folder.
4) In the Create folder panel, give your folder a Name and optionally, a Description.
5) Click **Create**.