Data Science Setup Steps

Instructions:

For each step, check whether the described resources are set up. If not, have a tenancy admin set them up.

Table of Contents

[IAM Setup 1](#_Toc93652306)

[Compartment 1](#_Toc93652307)

[User Group 1](#_Toc93652308)

[Dynamic Group Matching Rules 1](#_Toc93652309)

[Policies 2](#_Toc93652311)

[Infrastructure Setup 2](#_Toc93652312)

[Network 2](#_Toc93652313)

[Data Science 3](#_Toc93652314)

[Import a Dataset into a Notebook 3](#_Toc93652315)

[From an Object Storage Bucket 3](#_Toc93652316)

[From an Autonomous Data Warehouse (ADW) 3](#_Toc93652317)

# IAM Setup

## Compartment

*Create a Compartment to isolate Data-Science-related resources.*

Create Compartment DataScienceCompartment.

## User Group

*Group for users who manage Data-Science-related operations.*

Create Group DataScienceGroup. Add your OCI user to the group.

## Dynamic Group Matching Rules

*Group defined by a rule that matches particular resource-types within a given compartment. Enables management of Data-Science-related operations.*

Create Dynamic Group DataScienceDynamicGroup. Add the following as a matching rule:

## Any {fnfunc.compartment.id='<OCID of DataScienceCompartment>', ApiGateway.compartment.id='<OCID of DataScienceCompartment>', datasciencemodeldeployment.compartment.id='<OCID of DataScienceCompartment>', datasciencejobrun.compartment.id='<OCID of DataScienceCompartment>', datasciencenotebooksession.compartment.id='<OCID of DataScienceCompartment>'}

## Policies

*Collection of policy statements that grant permissions for user groups, dynamic groups, and services. Enables management of Data-Science-related operations.*

At the root (tenancy) level, create policy DataSciencePolicy with the following statements.

Note: If the root compartment is not the immediate parent of DataScienceCompartment, replace DataScienceCompartment with the path to DataScienceCompartment, e.g. ParentCompartment3:ParentCompartment2:ParentCompartment1:DataScienceCompartment

Note: The aggregate resource type 'data-science-family' expands to multiple resource-types, described [here](https://docs.oracle.com/en-us/iaas/data-science/using/policies.htm#resource-types).

Allow group DataScienceGroup to manage data-science-family in compartment DataScienceCompartment

Allow group DataScienceGroup to use virtual-network-family in compartment DataScienceCompartment

Allow group DataScienceGroup to manage functions-family in compartment DataScienceCompartment

Allow group DataScienceGroup to manage api-gateway-family in compartment DataScienceCompartment

Allow group DataScienceGroup to manage object-family in compartment DataScienceCompartment

Allow group DataScienceGroup to use autonomous-database-family in compartment DataScienceCompartment

Allow group DataScienceGroup to manage repos in tenancy

Allow dynamic-group DataScienceDynamicGroup to use virtual-network-family in compartment DataScienceCompartment

Allow dynamic-group DataScienceDynamicGroup to use functions-family in compartment DataScienceCompartment

Allow dynamic-group DataScienceDynamicGroup to manage public-ips in compartment DataScienceCompartment

Allow dynamic-group DataScienceDynamicGroup to manage data-science-family in compartment DataScienceCompartment

Allow dynamic-group DataScienceDynamicGroup to manage object-family in compartment DataScienceCompartment

Allow dynamic-group DataScienceDynamicGroup to use autonomous-database-family in compartment DataScienceCompartment

Allow service datascience to use virtual-network-family in compartment DataScienceCompartment

Allow service FaaS to use virtual-network-family in compartment DataScienceCompartment

Allow service FaaS to read repos in tenancy

# Infrastructure Setup

## Network

*Network that will host the OCI Data Science instance.*

In a Compartment for networking resources, create a VCN (e.g. DS VCN), and within that VCN, create:

* A private Subnet (e.g. DS Subnet)
* A Service Gateway (e.g. Service Gateway) that allows access to all services
* A NAT Gateway (e.g. NAT Gateway)
* A Route Table (e.g. DS Route Table) with a rule that enables access to all services via the Service Gateway, and a rule that enables access via 0.0.0.0/0 via the NAT Gateway. The Route Table should be assigned to the private Subnet.
* A Security List (e.g. DS Security List) with rules that allow ingress, egress via All Protocols with source, destination 0.0.0.0/0. The Security List should be assigned to the private Subnet.

## Data Science

*OCI Data Science Platform resources.*

In Compartment DataScienceCompartment, create a Data Science Project, and within that Project, create:

* Data Science Notebook Session: [VM shape](https://docs.oracle.com/en-us/iaas/Content/Compute/References/computeshapes.htm) (e.g. VM.StandardE2.2), suitable Block Volume size (e.g. 100 GB), provision in the private subnet

# Import a Dataset into a Notebook

[ADS Documentation](https://docs.oracle.com/en-us/iaas/tools/ads-sdk/latest/index.html)

[ADS Documentation - Quick Start Guide](https://docs.oracle.com/en-us/iaas/tools/ads-sdk/latest/user_guide/quickstart/quickstart.html)

[ADS Documentation - Connecting to Data Sources](https://docs.oracle.com/en-us/iaas/tools/ads-sdk/latest/user_guide/loading_data/connect.html)

If a sample .csv dataset is needed, [this Iris dataset](https://objectstorage.us-ashburn-1.oraclecloud.com/p/PjnbwhhdixsEAnKuWJJ5a5YdcUgXBK0X8wvkbRJOQAstqJD9Ov1yUDEYfSv8-H7Z/n/orasenatdpltintegration03/b/samcac/o/Iris.csv) can be used.

## From an Object Storage Bucket

Install and use Conda Environment: dataexpl\_p37\_cpu\_v2

Upload [this file](https://objectstorage.us-ashburn-1.oraclecloud.com/p/FP6mD8uri0D-IQNcjNTK5MK-aGELa4YpZoZ-3Kgj65E2nJLnL0UPPwgiyqgIqFzV/n/orasenatdpltintegration03/b/samcac/o/ImportExport_ObjS.ipynb) into your Notebook Session, and replace <values> with your own.

You can also import the file by opening Terminal and running the following command:

wget https://objectstorage.us-ashburn-1.oraclecloud.com/p/FP6mD8uri0D-IQNcjNTK5MK-aGELa4YpZoZ-3Kgj65E2nJLnL0UPPwgiyqgIqFzV/n/orasenatdpltintegration03/b/samcac/o/ImportExport\_ObjS.ipynb

## From an Autonomous Data Warehouse (ADW)

Install and use Conda Environment: dataexpl\_p37\_cpu\_v2

Upload [this file](https://objectstorage.us-ashburn-1.oraclecloud.com/p/VwIIUvgakVY3mmn-0Et-LHFp5R2L6961-CJKuuwT2G6rbrtgt5GD5djeWzkoW36R/n/orasenatdpltintegration03/b/samcac/o/ImportExport_ADW.ipynb) into your Notebook Session, and replace <values> with your own.

You can also import the file by opening Terminal and running the following command:

wget https://objectstorage.us-ashburn-1.oraclecloud.com/p/VwIIUvgakVY3mmn-0Et-LHFp5R2L6961-CJKuuwT2G6rbrtgt5GD5djeWzkoW36R/n/orasenatdpltintegration03/b/samcac/o/ImportExport\_ADW.ipynb