Workspace Design and Setup: Data Science Project & Notebooks

Lab 01-1 Practices

Estimated Time: 30 mins

Setup Instructions for the Lab

In this practice an environment is pre-created to work with labs. This includes:

* 1. Creating a User
  2. Adding User to a User Group
  3. Creation of Dynamic Group
  4. Creation of Policies that allows User Groups and Dynamic Groups permissions to conduct operations in a tenancy, usually a user compartment.

**The following setup is already done for you.** You may read through it to familiarize with the required setup.

Creating a User

* 1. From the Navigation Menu**,** click **Identity** under Identity & Security.
  2. Click **Users**.
  3. Fill out the dialog box with the following details:
* **Name**: Provide a name
* **Description:** Provide a suitable description
* **Email**: Optional
* **Confirm Email**: Provide the same email id
* Click **Create.**

Adding a user to a user group

* 1. From the Navigation Menu**,** click **Groups** under Identity & Security.
  2. Click **Create Group**.
  3. Fill out the dialog box with following details:
* **Name**: <DataScienceGroup>
* **Description**: Provide suitable description
  1. Add the user created to this group.

Create Dynamic Group

* 1. From the Navigation Menu**,** click **Dynamic Groups** under Identity & Security.
  2. Click **Create Dynamic Group**.
  3. Fill out the dialog box:
* **Name**: <DataScienceGroup>.
* **Description**: Provide suitable description
  1. Under matching rules add the following rules:

ALL {resource.type='datasciencenotebooksession', resource.compartment.id=‘<replace-with-your-compartment-ocid>’}

ALL {resource.type='datasciencejobrun', resource.compartment.id=‘<replace-with-your-compartment-ocid>’}

ALL {resource.type='datasciencemodeldeployment', resource.compartment.id=‘<replace-with-your-compartment-ocid>’}

**Note:** Replace with the OCID of the compartment where you will create your notebook session. You can retrieve the OCID of the compartment by copying the partial displayed OCID when you click your **Compartment** name

**Note:** Refer to Dynamic Groups documentation if required. <https://docs.oracle.com/en-us/iaas/Content/Identity/Tasks/managingdynamicgroups.htm>

IAM Policies

* 1. From the Navigation Menu**,** click **Policies** under Identity & Security.
  2. Click **Create a policy**.
  3. Fill out the dialog box with the following details:
* **Name**: <Policy Name>.
* **Description**: Provide suitable description for policy.
* **Compartment**: *<your-compartment>.*
  1. Toggle to Show manual editor, and include the following statements:

Allow group <student group name> to manage data-science-projects in compartment <student compartment name>

Allow group <student group name> to manage data-science-notebook-sessions in compartment <student compartment name>

Allow group <student group name> to manage data-science-models in compartment <student compartment name>

Allow group <student group name> to manage data-science-model-deployments in compartment <student compartment name>

Allow group <student group name> to manage data-science-job-runs in compartment <student compartment name>

Allow group <student group name> to manage data-science-jobs in compartment <student compartment name>

Allow service datascience to use virtual-network-family in compartment <student compartment name>

Allow group <student group name> to manage log-groups in compartment <student compartment name>

Allow group <student group name> to use log-content in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to read data-science-projects in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to manage data-science-models in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to manage data-science-model-deployments in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to manage log-groups in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to use log-content in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to manage data-science-jobs in compartment <student compartment name>

Allow dynamic-group <student dynamic group name> to manage data-science-job-runs in compartment <student compartment name>

Allow group <student group name> to manage vaults in compartment <student compartment name>

Allow group <student group name> to manage buckets in compartment <student compartment name>

Allow group <student group name> to manage objects in compartment <student compartment name>

Allow group <student group name> to manage keys in compartment <student compartment name>

Allow group <student group name> to manage secrets in compartment <student compartment name>

Allow group <student group name> to manage objectstorage-namespaces in compartment <student compartment name>

Allow group <student group name> to manage data-transfer-jobs in compartment <student compartment name>

Allow group <student group name> to manage ai-service-anomaly-detection-project in compartment <student compartment name>

Allow group <student group name> to manage ai-service-anomaly-detection-data-asset in compartment <student compartment name>

Allow group <student group name> to manage ai-service-anomaly-detection-model in compartment <student compartment name>

Allow group <student group name> to manage ai-service-vision-model in compartment <student compartment name>

Get Started

Overview

Data Science is a fully managed and serverless platform for data science teams to build, train, and manage machine learning models in the Oracle Cloud Infrastructure.

Projects are collaborative workspaces for organizing and documenting Data Science assets, such as notebook sessions and models. In this workshop, each participant creates their own project.

Data Science notebook sessions are interactive coding environments for building and training models. Notebook sessions come with many pre-installed open source, and Oracle-developed machine learning and data science packages. In this workshop, each user has a notebook session containing several notebooks.

Condas are collections of specific libraries and other resources that simplify library management. They are used to run notebooks and deployed models. Conda Environment as somewhere between a Docker image and a Python virtual environment. Conda is like a virtual environment that lets you run Python processes in different environments with different versions of the same library.

Two block volume drives are associated with a notebook session. There is a boot volume that is initialized each time the notebook session is activated. Any data on the boot volume is lost when the notebook session is deactivated or terminated. There is an additional block storage unit that is preserved when a notebook session is deactivated, but it is destroyed when a notebook session is terminated. This block volume is represented by the /home/datascience directory, and it is where the JupyterLab notebooks, data files, installed custom software, and other files should be stored.

When a notebook session is created or activated, the compute instance, block storage, VCN, and subnet are provisioned and configured. These resources can be changed by deactivating a notebook session, and then selecting a different configuration when activating the notebook session again.

**Note:** The size of the block storage can only be increased.

This lab will guide you through creating a Data Science Project, creating a notebook session in your project, and then copying a repository to your newly created data science notebook session environment.

|  |
| --- |
|  |

In this lab, you’ll:

* Create your first Data Science project.
* Create a Data Science notebook session.
* Create API signing keys and configuration files.
* Download course notebooks and datasets to your Notebook Session.
* Install and use a conda environment.

Prerequisites

* The required IAM policies have been are already set for you.
* You must have access to the OCI Console.

Create your first Data Science project

In this practice, you will learn to create a Data Science Project.

Tasks

1. In the top right of Console, under **Regions**, select **US-West (Phoenix)** Region
2. From the Navigation Menu, under **Analytics & AI**, select **Data Science** under Machine Learning.
3. In the left pane, under **List Scope**, note that, your allocated compartment is pre-selected.
4. Click **Create Project**, and fill out the dialog box with the following details:

* **Name**: <*Data Science Project>*
* **Description**: <*Provide suitable Description>*

1. Click **Create.**
2. The project gets created and listed on the Projects page.

**Note**: The project page lists project information along with resources contained in the project, such as notebook sessions, models, jobs, and model deployments. We'll work with all of these in our upcoming labs.

Create and Open Notebook Session

In this practice, you will learn how to create a Notebook session and then open a Notebook within it.

Tasks

1. From the Navigation Menu, click **Data Science** under Analytics & AI.
2. Click *< your -Data Science Project>*.
3. Click **Notebook Sessions** under Resources.
4. Click **Create Notebook Session**.
5. The Create notebook session’s dialog box opens. Fill out the details:
   1. **Compartment**: <*your-compartment*>
   2. **Name**: <*DataScience\_Labs*>
   3. **Compute Shape:** Click **Change shape**.
   4. **Shape Series:** Intel
   5. **Shape name: VM. Standard.E4. Flex**
   6. Click **Select shape**.
   7. **Number of OCPUs**:2
   8. **Block storage size:** 50 GB
   9. **Networking type:** Default Networking
6. Click **Create.**
7. Check the **View detail page** on clicking Create.
8. While the notebook session is being created, you will be taken to the detail page.  The notebook session will take about 4 minutes to provision. When it is ready, the Open button will be enabled.
9. Click **Open** to open your notebook session.

Create API Signing Keys and Configuration Files

Here, you will see how to create API keys. You will also learn to create your configuration file and save it in your Terminal once you open your notebook session. The API keys and configuration files are required in some subsequent lab practices.

Tasks

1. Setup API Signing Key
   1. From the OCI console, click **Profile** icon on the top right and then select User Settings.
   2. Under resources, navigate to API key and click **Add API Key.**
   3. In the dialog, select **Generate API Key Pair**. Click **Download Private Key** and save the key to your local computer, and you will upload it later to the Terminal.
   4. Rename this file as oci-api-key.pem.
   5. Click **Add**.
   6. A configuration file window will pop up. Copy the content of **Configuration File Preview** and save it in your local computer.
2. Setting up API Key and Configuration File in the Terminal.
   1. From the notebook session, open the **Terminal**.
   2. Make sure you are in the </home/datascience> directory.
   3. In the Terminal, type the following command to create .oci folder:

$ mkdir ~/.oci

* 1. In the left pane, click **Up Arrow** icon and upload oci-api-key.pem.
  2. Move the key to the .oci folder. Give permission only to the user by using the following command in the Terminal:

$ mv oci-api-key.pem ~/.oci/

$ chmod 600 ~/.oci/oci-api-key.pem

* 1. From the Launcher in your Notebook session, click Terminal to start a new terminal session.
  2. Using your favorite editor, create a config file and paste the earlier copied configuration file preview from step 1 f. For example, you can use vi editor which comes installed as follow

$ vi ~/.oci/config

* 1. You can check and confirm that the file has been created and written successfully by

$ cat ~/.oci/config

It should look something like

[DEFAULT]

user=ocid1.user.oc1..xxxxxxxxxxxxxxxxxxxxxxxx

fingerprint=xx:xx:xx:xx:xx:xx

tenancy=ocid1.tenancy.oc1..xxxxxxxxxxxxxxxxxx

region=<us-ashburn-1>

key\_file=<path to your private key file which you stored in .oci folder in step 2e>

Confirm that the details are correct. You can update the file permission to be accessible by owner only by

$ chmod 600 ~/.oci/config

Download course notebooks and datasets to your Notebook Session

After completing the earlier tasks (create and open notebooks session), use the following steps to download the notebooks and datasets from the bucket.

Tasks

1. Open a terminal window.
2. Execute the following command in the terminal window:

$ oci os object bulk-download -ns <your-tenancy> -bn DSP-LABS --download-dir ./labs --auth 'resource\_principal'

Note: Replace with your tenancy name.

1. You should see the lab folder in the JupyterLab file browser window on the left. The content of this practice is under the following path:

</home/datascience/labs/>

Install and Use a Conda Environment

In this practice you will learn to Install and use conda environments within the Data Science Notebook. Use the Conda Environment Explorer tool in the notebook session to install the required conda environment. Once the environment is installed, you will run a notebook in a conda environment kernel.

Tasks

Before you can use a conda environment in your notebook session, you need to install it. You will install 3 pre-built Data Science Conda Environments.

1. Installation of Published Conda Environment
2. Install a conda environment <env name> from the bucket <bucket>
   1. Go to the Launchertab and select **Terminal** to open a terminal window.

**Paste the command** into the terminal window and press **Enter** to execute it. The command that you previously copied is:  
$ odsc conda install --uri oci://LAB\_Conda@<your-tenancy>/data-science-gmlv1\_0

Note: Replace with your tenancy name

* 1. Note: At the end of the above install, you will see instructions to activate the conda environment. Make sure the conda environment is activated before using it.
  2. You can confirm that the conda environment has been successfully installed by going back to the Launcher tab and see that a new kernel has been created with the slugname of the conda environment. It takes about 3-5 minutes for the conda package to be installed.
  3. Select the kernel and click **Create Notebook.**

1. Installation of Conda Environment for TensorFlow 2.8 for CPU on Python 3.8.
   1. On the Launcher tab, click **Environment Explorer**.
   2. On the **Environment Explorer** tab, scroll down until you find the **TensorFlow 2.8 for CPU on Python 3.8**. (If you see no results, use the Refresh button on the right side of the filter bar of Environment Explorer.)
   3. Click the **caret** on the right side and copy the install command.
   4. Go back to the Launchertab and select **Terminal** to open a terminal window.
   5. **Paste the command** into the terminal window and press **Enter** to execute it. The command that you previously copied is:  
      $ odsc conda install -s tensorflow28\_p38\_cpu\_v1
   6. You will receive a prompt related to what version number you want. Press **Enter** to select the default.
   7. It takes about 3-5 minutes for the conda package to be installed.
   8. You can confirm that the conda environment has been successfully installed by going back to the Launcher tab and see that a new kernel has been created with the slugname of the conda environment.
2. Installation of Conda Environment for automlx\_p38\_cpu\_v5 for CPU on Python 3.8.
   1. On the Launcher tab, click **Environment Explorer**.
   2. On the **Environment Explorer** tab, scroll down until you find the **Oracle AutoML and Model Explanation for Python 3.8**. (If you see no results, use the Refresh button on the right side of the filter bar of Environment Explorer.)
   3. Click the **caret** on the right side and copy the install command.
   4. Go back to the Launchertab and select **Terminal** to open a terminal window.
   5. **Paste the command** into the terminal window and press **Enter** to execute it. The command that you previously copied is:  
      $ odsc conda install -s automlx\_p38\_cpu\_v5
   6. You will receive a prompt related to what version number you want. Press **Enter** to select the default.
   7. It takes about 3-5 minutes for the conda package to be installed.
   8. You can confirm that the conda environment has been successfully installed by going back to the Launcher tab and see that a new kernel has been created with the slugname of the conda environment.

Select the kernel and click **Create Notebook.**

**Note:** You will require these two conda environments for the subsequent labs. If any other condo environment needs to be installed, follow the above instructions to set up and install the conda environment.

Purge Instructions

Deactivating Notebook Sessions

1. From the navigation menu, under **Analytics & AI** and click on **Data Science**.
2. You are now on Projects page.
3. Click *<your-Data Science-project>*.
4. Click **Notebook sessions** under Resources**.**
5. Click **Actions** icon under *<your-notebook-session>* and then click **Deactivate.**
6. When prompted for confirmation, click **Deactivate**.

We will be using the same project and notebook sessions for our subsequent labs. If you happen to delete the artifact, kindly follow this lab instructions again to create a new project and notebook session.