**Machine Learning before the hands-on lab setup guide**

**Requirements**

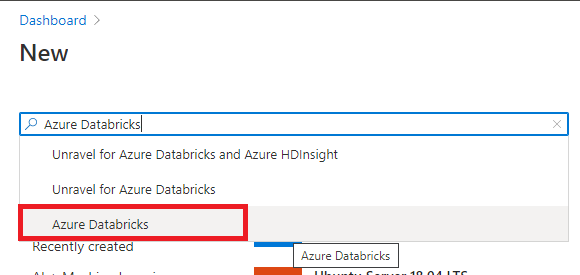
1. Microsoft Azure subscription must be pay-as-you-go or MSDN
   * Trial subscriptions will not work. You will run into issues with Azure resource quota limits.
   * Subscriptions with access limited to a single resource group will not work. You will need the ability to deploy multiple resource groups.

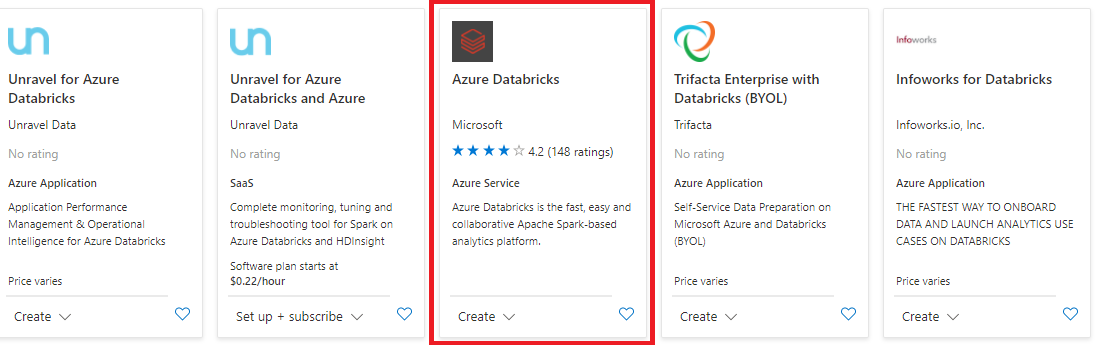
**Before the hands-on lab**

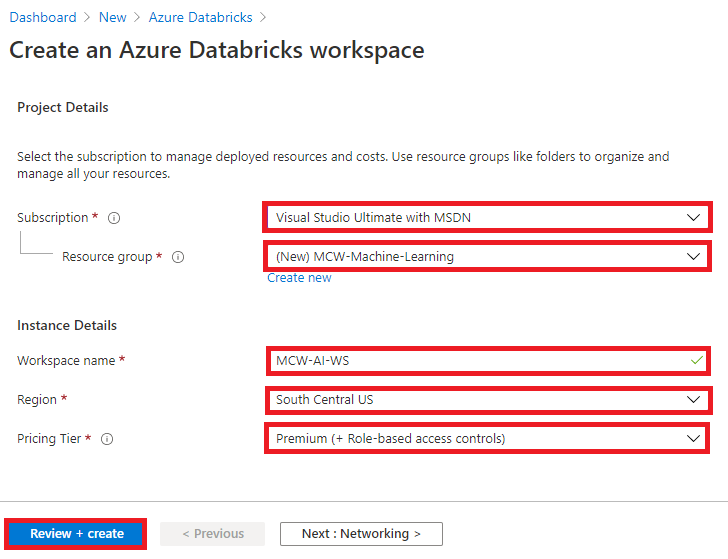
Duration: 30 minutes

**Task 1: Create your Azure Databricks Account**

Azure Databricks is an Apache Spark-based analytics platform optimized for Azure that supports data engineering and machine learning and deep learning workloads.

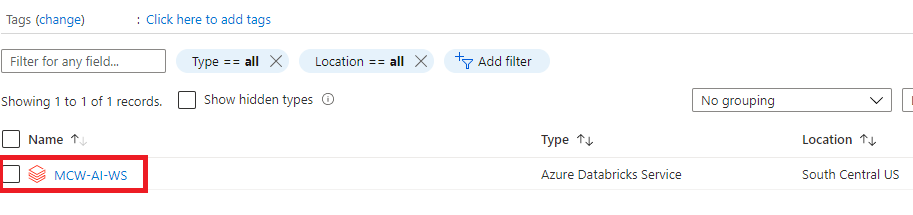
1. In the [Azure Portal](https://portal.azure.com/), select **+ Create a resource**, then type "Azure Databricks" into the seach bar. Select Azure Databricks from the results. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-1.png)

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-2.png)

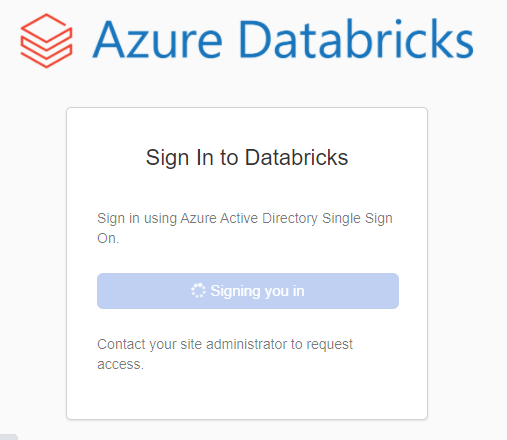
1. Select **Create** on the bottom of the blade that follows.
2. Set the following configuration on the Azure Databricks Service creation forms:
   * **Workspace name** : Enter a unique name as indicated by a green checkmark.
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource Group**: Select **Create new** and provide the name MCW-Machine-Learning
   * **Region**: Select a region close to you. ***(If you are using Azure Pass, select South Central US.)***
   * **Pricing**: Select Premium
3. Select **Review + Create** and then select **Create** when the form values passes validation. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-3.png)

**Task 2: Create an Azure Databricks cluster**

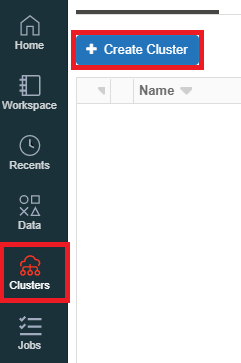
You have provisioned an Azure Databricks workspace, and now you need to create a new cluster within the workspace.

1. From within Azure portal navigate to your resource group name (e.g., MCW-Machine-Learning).
2. Next, select your Azure Databricks service from the list. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-4.png)
3. In the Overview pane of the Azure Databricks service, **Launch Workspace**.

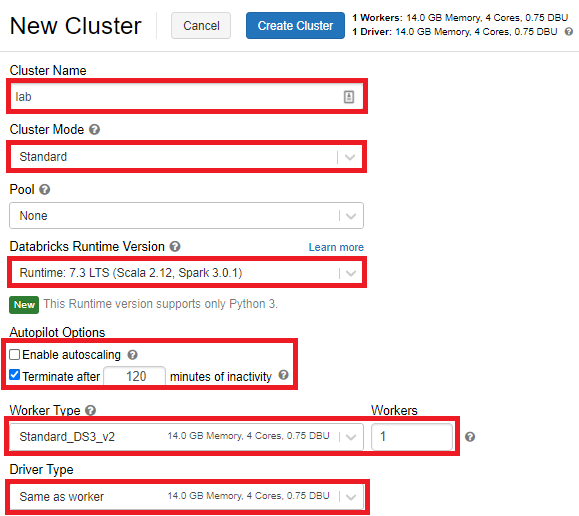
[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-5.png)

Azure Databricks will automatically log you in using Azure Active Directory Single Sign On. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-6.png)

1. Select **Clusters** (1) from the menu, then select **Create Cluster** (2).

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-7.png)

1. On the Create New Cluster form, provide the following:
   * **Cluster Name**: lab
   * **Cluster Mode**: **Standard**
   * **Databricks Runtime Version**: **Runtime: 7.3 LTS (Scala 2.12, Spark 3.0.1)**
   * **Enable autoscaling**: **Uncheck** this option.
   * **Terminate after \_\_\_ minutes of inactivity**: Leave **checked** and in the text box enter 120.
   * **Worker Type**: **Standard\_DS3\_v2**
   * **Workers**: 1
   * **Driver Type**: **Same as worker**

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-8.png)

1. Select **Create Cluster**. It will take few minutes to create the cluster. Please ensure that the cluster state is running before proceeding further.

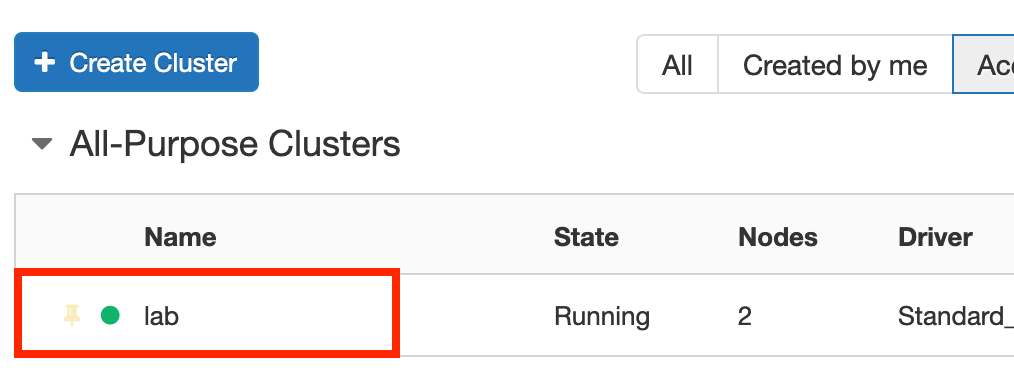
[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-9.png)

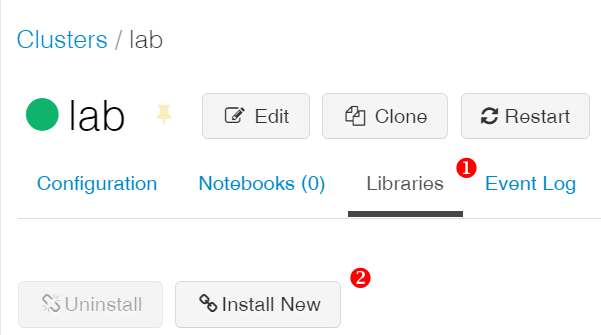
**Task 3: Install libraries on the Azure Databricks Cluster**

The notebooks you will run depends on certain Python libraries that will need to be installed in your cluster. The following steps walk you through adding these dependencies.

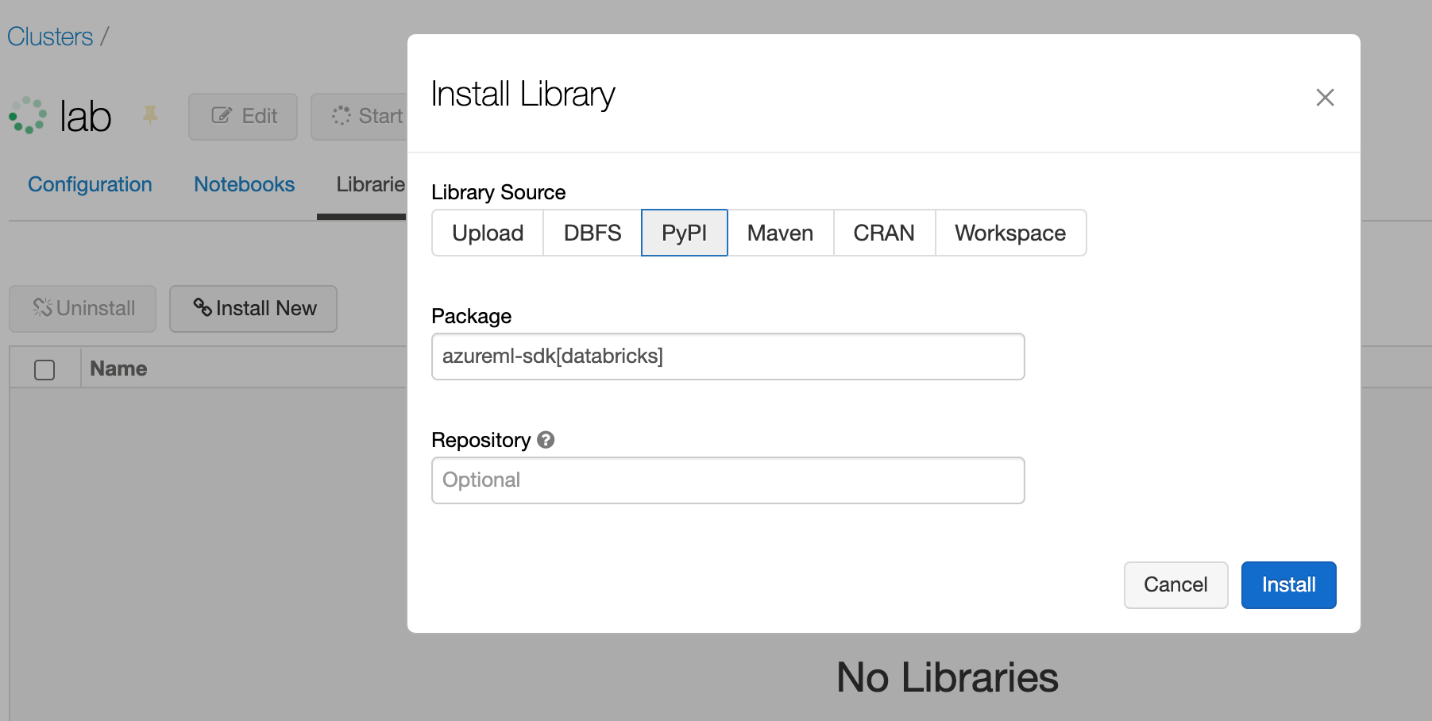
1. From the left-hand menu in your Workspace, select **Clusters**.

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-10.png)

1. In the list of clusters, select your cluster. Make sure the state of the cluster is Running. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-11.png)
2. Select the **Libraries** link and then select **Install New**.

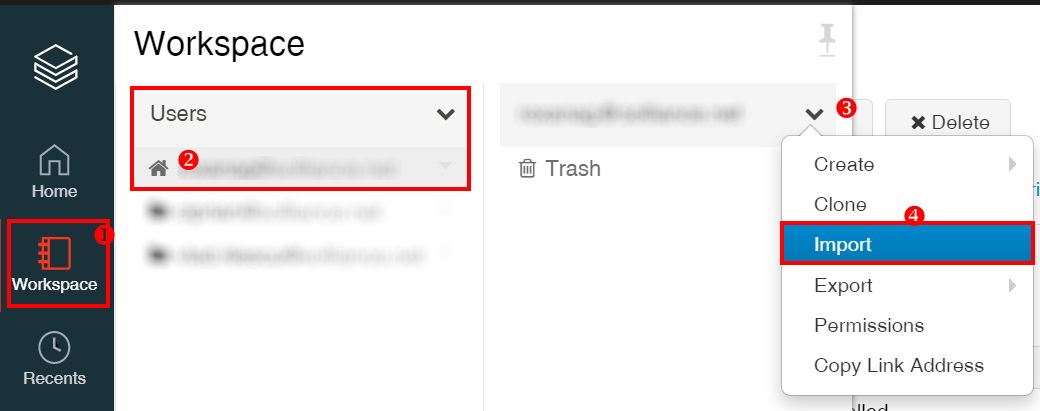
[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-12.png)

1. In the Library Source, select **PyPi** and in the Package text box type azureml-sdk[databricks] and select **Install**.

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-13.png)

1. An entry for azureml-sdk will appear in the list. The install will progress first with a status of installing followed by the status of installed.

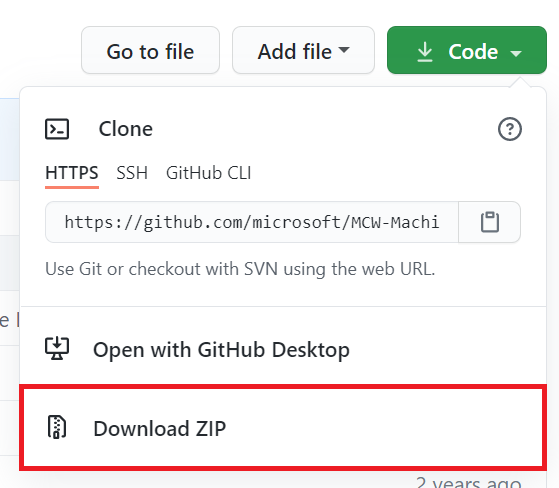
**Task 4: Upload the Databricks notebook archive**

1. Within the Azure Databricks Workspace, using the command bar on the left, select Workspace, Users and select your username (the entry with house icon).
2. In the blade that appears, select the downwards pointing chevron next to your name, and select **Import**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-14.png)
3. On the Import Notebooks dialog, select **URL** and paste in the following URL:

https://github.com/microsoft/MCW-Machine-Learning/blob/master/Hands-on%20lab/notebooks/AI%20with%20Databricks%20and%20AML.dbc?raw=true

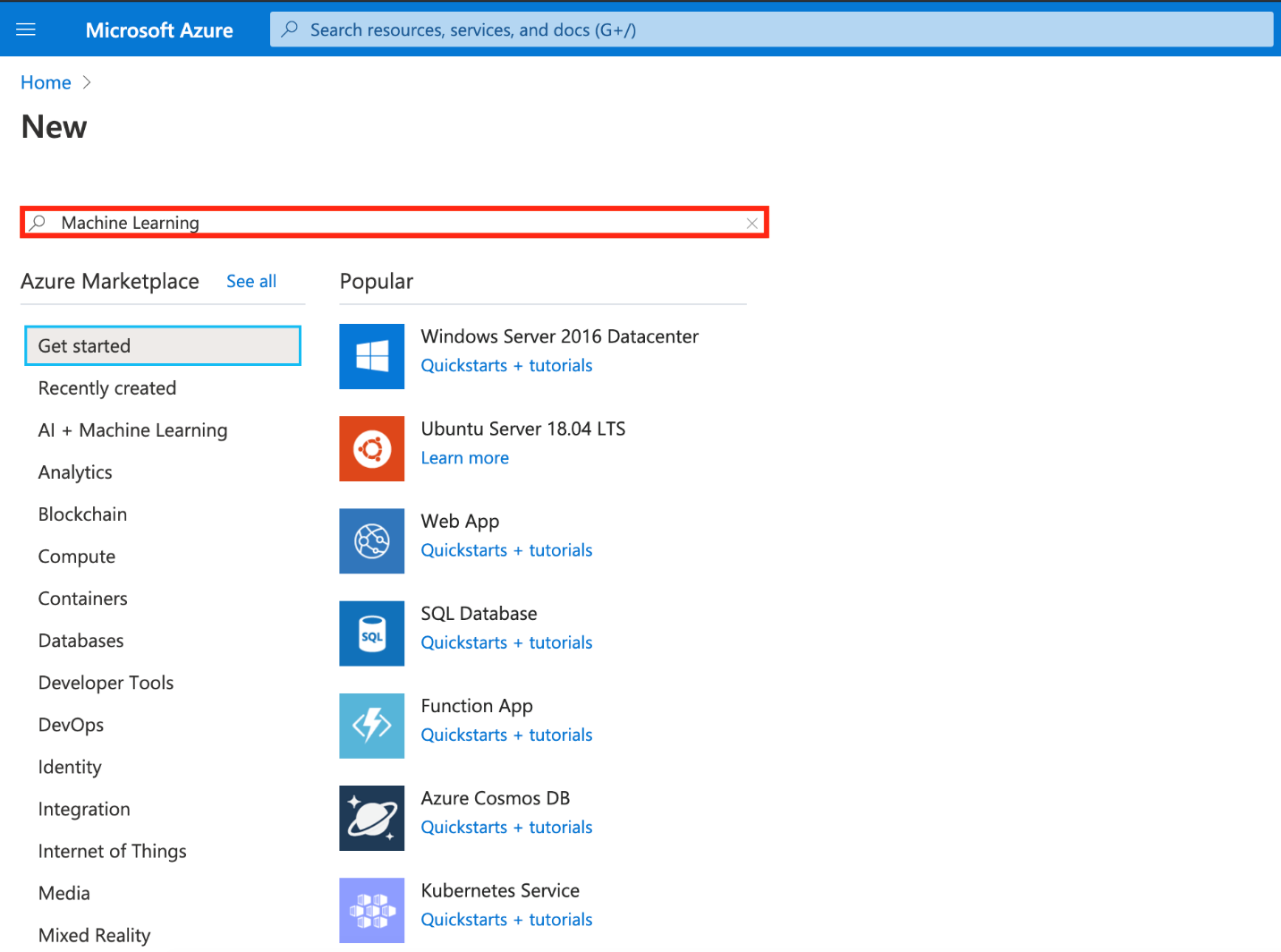
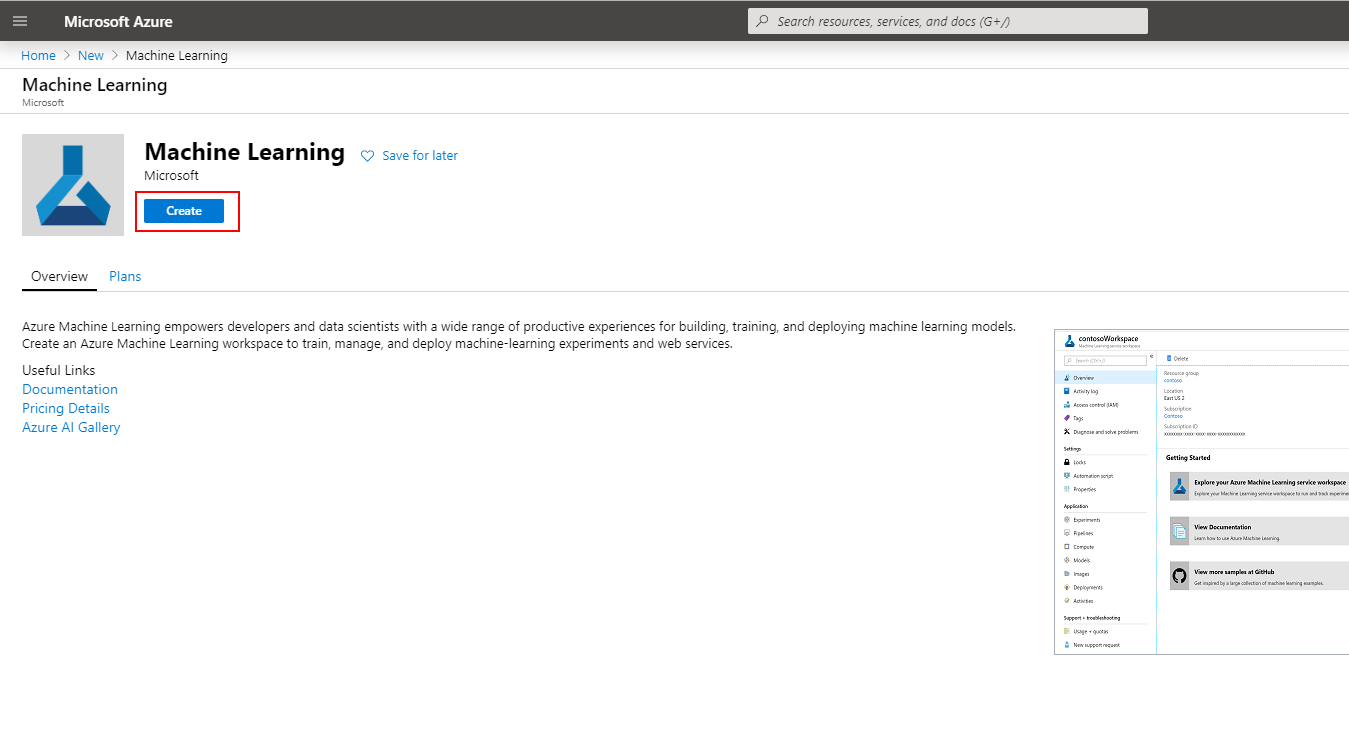
1. Select **Import**

**Note**: If you are facing difficulty with this URL, you can also consider cloning this repository locally, and uploading the file from your computer to databricks. To do this, download the zip archive of this repository, and extract it to a known location. Then, from the repository root, navigate to Hands-on lab/notebooks/AI with Databricks and AML.dbc.

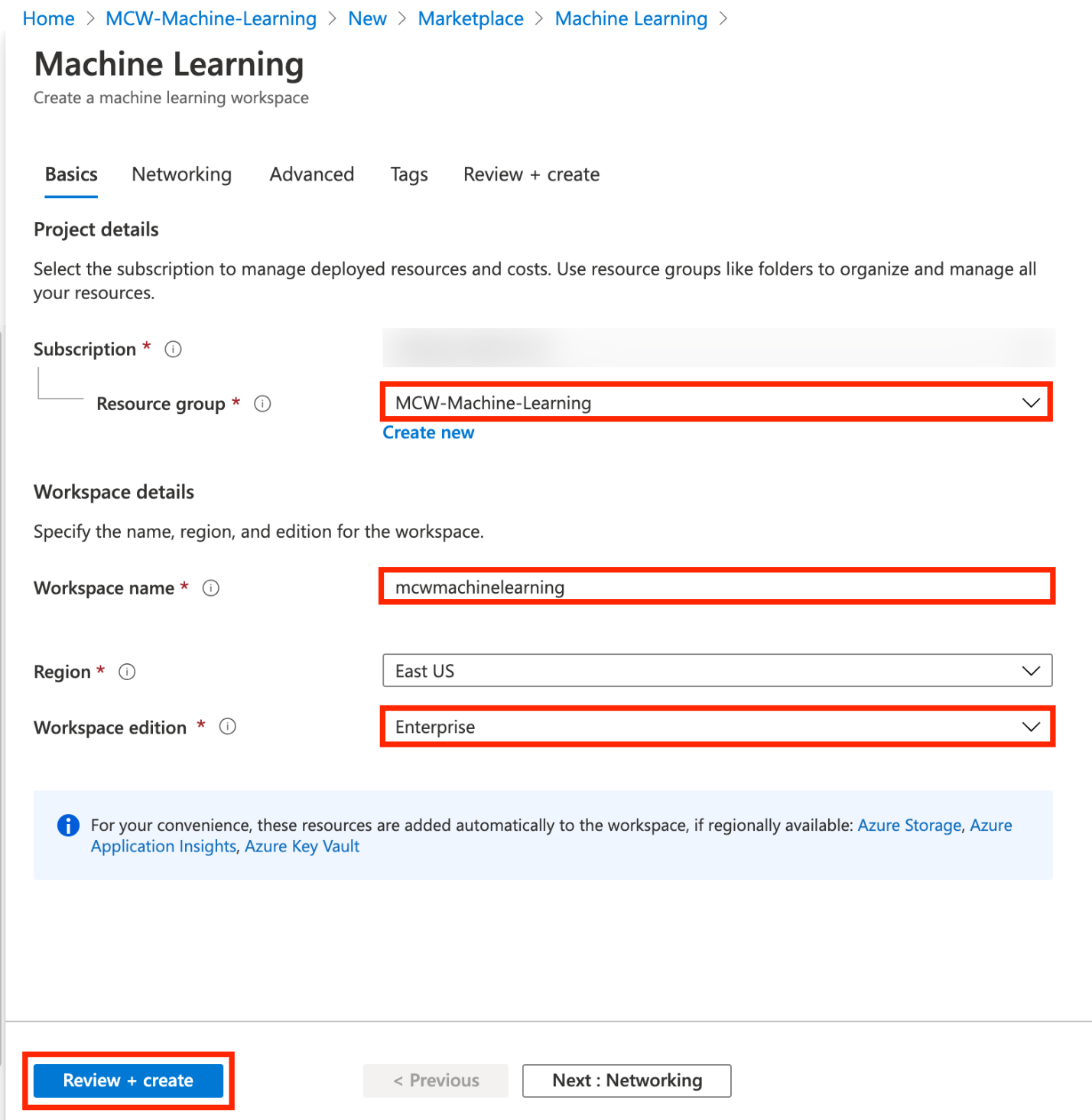
[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-15.png)

1. A folder named after the archive should appear. Select that folder.
2. The folder will contain one or more notebooks. These are the notebooks you will use in the completing this lab.

**Task 5: Create your Azure Machine Learning Workspace**

1. In the [Azure Portal](https://portal.azure.com/), select **+ Create a resource**, then type Machine Learning into the search bar. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-16.png)
2. Select **Create**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-17.png)
3. In the Create Machine Learning Workspace dialog that appears, provide the following values:
   * **Workspace Name**: mcwmachinelearning
   * **Subscription**: Choose your Azure subscription.
   * **Resource group**: Select the resource group in which you deployed your Azure Databricks workspace.
   * **Location**: Choose a region closest to you (it is OK if the Azure Databricks Workspace and the Azure Machine Learning Workspace are in different locations).
   * **Workspace edition**: Select Enterprise.

If did not showing **Workspace edition**, please ignore it.

1. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T1-18.png)
2. **Note**: The option to set the **Workspace edition** may not be present.
3. Select **Review + Create** and then select **Create** when the form values passes validation.

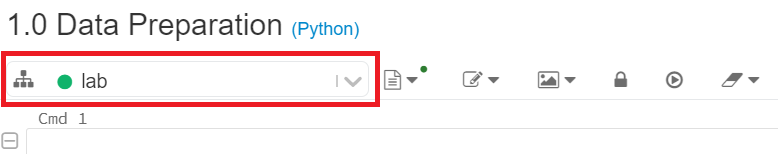
You should follow all this steps provided *before* attending the Hands-on lab.

**Exercise 1: Data exploration and preprocessing**

Duration: 40 minutes

Understanding data through data exploration is one of the core challenges faced today by data engineers and data scientists. Using raw data for modeling can produce misleading results, since data is often noisy and unreliable, and may be missing values. In this exercise, you will explore the raw data, transform and register the dataset in the Datastore. You will use it to train a forecasting model later in this hands-on-lab. The data preparation steps will be performed on the Azure Databricks cluster.

**Task 1: Load, explore and prepare the dataset using an Azure Databricks notebook**

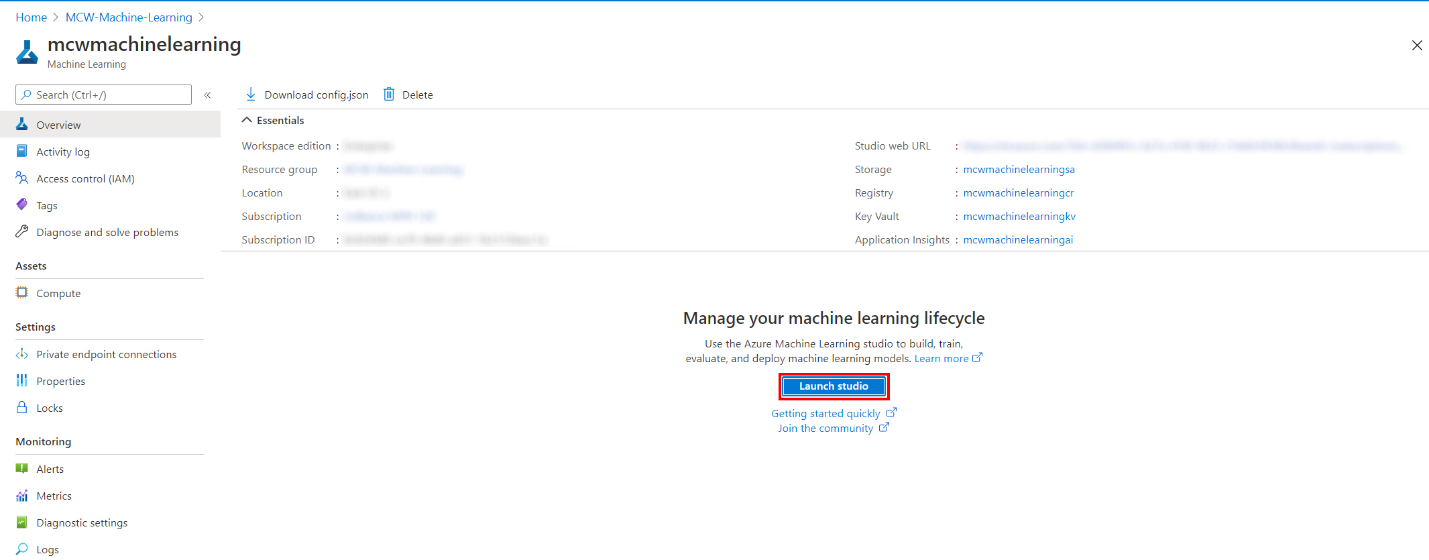
1. Browse to your Azure Databricks Workspace and navigate to AI with Databricks and AML \ 1.0 Data Preparation. This is the notebook you will step through executing in this exercise.
2. When you execute a notebook, you will need to attach it to a cluster. You can specify that the **lab** cluster should be used in the upper left-hand corner of the notebook. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T2-1.png)
3. Follow the instructions within the notebook to complete the exercise. Press Shift+Enter to execute a cell.

**Exercise 2: Creating a forecast model using automated machine learning**

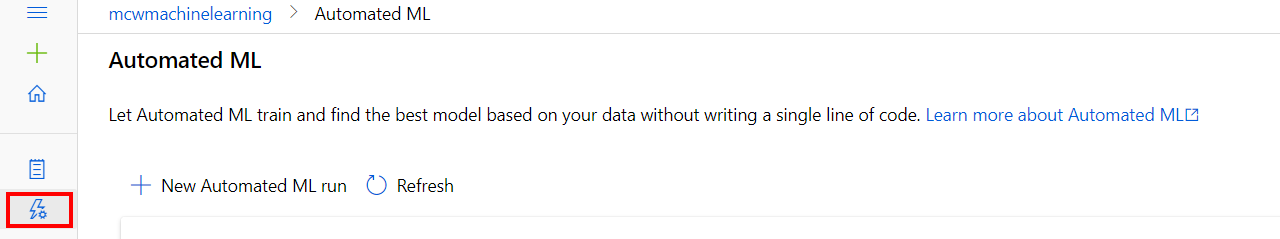
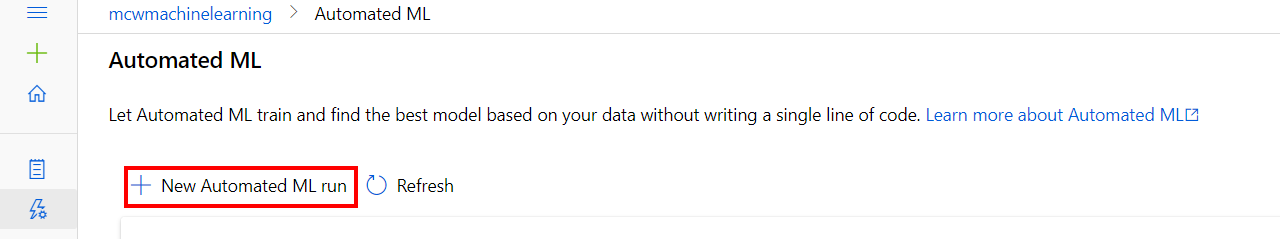
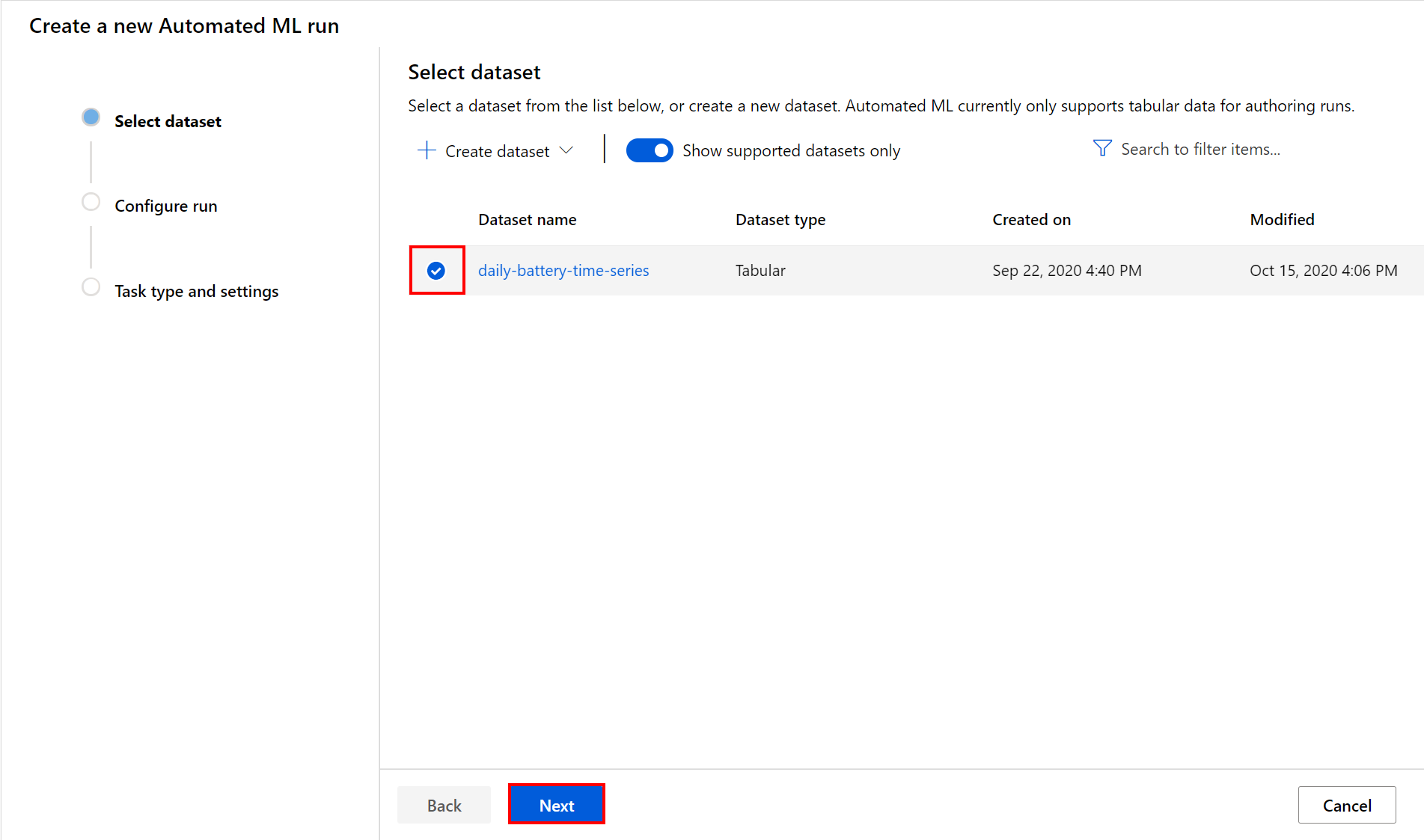
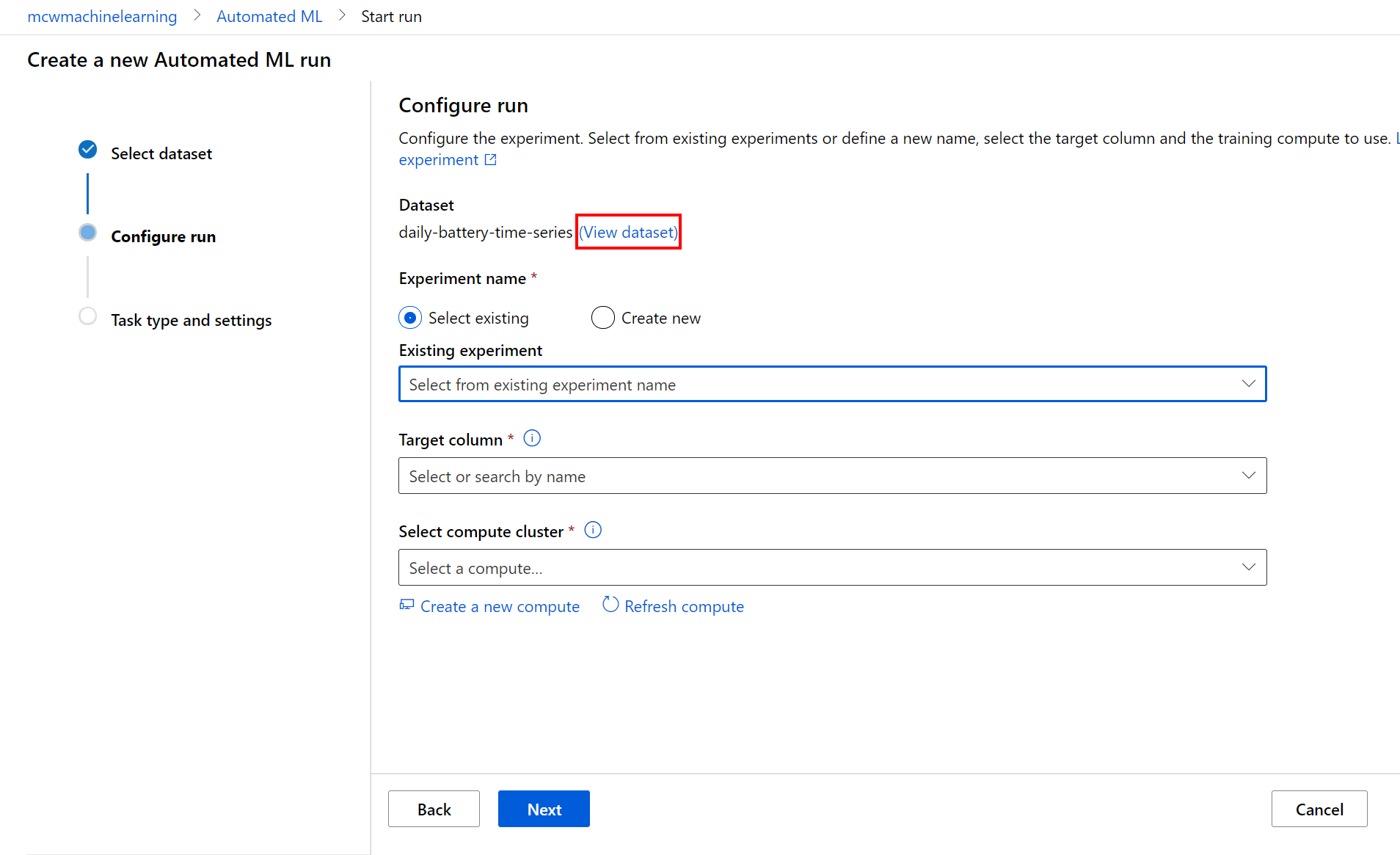
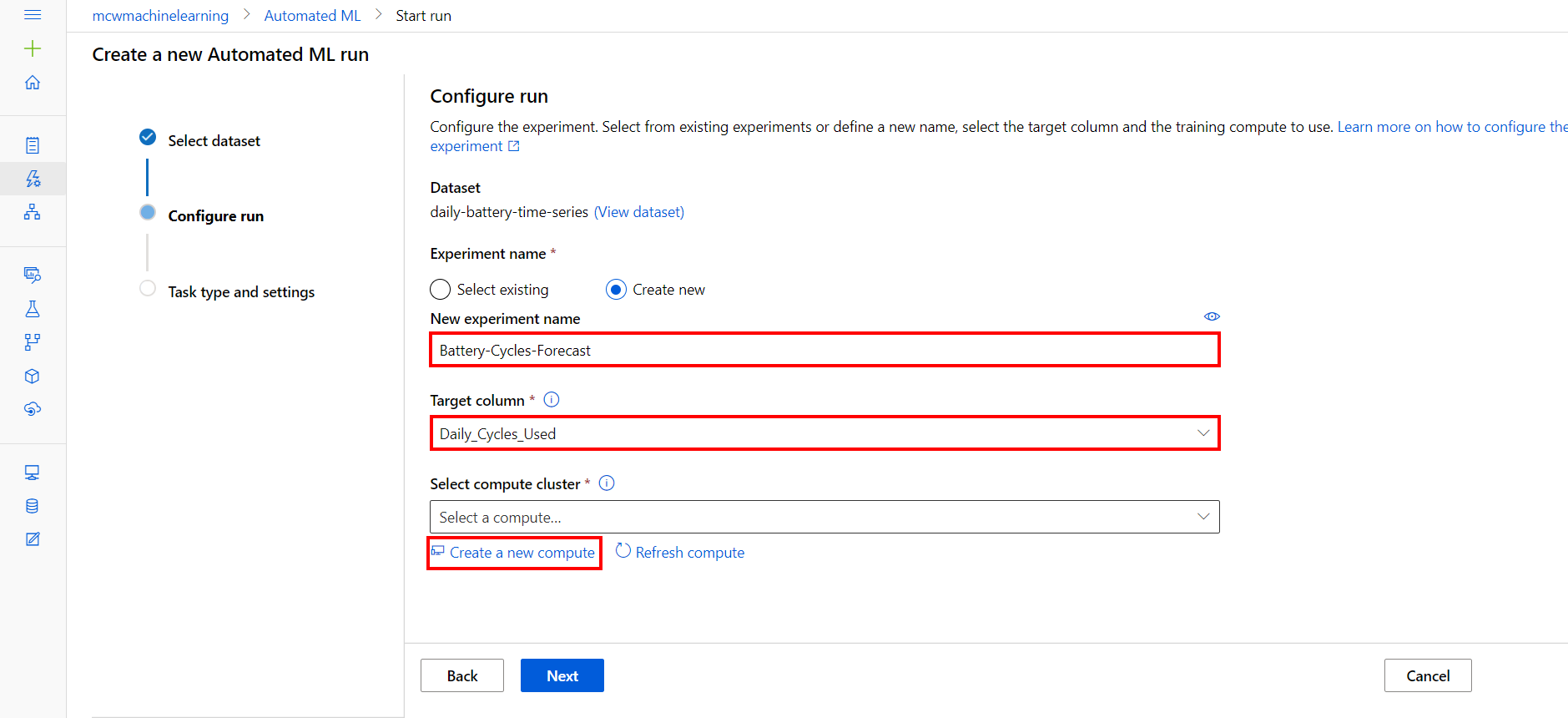
Duration: 60 minutes

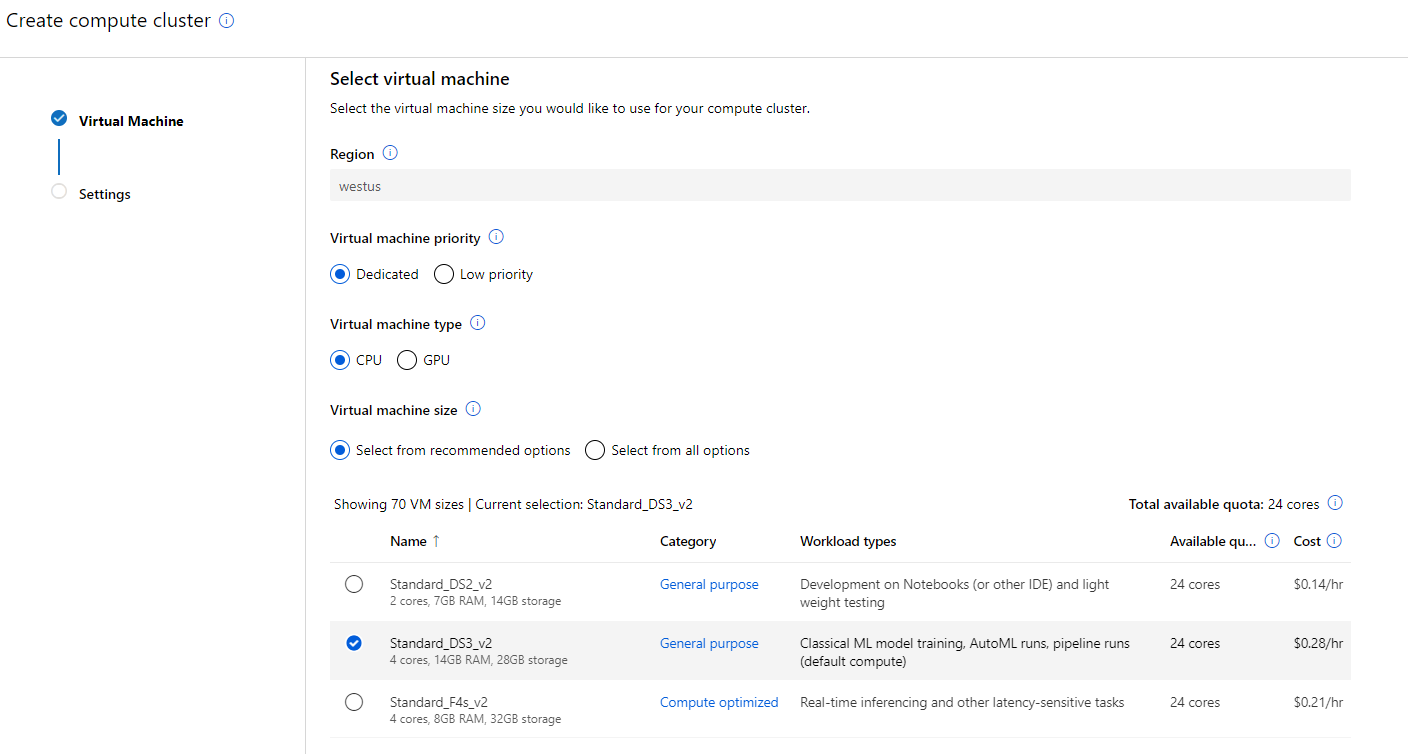
In this exercise, you will create a model that predicts battery failure from time-series data using the visual interface to automated machine learning in an Azure Machine Learning workspace. You will deploy the best model to Azure Container Instances (ACI), and you will also perform batch scoring from a Databricks notebooks.

**Task 1: Create an automated machine learning experiement**

1. Navigate to your Azure Machine Learning workspace in the Azure Portal. Select **Try the new Azure Machine Learning studio, Launch now**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-1.png)

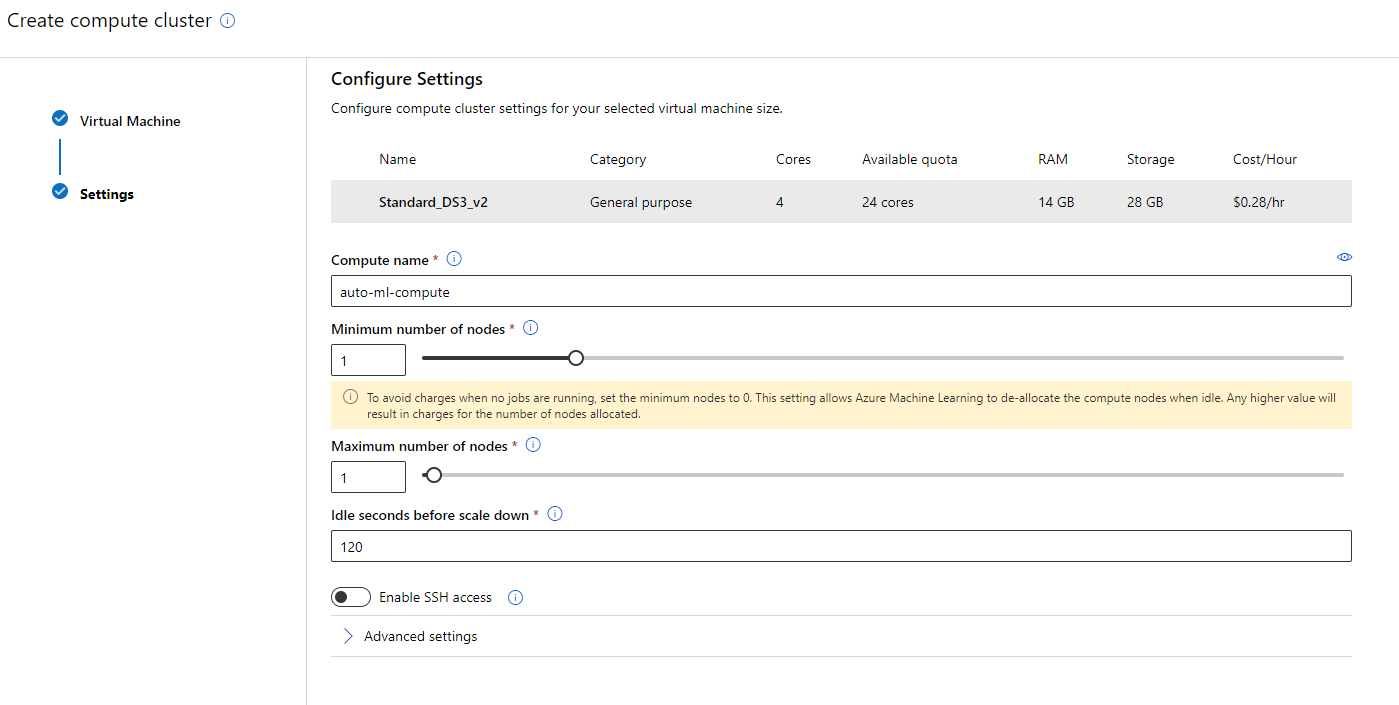
**Note**: Alternatively, you can sign-in directly to the [Azure Machine Learning studio portal](https://ml.azure.com/).

1. Select **Automated ML icon** in the left navigation bar. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-2.png)
2. Select **+ New automated ML run**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-3.png)
3. Select the daily-battery-time-series dataset from the list of registered datasets and then select **Next**. (This dataset was registered as a final step of the previous exercise, from the Azure Databricks notebook.) [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-4.png)
4. Review the dataset details in the Configure run section, by selecting the **View dataset** link next to the dataset name. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-5.png)
5. Provide the experiment name: Battery-Cycles-Forecast and select **Daily\_Cycles\_Used** as target column. Select **Create a new compute**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-6.png)
6. For the new compute, provide the following values and then select **Next**:
   * **Virtual machine priority**: Dedicated
   * **Virtual machine type**: CPU
   * **Virtual machine size**: Select from recommended options --> Standard\_DS3\_v2

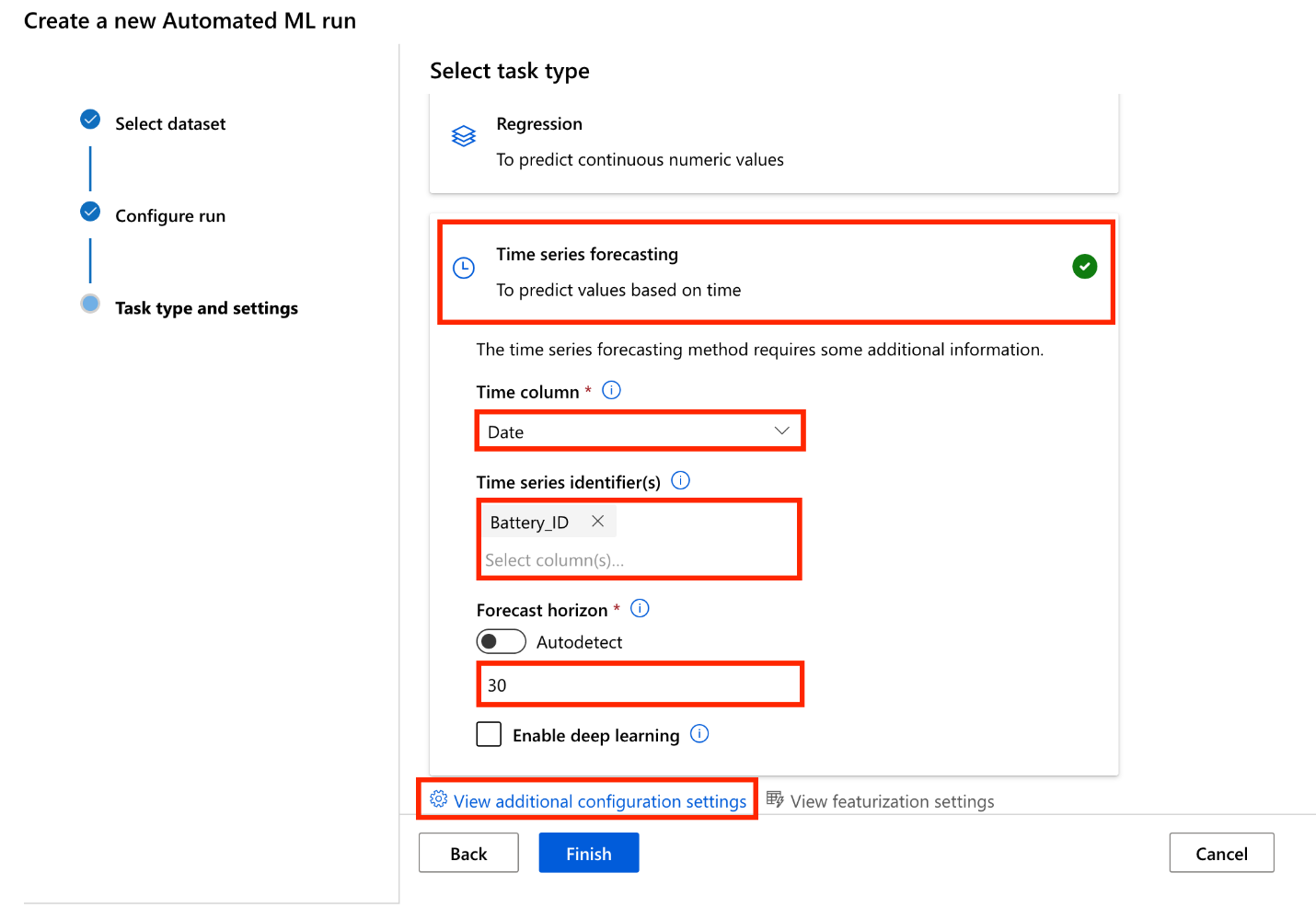
[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-7.png)

1. On the next page, provide the following parameters for your compute cluster. Then, select **Create**.
   * **Compute name**: auto-ml-compute
   * **Minimum number of nodes**: 1
   * **Maximum number of nodes**: 1
   * **Idle seconds before scale down**: 120

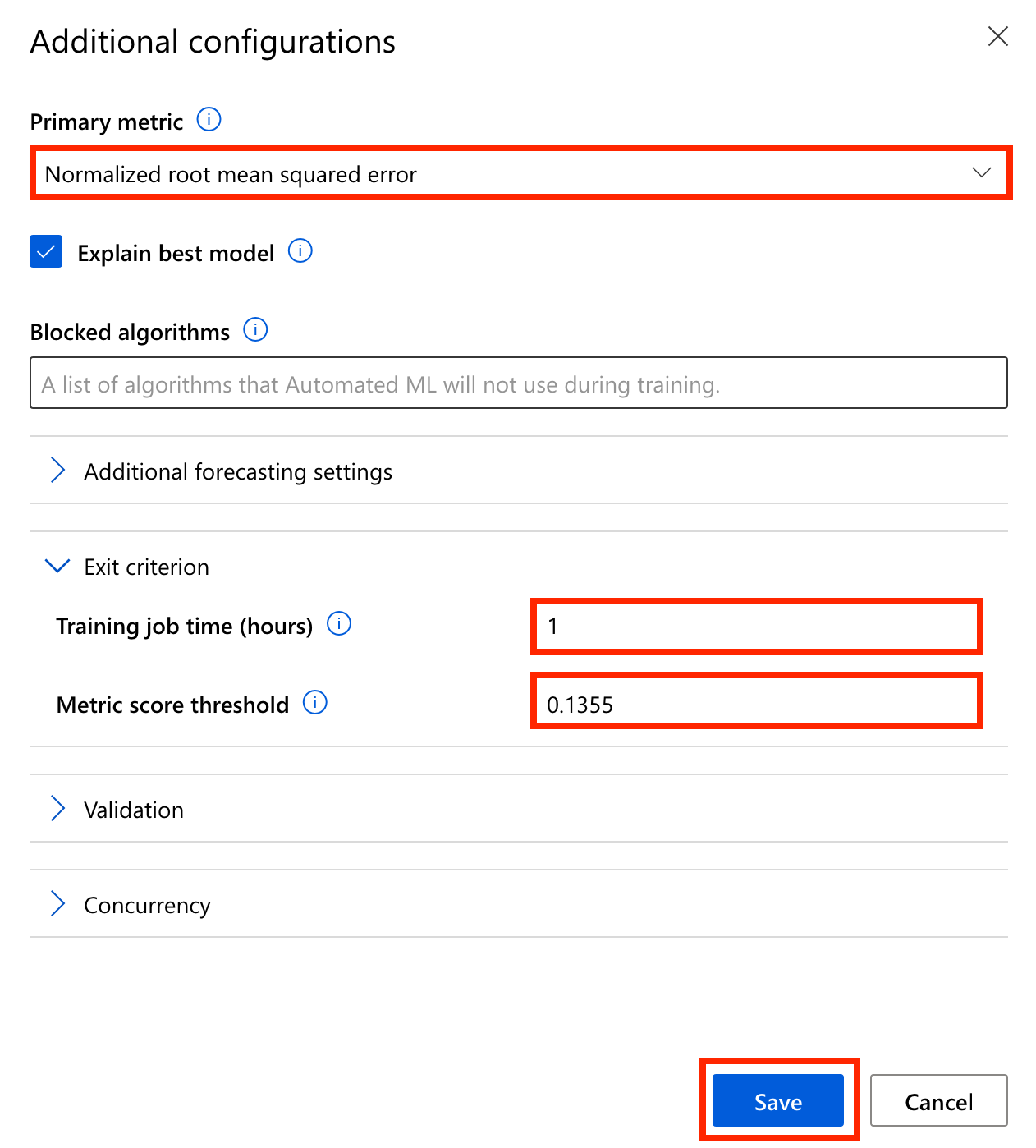
**Note**: The creation of the new compute may take several minutes. Once the process is completed, select **Next** in the Configure run section.

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-8.png)

1. Select the Time series forecasting task type and provide the following values and then select **View additional configuration settings**:
   * **Time column**: Date
   * **Time series identifier(s)**: Battery\_ID
   * **Forecast horizon**: 30

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-9.png)

1. For the automated machine learning run additional configurations, provide the following values and then select **Save**:
   * **Primary metric**: Normalized root mean squared error
   * **Explain best model**: Selected
   * **Training job time (hours)** (in the Exit criterion section): enter 1 as this is the lowest value currently accepted.
   * **Metric score threshold**: enter 0.1355. When this threshold value will be reached for an iteration metric the training job will terminate.

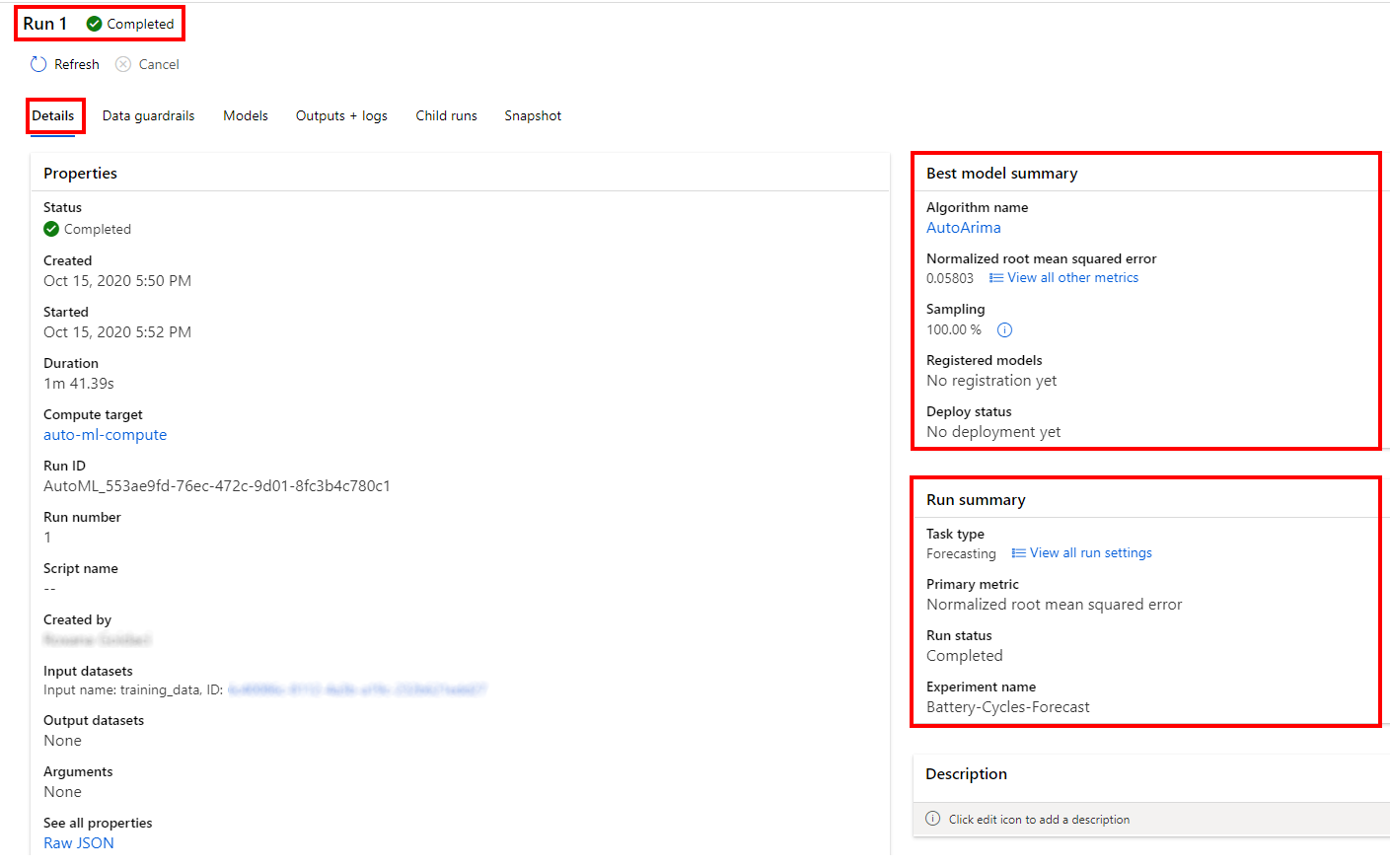
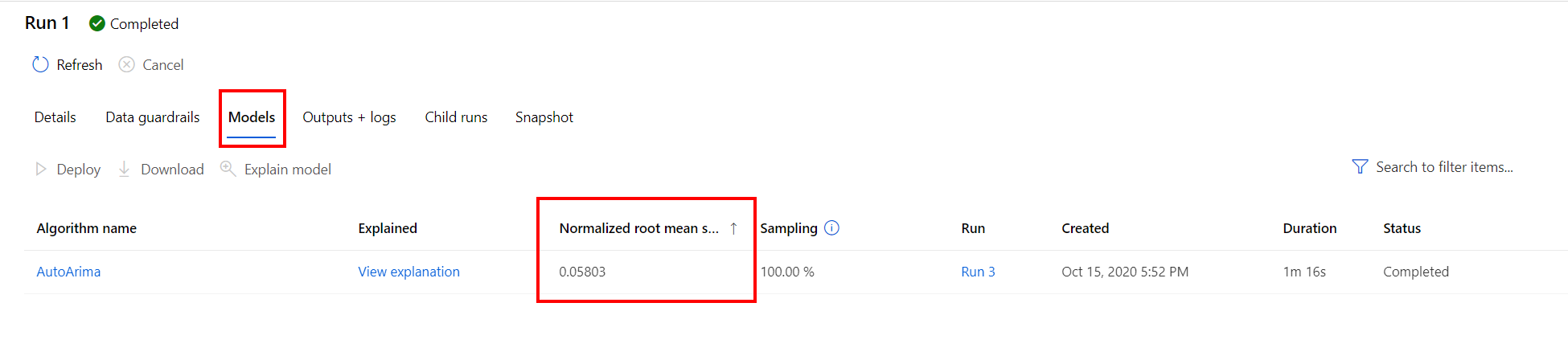
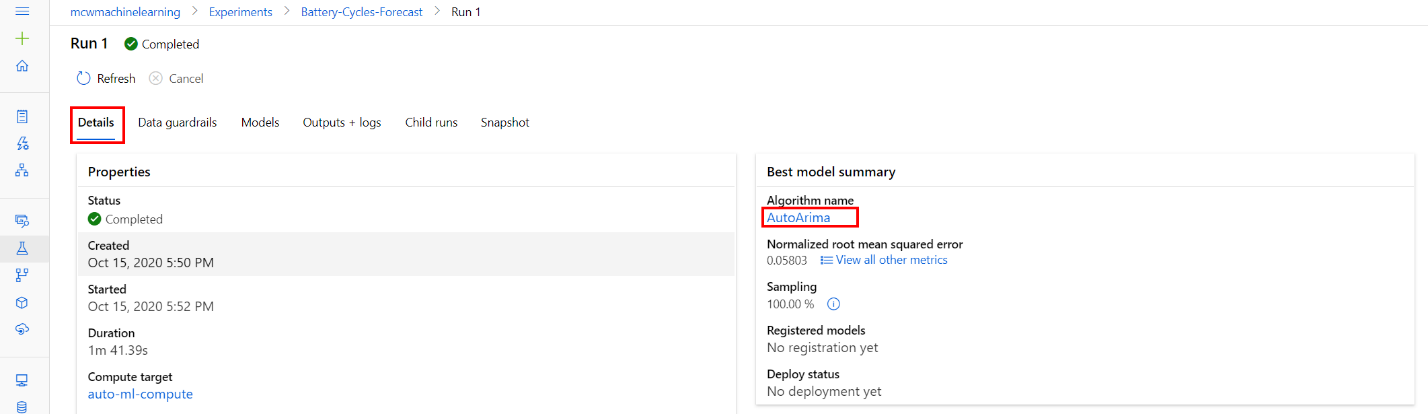
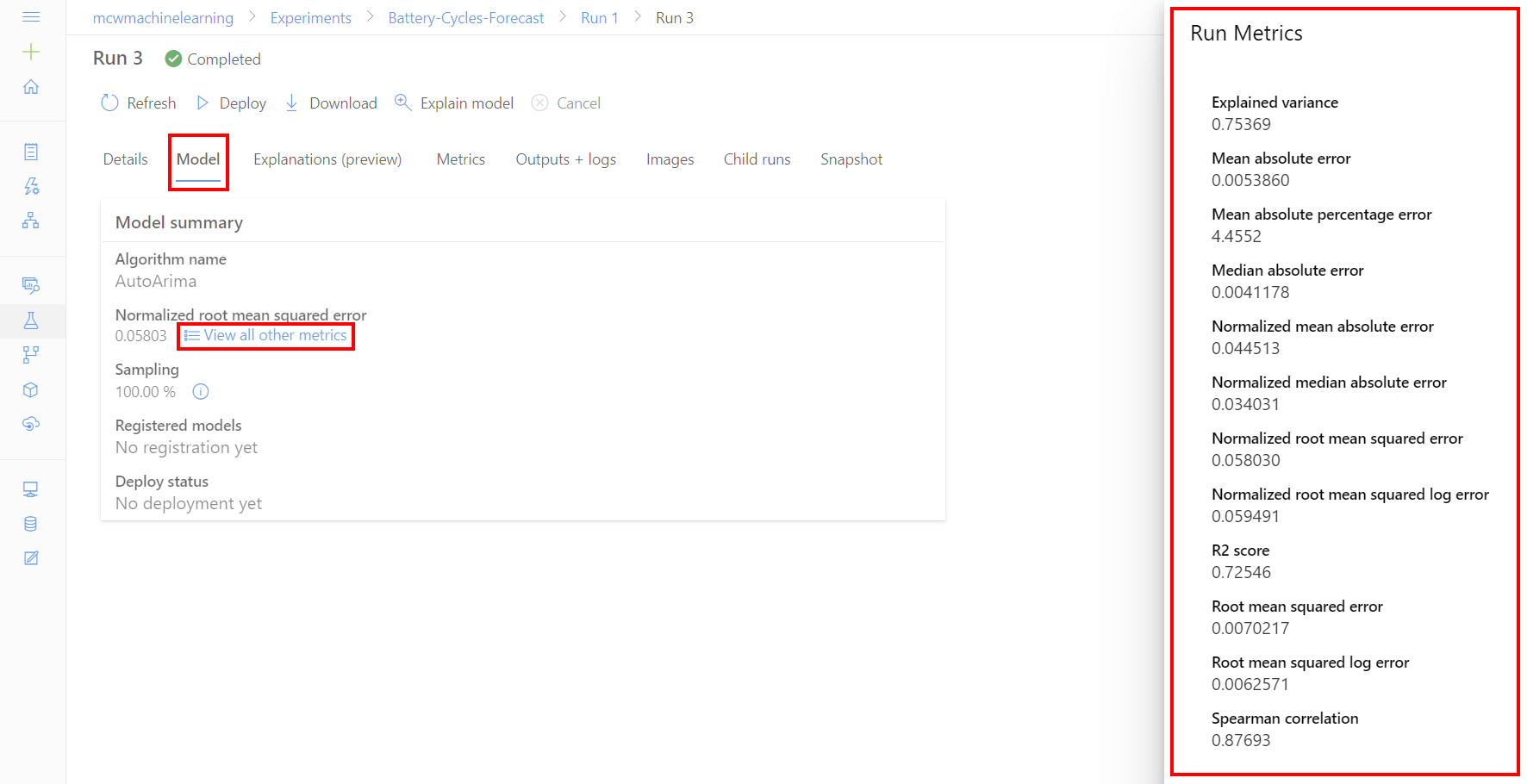
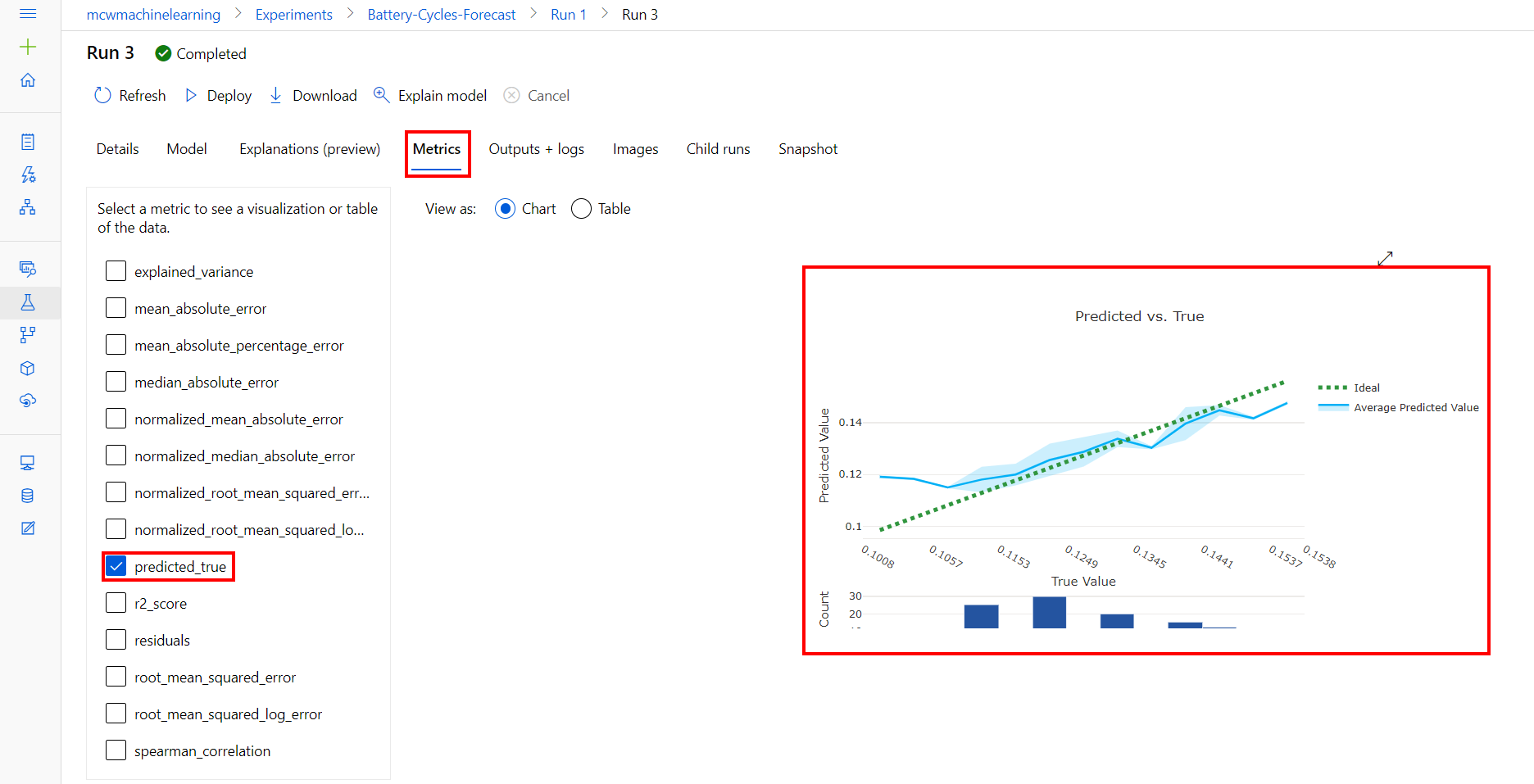
[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-10.png)

**Note**: We are setting a metric score threshold to limit the training time. In practice, for initial experiments, you will typically only set the training job time to allow AutoML to discover the best algorithm to use for your specific data.

1. Select **Finish** to start the new automated machine learning run.

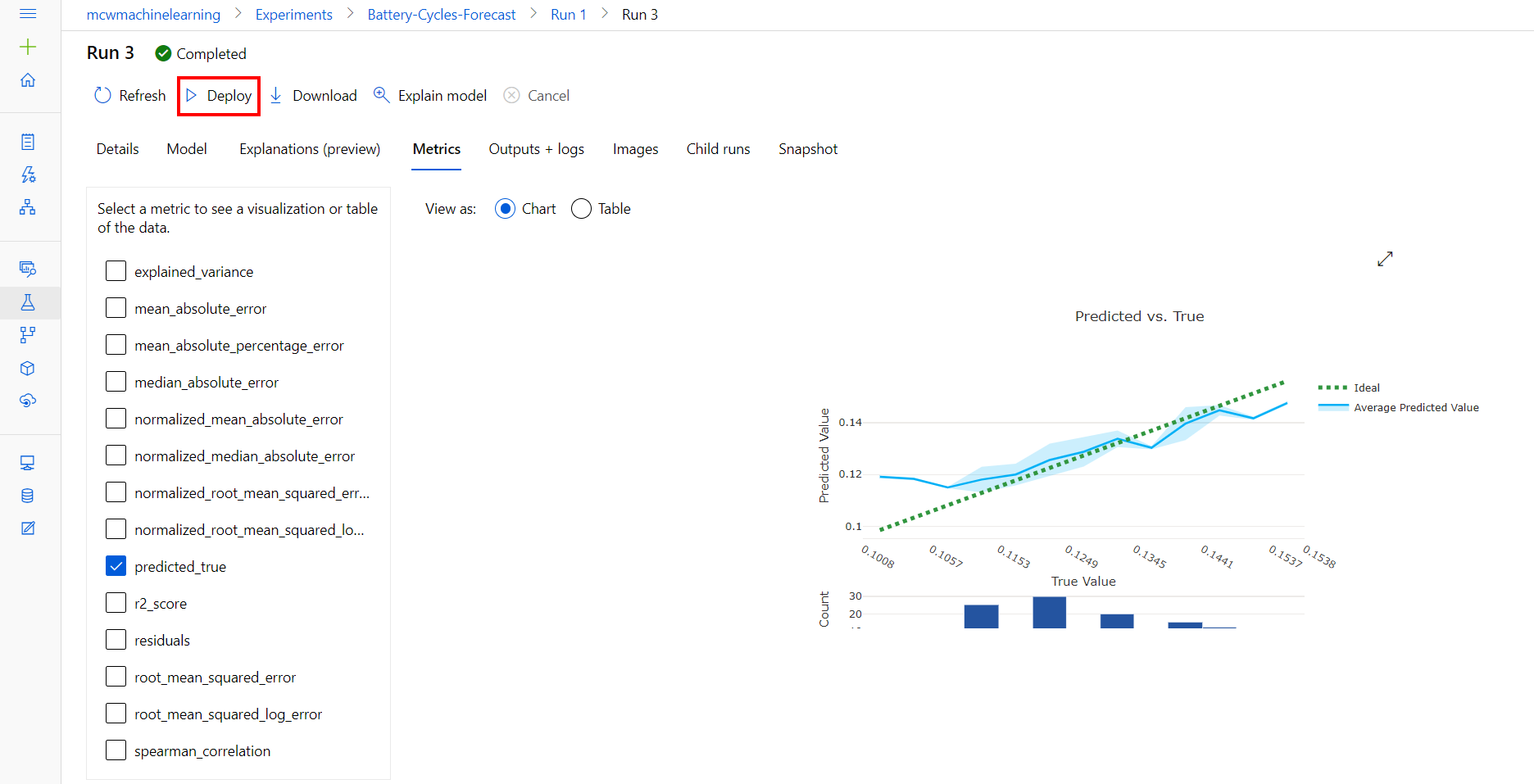
**Note**: The experiment should run for up to 10 minutes. If the run time exceeds 15 minutes, cancel the run and start a new one (steps 3, 9, 10). Make sure you provide a higher value for Metric score threshold in step 10.

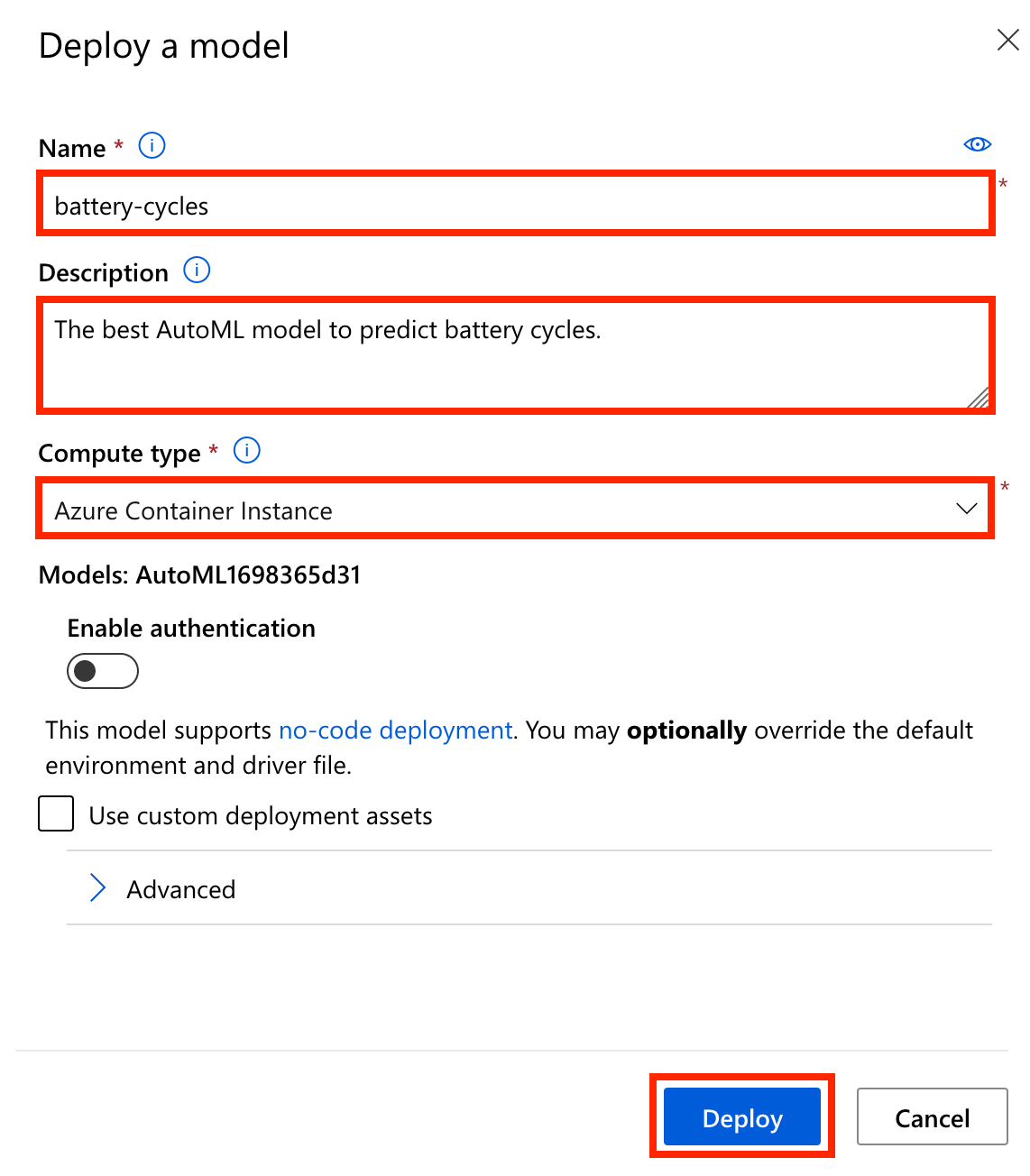
**Task 2: Review the experiment run results**

1. Once the experiment completes, select Details to examine the details of the run containing information about the best model and the run summary. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-11.png)
2. Select Models to see a table view of different iterations and the Normalized root mean squared error score for each iteration. Note that the normalized root mean square error measures the error between the predicted value and actual value. In this case, the model with the lowest normalized root mean square error is the best model. Note that Azure Machine Learning Python SDK updates over time and gives you the best performing model at the time you run the experiment. Thus, it is possible that the best model you observe can be different than the one shown below. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-12.png)
3. Return to the details of your experiment run and select the best model **Algorithm name**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-13.png)
4. From the Model tab, select **View all other metrics** to review the various Run Metrics to evaluate the model performance. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-14.png)
5. Next, select **Metrics, predicted\_true** to review the model performance curve: Predicted vs True. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-15.png)

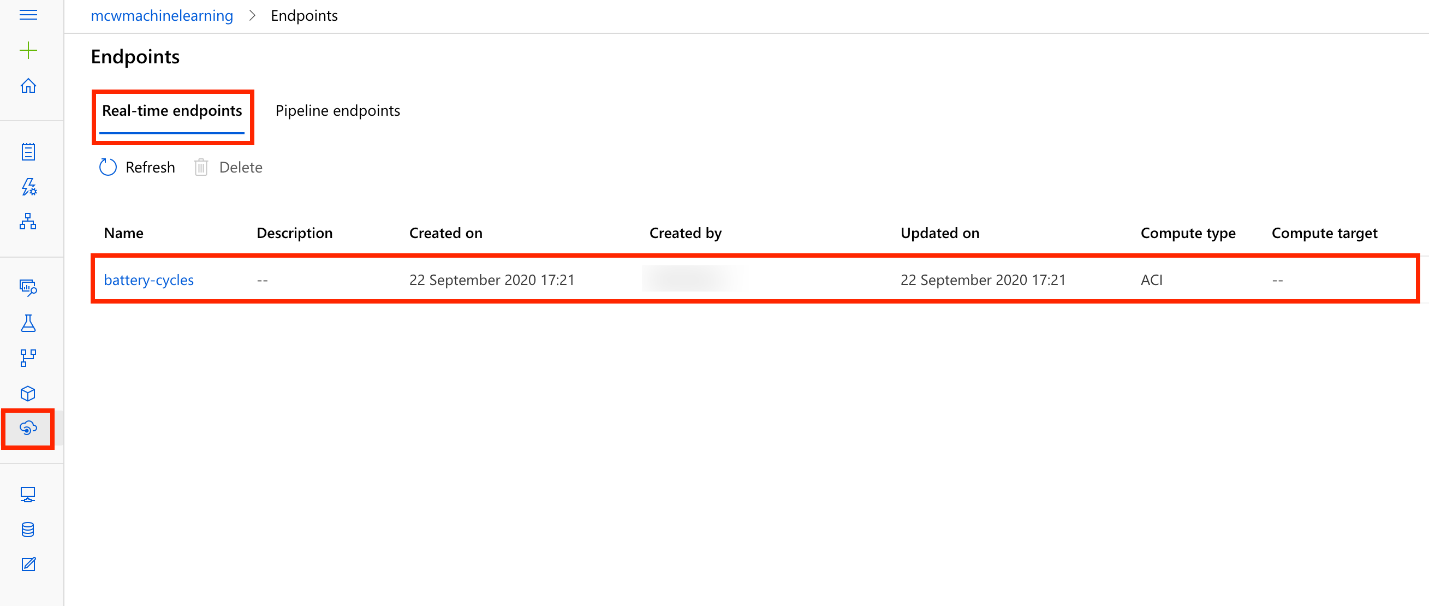
**Note**: You may need to deselect the other metrics.

**Task 3: Deploy the Best Model**

1. From the top toolbar select **Deploy**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-16.png)
2. Provide the Name, Description and Compute type, and then select **Deploy**:
   * **Name**: **battery-cycles**
   * **Description**: **The best AutoML model to predict battery cycles.**
   * **Compute type**: Select ACI.

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-17.png)

1. The model deployment process will register the model, create the deployment image, and deploy it as a scoring webservice in an Azure Container Instance (ACI). To view the deployed model, from Azure Machine Learning studio select **Endpoints icon, Real-time endpoints**.

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T3-18.png)

**Note**: The battery-cycles endpoint will show up in a matter of seconds, but the actual deployment takes several minutes. You can check the deployment state of the endpoint by selecting it and then selecting the Details tab. A successful deployment will have a state of Healthy.

1. If you see your model deployed in the above list, you are now ready to continue on to the next exercise.

**Task 4: Perform batch scoring in Azure Databricks**

1. Browse to your Azure Databricks Workspace and navigate to AI with Databricks and AML \ 2.0 Batch Scoring for Timeseries. This is the notebook you will step through executing in this exercise. Again, remember that you may need to reconnect to the **lab** cluster.
2. Follow the instructions within the notebook to complete the exercise.

**Exercise 3: Creating a deep learning model (RNN) for time series data**

Duration: 45 minutes

In this exercise, you will create a deep learning model (using a RNN - Recurrent Neural Network), and you will apply the forecast model to a Spark streaming job in order to make predictions against streaming data.

**Task 1: Create the deep learning model and start a streaming job using a notebook**

1. Browse to your Azure Databricks Workspace and navigate to AI with Databricks and AML \ 3.0 Deep Learning with Time Series. This is the notebook you will step through executing in this exercise.
2. Follow the instructions within the notebook to complete the exercise.

**Exercise 4: Creating, training and tracking a deep learning text classification model with MLflow and Azure Machine Learning**

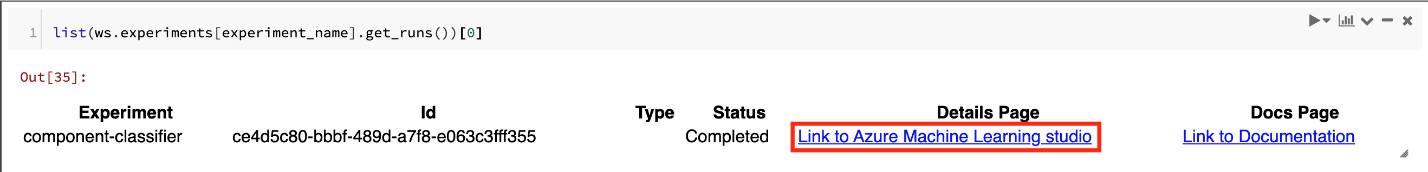
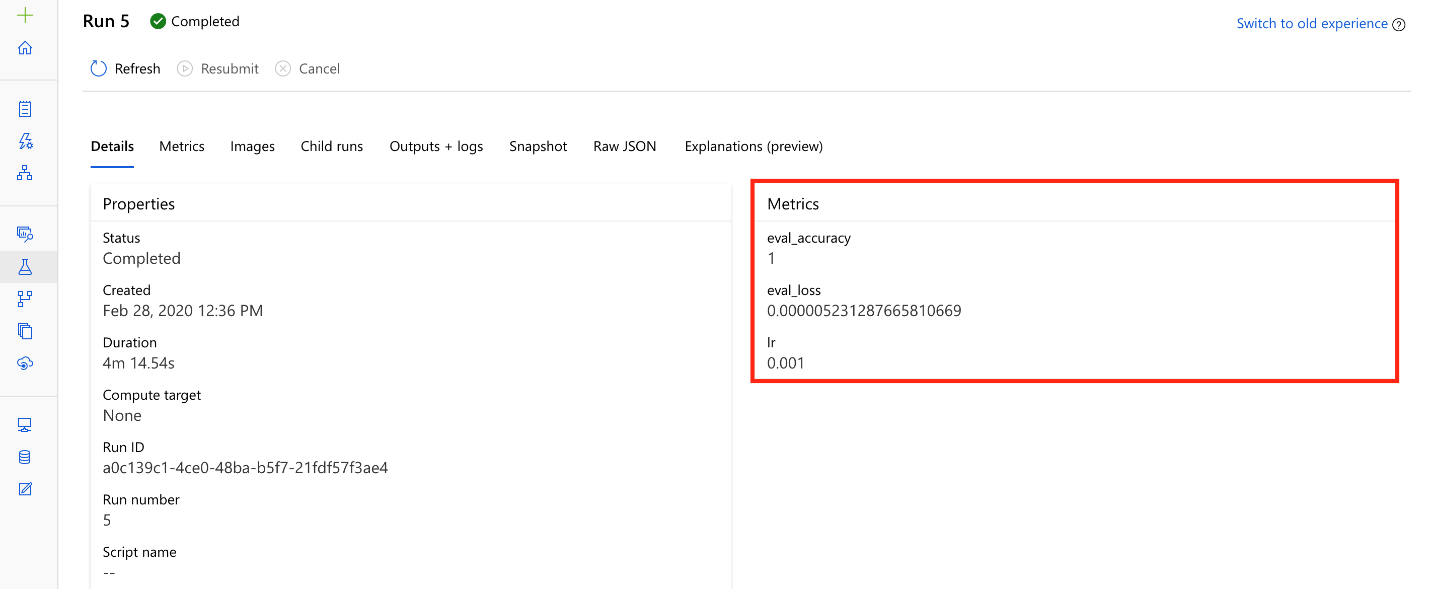
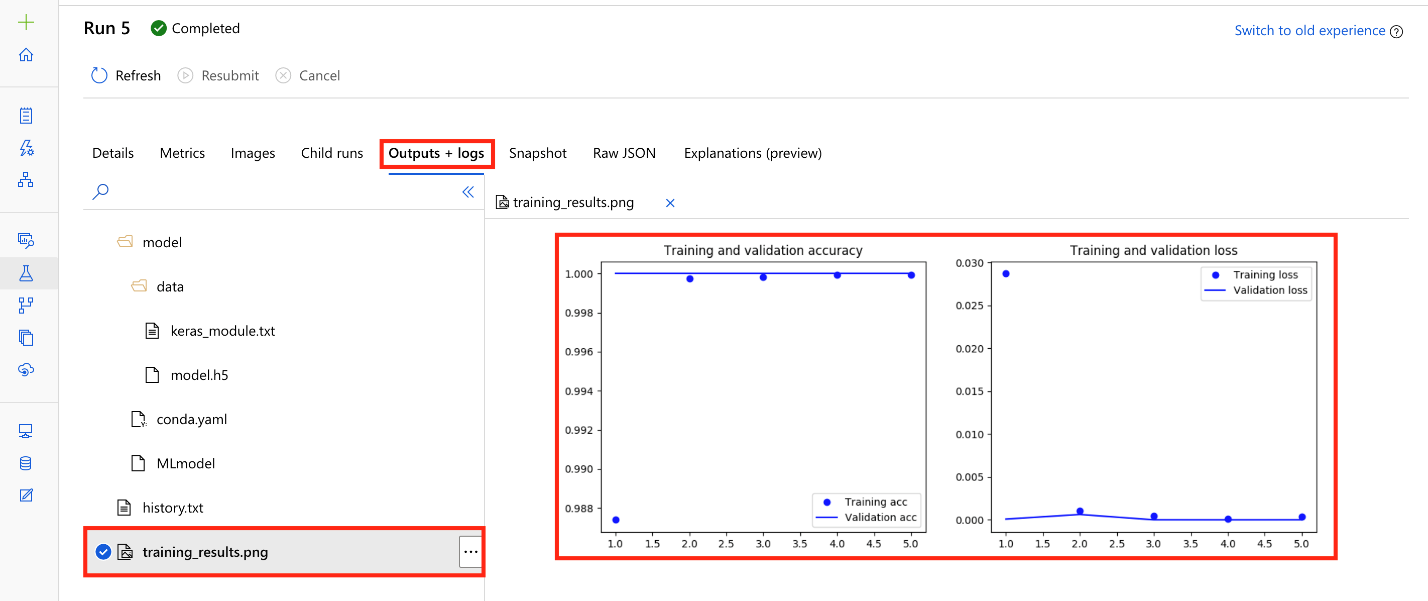
Duration: 45 minutes

In this exercise, you create a model for classifying component text as compliant or non-compliant. You will train the model Azure Machine Learning and use MLflow integration with Azure Machine Learning to track and log experiment metrics and artifacts in the Azure Machine Learning workspace.

**Task 1: Create, train and track the classification model using a notebook**

1. Browse to your Azure Databricks Workspace and navigate to AIT with Databricks and AML \ 4.0 Deep Learning with Text. This is the notebook you will step through executing in this exercise.
2. Follow the instructions within the notebook to complete exercise.

**Task 2: Review model performance metrics and training artifacts in Azure Machine Learning workspace**

1. Select the **Link to Azure Machine Learning studio** from the output of the last cell in the notebook to open the Run Details page in the Azure Machine Learning studio. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T5-1.png)
2. The **Run Details** page shows the three metrics that were logged via MLflow during the model training process: **Learning rate(lr)**, **evaluation loss(eval\_loss)**, and **evaluation accuracy (eval\_accuracy)**. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T5-2.png)
3. Next, select **Outputs + logs, tranining\_reseults.png** to review the model training artifacts logged using MLflow. In this section, you can review the curves showing both accuracy and loss as the model training progress. You can also observe that MLflow logs the trained model and the training history with Azure Machine Learning workspace. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T5-3.png)

**Exercise 5: Evaluate model interpretability**

Duration: 20 minutes

In this exercise, you will interpret the behavior of one of the models trained in previous exercises.

**Task 1: Create the deep learning model and start a streaming job using a notebook**

1. Browse to your Azure Databricks Workspace and navigate to AI with Databricks and AML \ 5.0 Model Interpretability. This is the notebook you will step through executing in this exercise.
2. Follow the instructions within the notebook to complete the exercise.

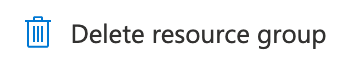
## After the hands-on lab

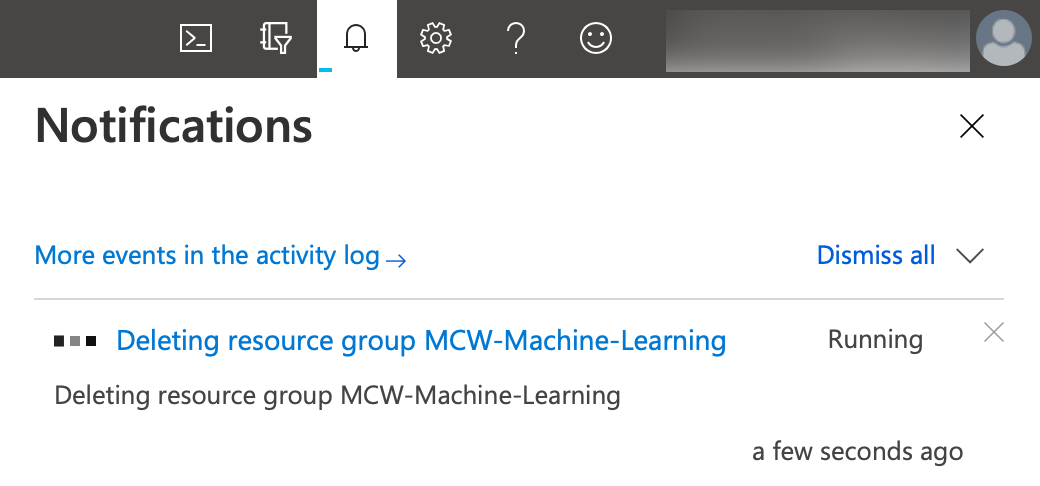
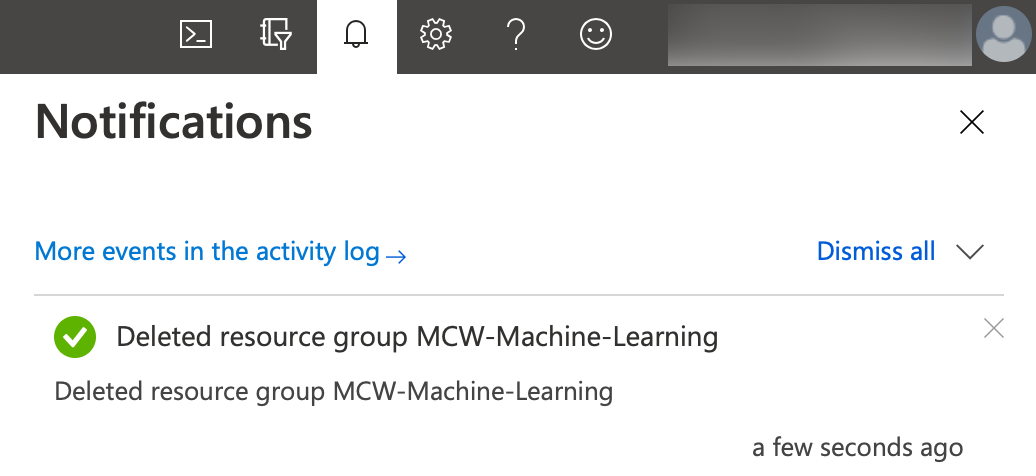
Duration: 5 minutes

To avoid unexpected charges, it is recommended that you clean up all of your lab resources when you complete the lab.

#### Task 1: Clean up lab resources

1. Navigate to the Azure Portal and locate the MCW-Machine-Learning Resources Group created for this lab.
2. Select **Delete resource group** from the command bar.

[](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T7-1.png)

1. In the confirmation dialog that appears, enter the name of the resource group and select **Delete**.
2. Wait for the confirmation that the Resource Group has been successfully deleted. If you don't wait, and the delete fails for some reason, you may be left with resources running that were not expected. You can monitor using the Notifications dialog, which is accessible from the Alarm icon. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T7-2.png)
3. When the Notification indicates success, the cleanup is complete. [](https://github.com/ceteongvanness/Cloud-Workshop-Machine-Learning/blob/main/Hands-on%20lab/images/T7-3.png)

You should follow all steps provided after attending the Hands-on lab.