

This document provides additional information for the labs.

## Exercise 1: Building a Machine Learning Model

### Task 2: Navigate to Machine Learning Studio

Use the **Chrome** browser loaded on the VM. Internet Explorer and Edge have a copy/paste issue inside of AzureML Studio.

---

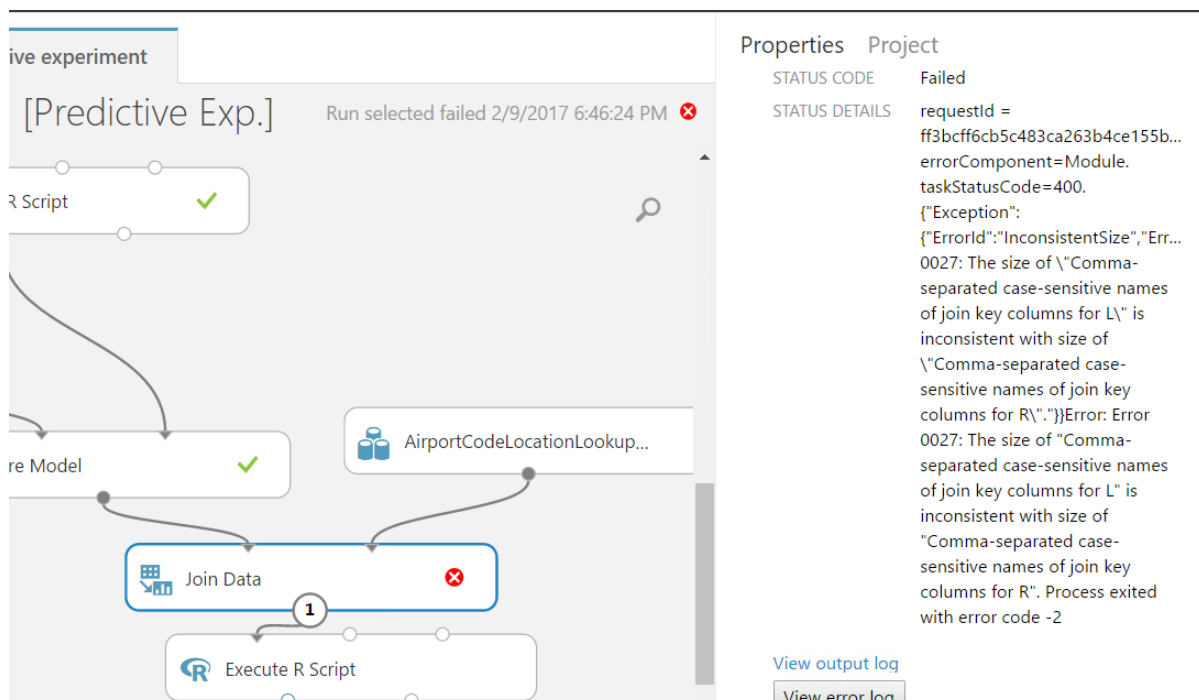
*Hint: At this point open the Cortana instructions in another window in Chrome. This will allow you to alt+tab to switch quickly between the instructions and AzureML Studio*

---

### Task 8: Operationalize the Experiment

Step 15. When setting the join columns, the columns list may not appear automatically. You may begin typing OriginAirportCode in the list box OR select "All Columns" under "Begin With". Then select the OriginAirportCode. Do not forget to set "Begin With" back to No Columns.

If you get the error below, it is because you have a different number of join columns between the left and right sides of the join. (You probably forgot to go back and select Begin With No Columns on the Left.)

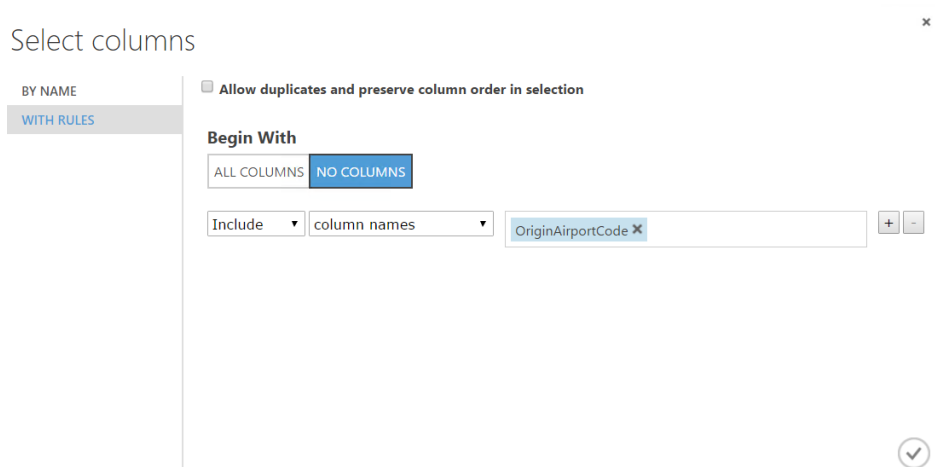


The screenshot shows the AzureML Studio interface. On the left, a workflow diagram for a predictive experiment is visible. It includes steps like 'R Script', 'Train Model', 'Join Data', and 'Execute R Script'. The 'Join Data' step is highlighted with a red 'X' and a circled '1', indicating an error. On the right, the 'Properties' pane shows the 'Project' tab with 'STATUS CODE' as 'Failed'. The 'STATUS DETAILS' section contains the following error message:

```
requestId = ff3bcff6cb5c483ca263b4ce155b...
errorComponent=Module.
taskStatusCode=400.
{"Exception":
{"ErrorId":"InconsistentSize","Err...
0027: The size of \"Comma-separated case-sensitive names of join key columns for L\" is inconsistent with size of \"Comma-separated case-sensitive names of join key columns for R\".}}Error: Error 0027: The size of \"Comma-separated case-sensitive names of join key columns for L\" is inconsistent with size of \"Comma-separated case-sensitive names of join key columns for R\". Process exited with error code -2
```

Below the error message are links for 'View output log' and 'View error log'.

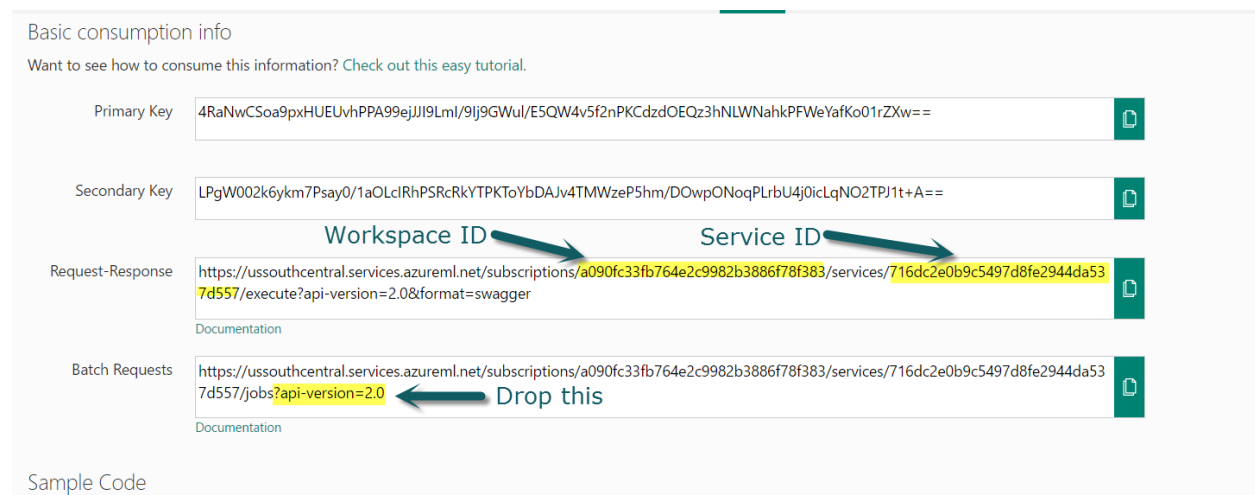
**Figure: Join Data Error**



**Figure: Select Begin With No Columns**

## Task 9: Deploy API and Note Web Service Information

If the keys are blank, just refresh the page.



Basic consumption info

Want to see how to consume this information? Check out this easy tutorial.

Primary Key: 4RaNwCSoa9pxHUEUvhPPA99ejJJl9Lml/9lj9GWul/E5QW4v5f2nPKCdzdOEQz3hNLWNahkPFWeYafKo01rZXw==

Secondary Key: LPgW002k6ykm7Psay0/1aOLclRhPSRcRkYTPKToYbDAJv4TMWzeP5hm/DOWpONoqPLrbU4j0icLqNO2TPJ1t+A==

Request-Response: <https://ussouthcentral.services.azureml.net/subscriptions/a090fc33fb764e2c9982b3886f78f383/services/716dc2e0b9c5497d8fe2944da537d557/execute?api-version=2.0&format=swagger>

Documentation

Batch Requests: <https://ussouthcentral.services.azureml.net/subscriptions/a090fc33fb764e2c9982b3886f78f383/services/716dc2e0b9c5497d8fe2944da537d557/jobs?api-version=2.0>

Documentation

Sample Code

**Figure: Identifying the keys**

Then Save your notepad.

## Exercise 2: Setup Azure Data Factory

### Task 2: Download and Stage Data to be Processed

When you click on the link, the file is placed in your This PC ->Downloads directory. Create the C:\Data folder and copy the file to that folder.

### Task 4: Log in to the Azure Portal

Make sure you do this inside the VM.

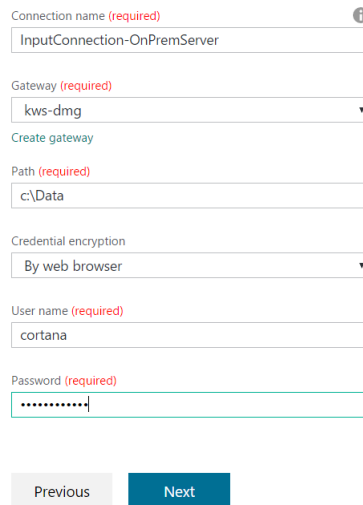
## Exercise 3: Develop Data Factory Pipeline for Data Movement

### Task 1: Create Copy Pipeline Using the Copy Data Wizard

Step 7. From the Source Data Store screen tab, select **File Server Share File System**.

Step 13. For the Gateway, it should already be filled in with the gateway connection you created in the previous exercise.

#### Specify File server share connection



Connection name (required) ⓘ  
 InputConnection-OnPremServer

Gateway (required)  
 kws-dmg ▼  
 Create gateway

Path (required)  
 c:\Data

Credential encryption  
 By web browser ▼

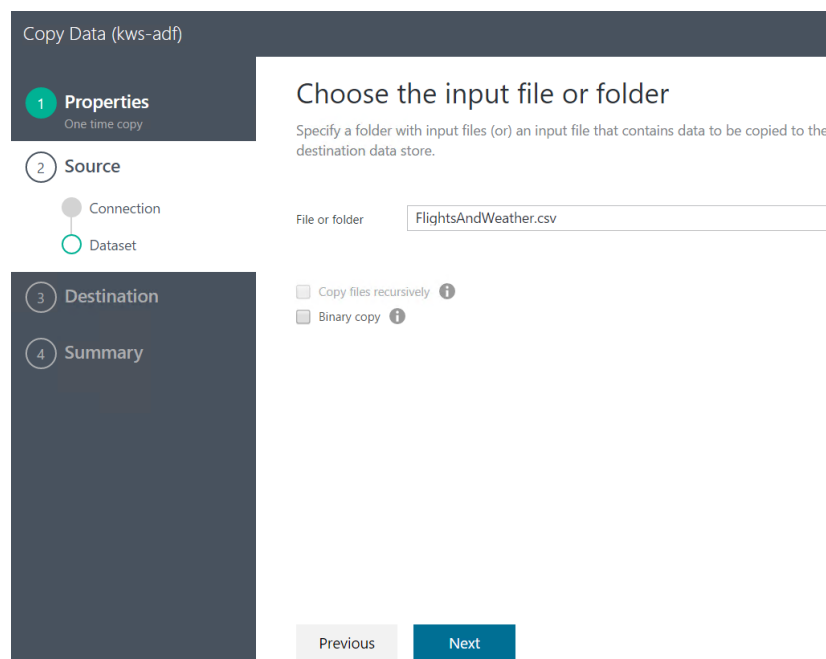
User name (required)  
 cortana

Password (required)  
 .....

Previous Next

**Figure: Specify File Server Share Connection**

17a. Click on the **Next** button from the bottom of the screen.



Copy Data (kws-adf)

1 Properties  
 One time copy

2 Source  
 Connection  
 Dataset

3 Destination

4 Summary

Choose the input file or folder

Specify a folder with input files (or) an input file that contains data to be copied to the destination data store.

File or folder  
 FlightsAndWeather.csv

☐ Copy files recursively ⓘ

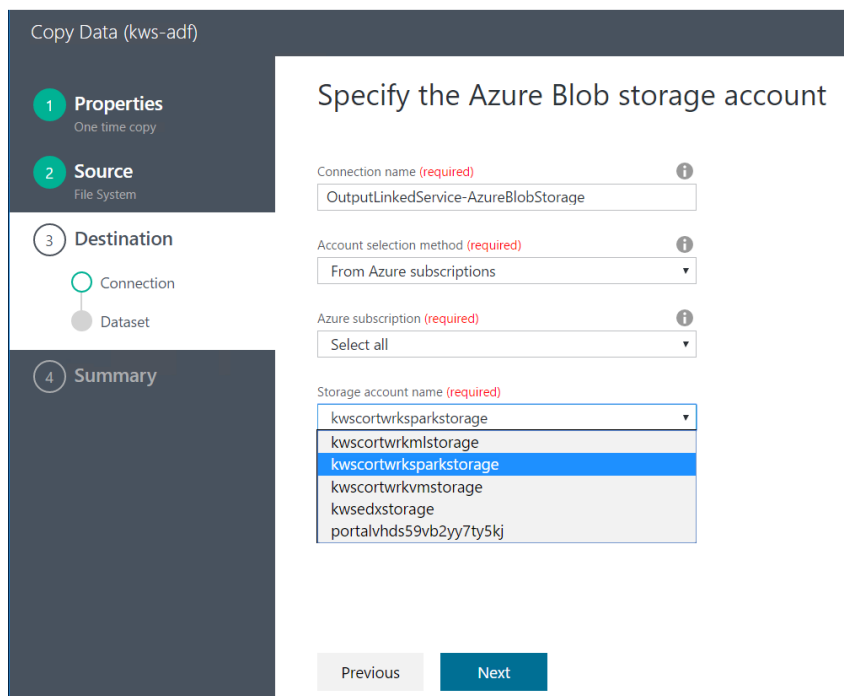
☐ Binary copy ⓘ

Previous Next

**Figure: Choose Input File**

Step 18. From the **File format settings** tab, leave everything as default. ~~except check the box~~ **Column name in the first data row**. **Column Name** in the first data row should be checked, which is the default. You can see the preview of the file from the bottom of the screen.

Step 24. Before clicking **Next** at the bottom of the screen, *please make sure you have selected the right storage account* (it will have the **sparkstorage** suffix). Finally, click **Next**.

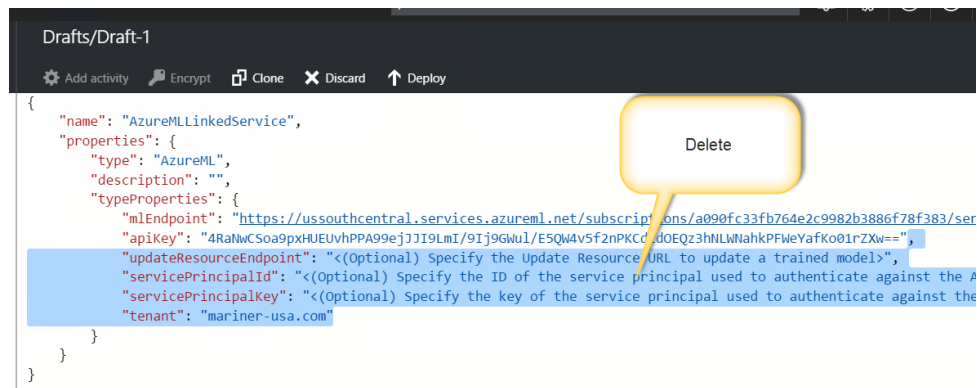


**Figure: Choose Azure Blob Storage Account**

## Exercise 4: Operationalize ML Scoring with Azure ML and Data Factory

### Task 1: Create Azure ML Linked Service

You will replace 2 items with Batch Key and Primary Key as the lab instructions indicate. Then you will delete the remaining items as highlighted in the picture below.



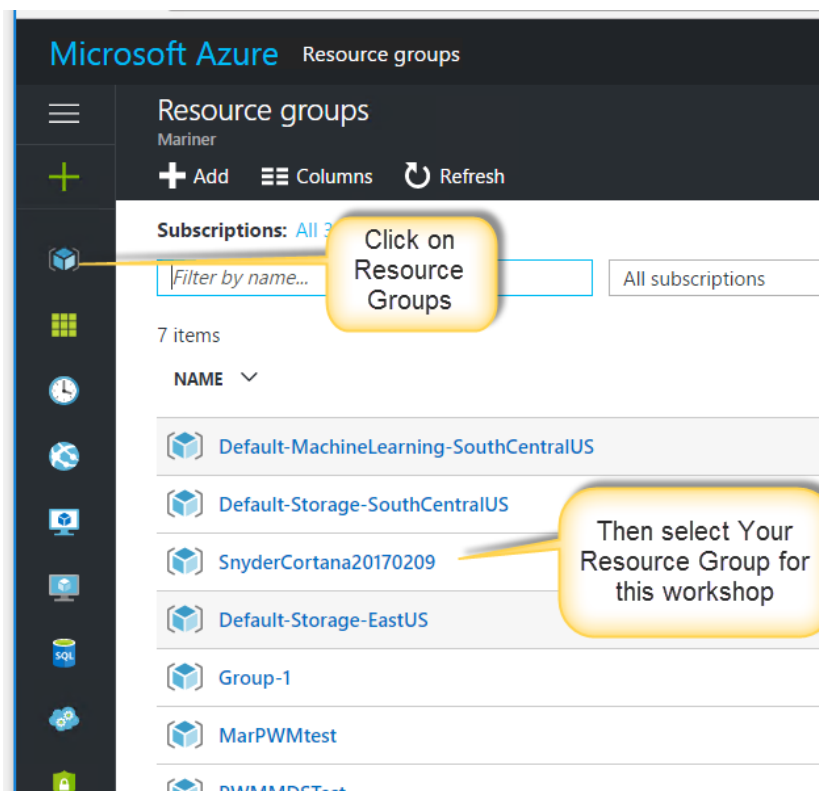
**Figure: Delete items on Default Linked Service**

## Exercise 5: Summarize Data Using HDInsight Spark

### Task 1: Summarize Delays by Airport

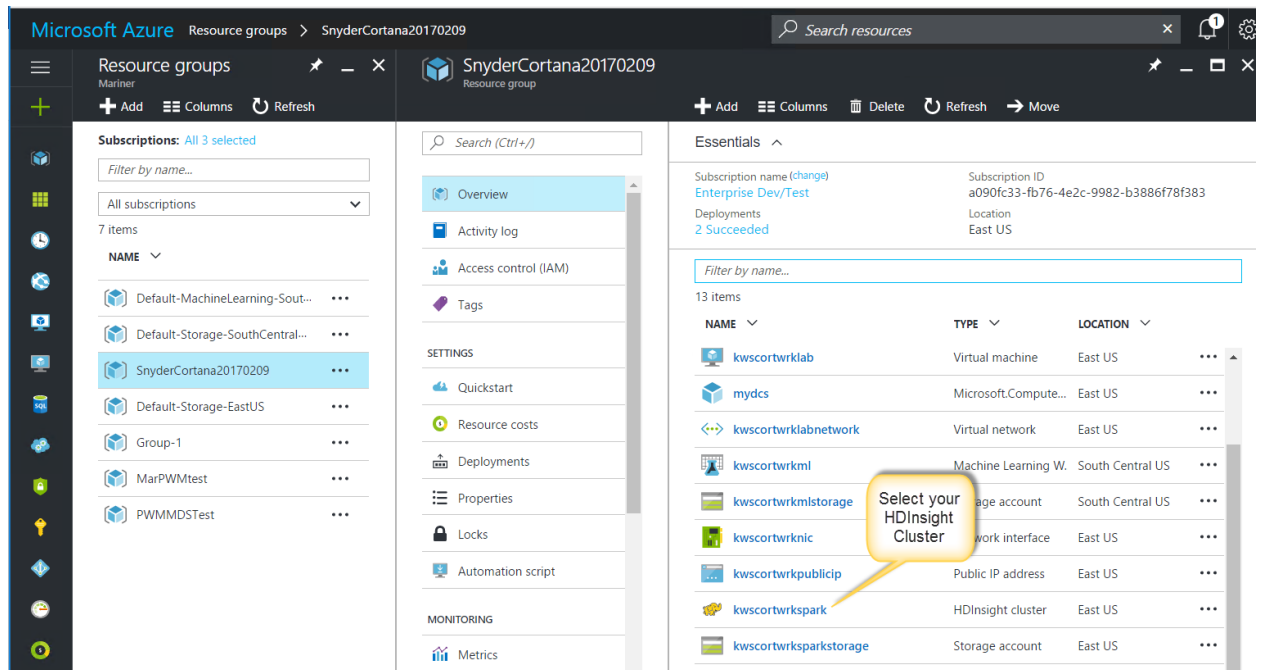
To Navigate to your HDInsight Cluster

Step1. Select Resource Groups, then select your resource group for this workshop.



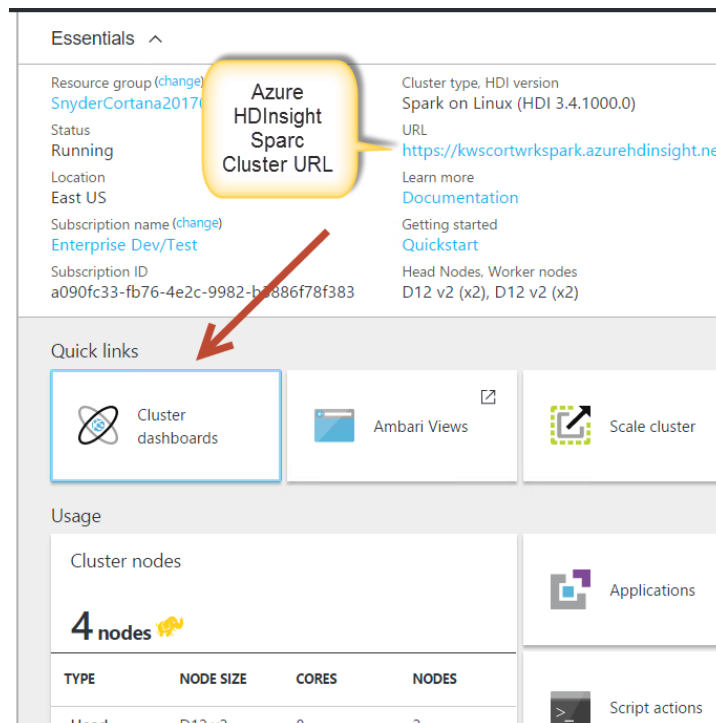
**Figure: Finding your HDInsight Cluster**

Step 2. Then select your HDInsight Cluster



**Figure: Select your HDInsight Cluster**

Step 3. NOTE the URL for your Azure HDInsight Sparc Cluster in Yellow below. Now you can select the Cluster Dashboards, in red.

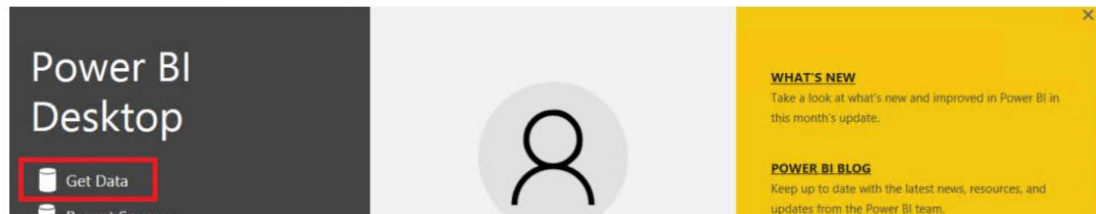


**Figure: Saving the cluster URL and navigating to Cluster Dashboards**

## Exercise 6: Visualizing in Power BI Desktop

### Task 2: Connect to the HDInsight Spark Using Power BI Desktop

1. Launch Power BI Desktop using the shortcut on the Desktop of the Lab VM.
2. Click on **Get Data** from the left side of the welcome window.

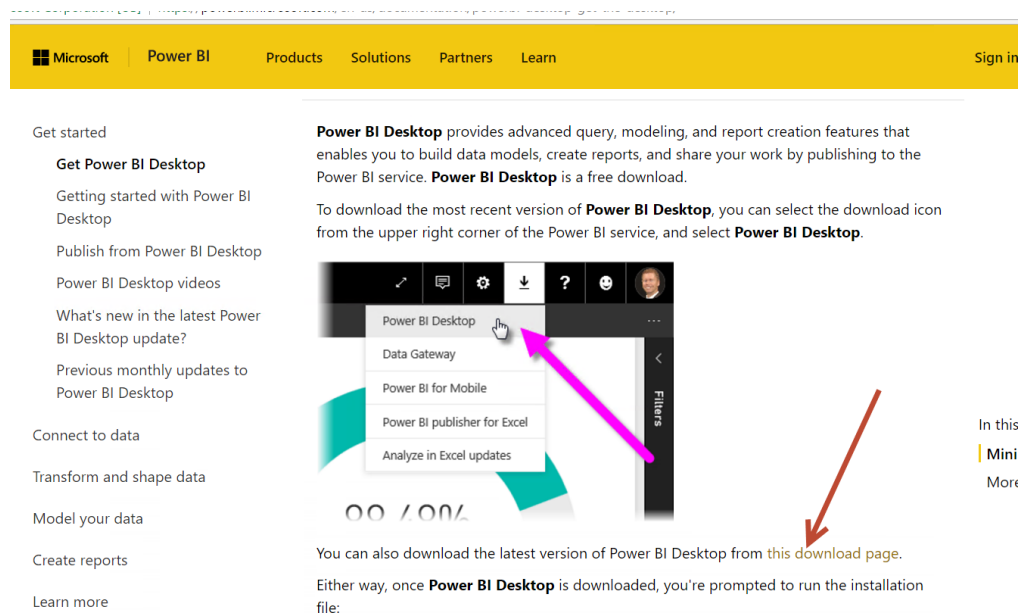


**Figure: Power BI Desktop must be installed on VM**

Power BI Desktop must be installed on the VM. For instructions, go to:

<https://powerbi.microsoft.com/en-us/documentation/powerbi-desktop-get-the-desktop/>

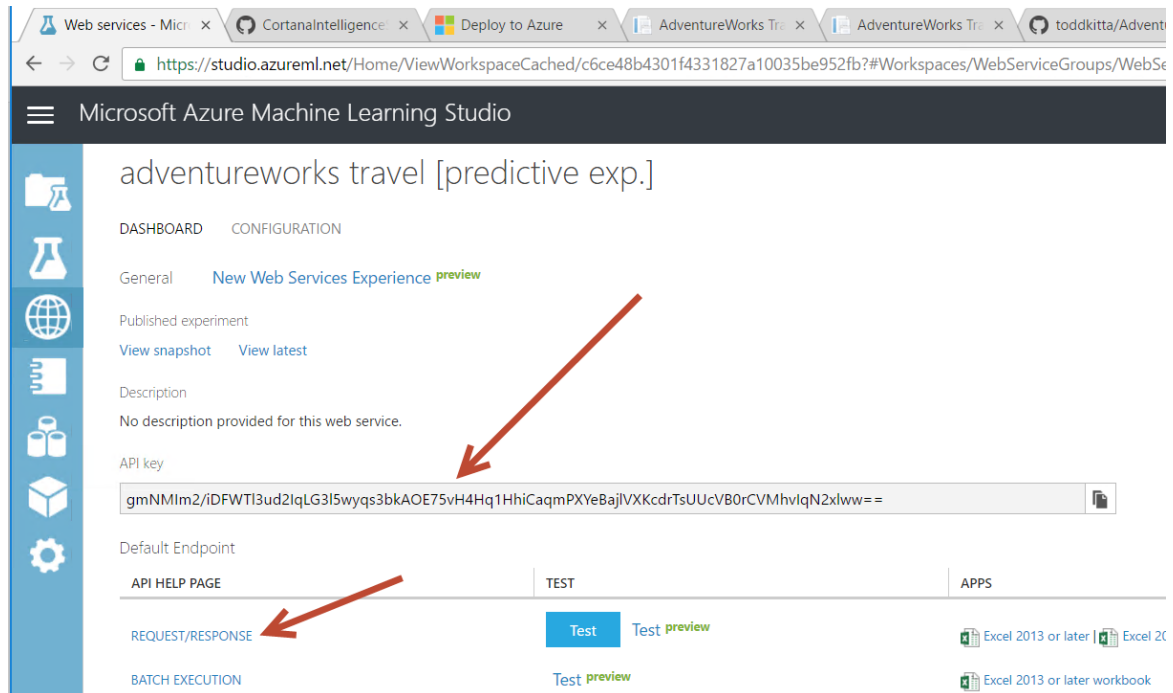
The download link is indicated by the RED arrow.



**Figure: Installing Power BI Desktop**

## Exercise 7: Deploy Intelligent Web App

The keys needed are the weather underground key, workspace id, service id, and what is referred to now, as the ML API key. In an earlier lab you saved this key, but then it was referred to as the machine learning web service api key. You may fetch that key by opening going to Studio.azureml.net in the web browser and opening your deployed web service.



**Figure: Finding the keys for deployment**

To retrieve the workspace and service id, click on the Request/Response link.

In the middle of the page you will see a request URI, from which you may capture the workspace and Service ID.

| Request |   |                                  |                                  |              |
|---------|---|----------------------------------|----------------------------------|--------------|
| Method  | Request URI   | Workspace id                     | Service ID                       | HTTP Version |
| POST    | https://ussouthcentral.services.azureml.net/workspaces/c6ce48b4301f4331827a10035be952fb/services/fde3498c41a8475db7a854a36c247788/exe | c6ce48b4301f4331827a10035be952fb | fde3498c41a8475db7a854a36c247788 | HTTP/1.1     |

**Figure: Finding the Workspace ID and Service ID keys**