

ADD NOTE TO SETUP about name uniqueness.

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.

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, 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. (You probably forgot to go back and Begin With No Columns on the Left.)

The screenshot displays the AzureML Studio interface. On the left, a workflow diagram for an experiment named "[Predictive Exp.]" is shown. The workflow includes the following steps: "R Script" (marked with a green check), "Train Model" (marked with a green check), "Join Data" (marked with a red X and a circled '1'), and "Execute R Script". A data asset named "AirportCodeLocationLookup..." is connected to the "Join Data" step. Above the workflow, a status bar indicates "Run selected failed 2/9/2017 6:46:24 PM" with a red X icon. On the right side, the "Properties" pane is open, showing the "Project" tab. Under "STATUS CODE", it says "Failed". Under "STATUS DETAILS", it provides the following error information:

```
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\"."}}  
  }  
}
```

At the bottom of the Properties pane, there are two links: "View output log" and "View error log".

Select columns

BY NAME

WITH RULES

☐ Allow duplicates and preserve column order in selection

Begin With

ALL COLUMNS

NO COLUMNS

Include

column names

OriginAirportCode

+

-



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

4RaNwCSoa9pxHUEUvhPPA99ejJI9Lml/9lj9GWul/E5QW4v5f2nPKCdzdOEQz3hNLWNahkPFWeYafKo01rZXw==



Secondary Key

LPgW002k6ykm7Psay0/1aOLclRhPSRcRkYTPKToYbDAJv4TMWzeP5hm/DOWpONoqPLrbU4j0icLqNO2TPJ1t+A==



Workspace ID

Service ID

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](#)

Drop this

Sample Code

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.

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

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

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

☐ Copy files recursively i

☐ Binary copy i

Previous Next

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.

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**

Copy Data (kws-adf)

1 Properties

One time copy

2 Source

File System

3 Destination

Connection

Dataset

4 Summary

Specify the Azure Blob storage account

Connection name (required)

OutputLinkedService-AzureBlobStorage

Account selection method (required)

From Azure subscriptions

Azure subscription (required)

Select all

Storage account name (required)

kWSCortwrksparkstorage

kWSCortwrkmlstorage

kWSCortwrksparkstorage

kWSCortwrkvmstorage

kWsedxstorage

portalvhds59vb2yy7ty5kj

Previous

Next

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.

Drafts/Draft-1

Add activity
Encrypt
Clone
Discard
Deploy

```

{
  "name": "AzureMLLinkedService",
  "properties": {
    "type": "AzureML",
    "description": "",
    "typeProperties": {
      "mlEndpoint": "https://ussouthcentral.services.azureml.net/subscriptions/a090fc33fb764e2c9982b3886f78f383/ser",
      "apiKey": "4Ra1wCSoa9pxHUEUvhPPA99ejJJi9LmI/9Ij9Gwul/E5QW4v5f2nPKcd0d0EQz3hNLWahkPFWeYafko01rZXw==",
      "updateResourceEndpoint": "<(Optional) Specify the Update Resource URL to update a trained model>",
      "servicePrincipalId": "<(Optional) Specify the ID of the service principal used to authenticate against the A",
      "servicePrincipalKey": "<(Optional) Specify the key of the service principal used to authenticate against the",
      "tenant": "mariner-usa.com"
    }
  }
}

```

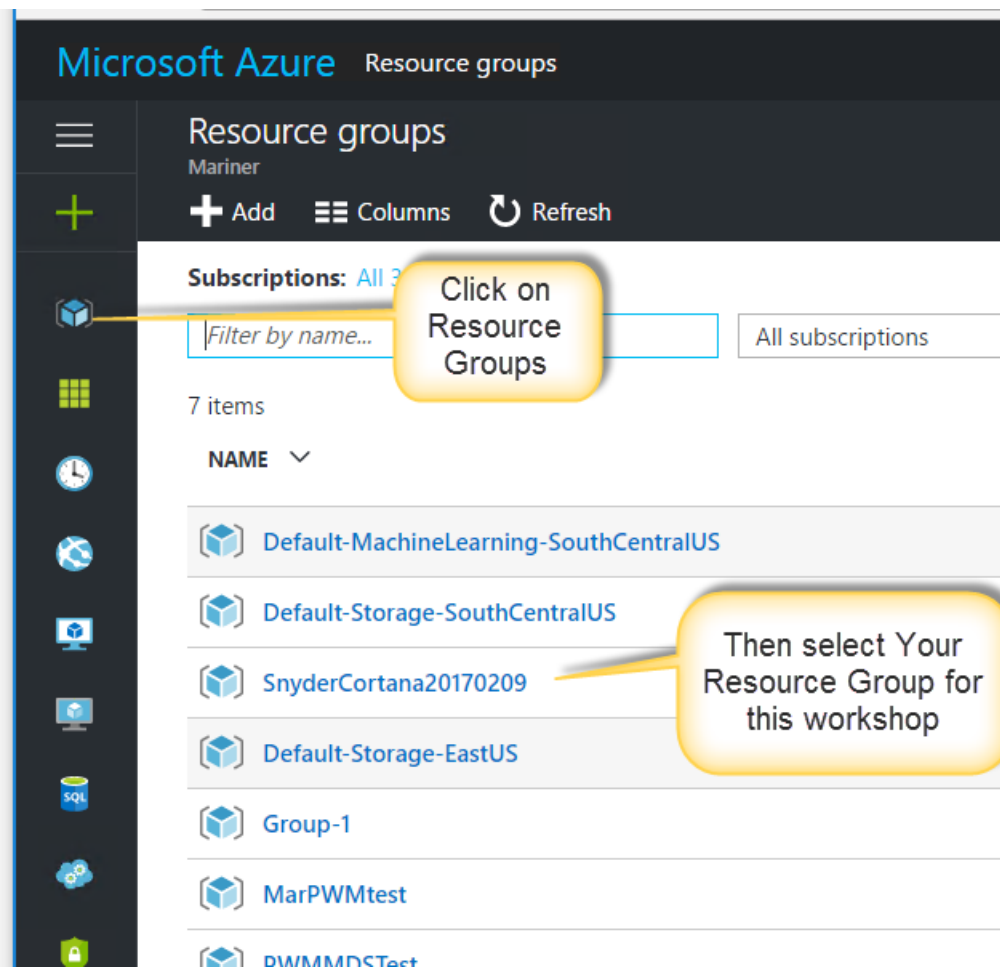
Delete

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.



Step 2: Then select your HDInsight Cluster

Microsoft Azure Resource groups > SnyderCortana20170209

Search resources

Resource groups

Mariner

+ Add Columns Refresh

Subscriptions: All 3 selected

Filter by name...

All subscriptions

7 items

NAME

- Default-MachineLearning-Sout...
- Default-Storage-SouthCentral...
- SnyderCortana20170209
- Default-Storage-EastUS
- Group-1
- MarPWMtest
- PWMMDSTest

Overview

Activity log

Access control (IAM)

Tags

SETTINGS

- Quickstart
- Resource costs
- Deployments
- Properties
- Locks
- Automation script

MONITORING

Metrics

Essentials

Subscription name (change)
Enterprise Dev/Test

Subscription ID
a090fc33-fb76-4e2c-9982-b3886f78f383

Deployments
2 Succeeded

Location
East US

Filter by name...

13 items

NAME	TYPE	LOCATION
kwscortwrklab	Virtual machine	East US
mydcs	Microsoft.Compute...	East US
kwscortwrklabnetwork	Virtual network	East US
kwscortwrkml	Machine Learning W.	South Central US
kwscortwrkmlstorage	Storage account	South Central US
kwscortwrknic	Network interface	East US
kwscortwrkpublicip	Public IP address	East US
kwscortwrkspark	HDInsight cluster	East US
kwscortwrksparkstorage	Storage account	East US

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 .

Essentials ^

Resource group (change)
[SnyderCortana20170](#)

Status
Running

Location
East US

Subscription name (change)
[Enterprise Dev/Test](#)

Subscription ID
a090fc33-fb76-4e2c-9982-b3886f78f383

Cluster type, HDI version
Spark on Linux (HDI 3.4.1000.0)

URL
<https://kwscortwrkspark.azurehdinsight.net>

Learn more
[Documentation](#)

Getting started
[Quickstart](#)

Head Nodes, Worker nodes
D12 v2 (x2), D12 v2 (x2)

Azure HDInsight Spark Cluster URL

Quick links

Cluster dashboards

Ambari Views

Scale cluster

Usage

Cluster nodes

4 nodes

TYPE	NODE SIZE	CORES	NODES
Head	D12 v2	8	2

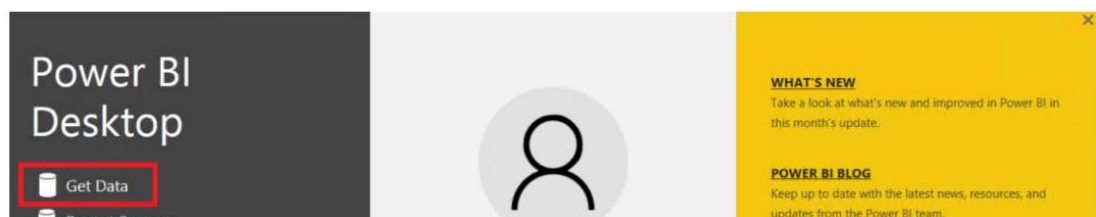
Applications

Script actions

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.



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

Microsoft

Power BI

Products

Solutions

Partners

Learn

Sign in

Get started

Get Power BI Desktop

Getting started with Power BI Desktop

Publish from Power BI Desktop

Power BI Desktop videos

What's new in the latest Power BI Desktop update?

Previous monthly updates to Power BI Desktop

Connect to data

Transform and shape data

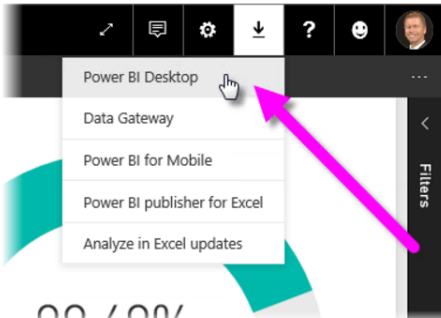
Model your data

Create reports

Learn more

Power BI Desktop provides advanced query, modeling, and report creation features that enables you to build data models, create reports, and share your work by publishing to the Power BI service. **Power BI Desktop** is a free download.

To download the most recent version of **Power BI Desktop**, you can select the download icon from the upper right corner of the Power BI service, and select **Power BI Desktop**.



You can also download the latest version of Power BI Desktop from [this download page](#).

Either way, once **Power BI Desktop** is downloaded, you're prompted to run the installation file:

In this

Mini

More

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 Studio.azureml.net, opening your deployed web service, and the key is visible.

Web services - Micro x CortanaIntelligence x Deploy to Azure x AdventureWorks Tr x AdventureWorks Tr x toddkitta/Advent

https://studio.azureml.net/Home/ViewWorkspaceCached/c6ce48b4301f4331827a10035be952fb?#Workspaces/WebServiceGroups/WebSe

Microsoft Azure Machine Learning Studio

adventureworks travel [predictive exp.]

DASHBOARD CONFIGURATION

General **New Web Services Experience** preview

Published experiment

[View snapshot](#) [View latest](#)

Description

No description provided for this web service.

API key

gmNMIm2/iDFWTl3ud2lqLG3l5wyqs3bkAOE75vH4Hq1HhiCaqmPXyeBajlVXXKcdrTsUUcVB0rCVMhvlqN2xlww==

Default Endpoint

API HELP PAGE	TEST	APPS
REQUEST/RESPONSE	Test <small>Test preview</small>	Excel 2013 or later Excel 20
BATCH EXECUTION	Test <small>Test preview</small>	Excel 2013 or later workbook

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 cute?api-version=2.0&details=true			HTTP/1.1