Azure Databricks workshop – TO DOs

# Validate access to Azure

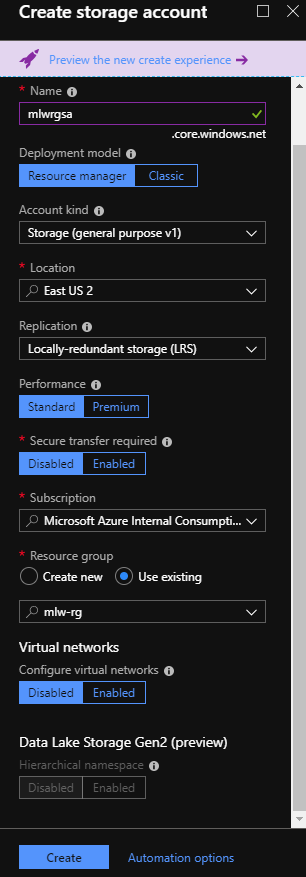
Ensure you can access your Azure pass, and connect to the Azure subscription provided to you for the workshop. Note: There is just enough credit to do the workshop.

# Create an Azure resource group

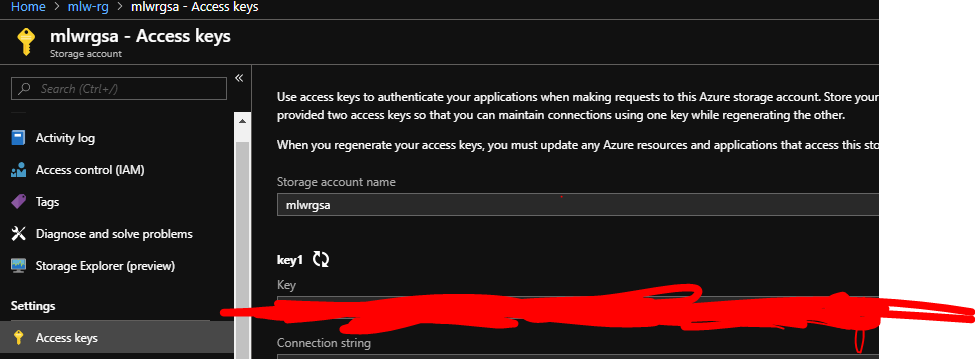
Name: mlw-rg  
Location: US East 2

# Create storage account

Resource group: mlw-rg  
Name: mlw<yourinitials>sa

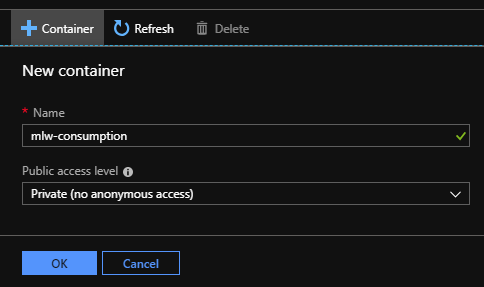


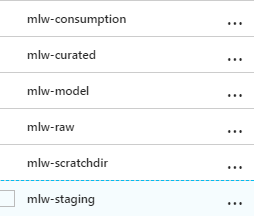
Go to Settings -> Access keys -> Copy and save key1 in a notepad



# Go to Blob and create storage containers in the storage account

Choose public access level as “Private”

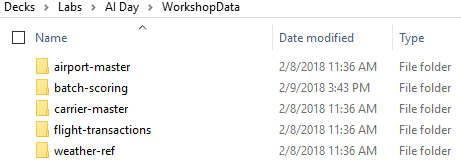




# Download the workshop data to your machine

<https://flightdelayprediction.blob.core.windows.net/workshop-data/WorkshopData.zip>

Unzip the downloaded datasets



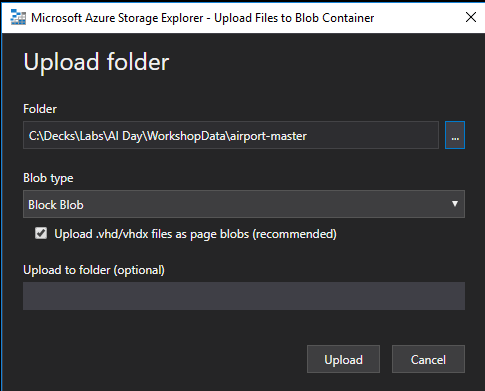
1. Install Azure storage explorer in your machine

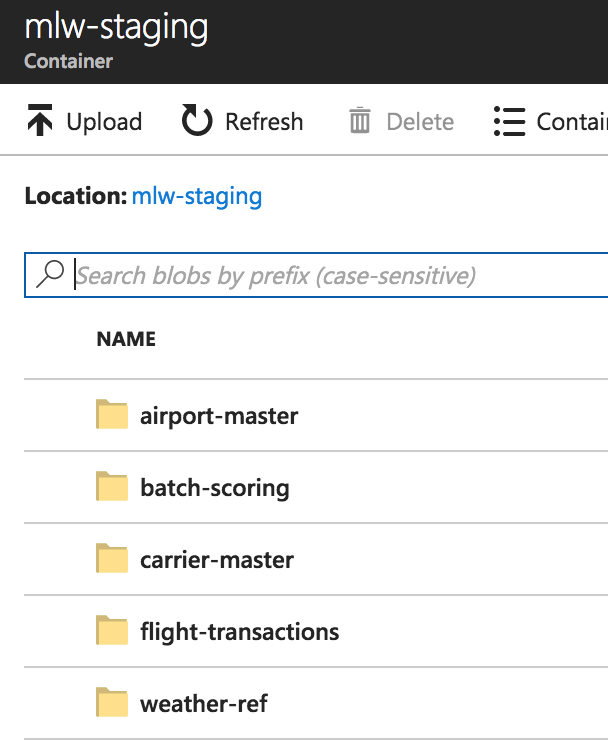
<https://azure.microsoft.com/en-us/features/storage-explorer/>

Connect to the Azure subscription and navigate to the storage account container, mlw-staging

# Upload the workshop data to the storage container mlw-staging

Using storage explorer, copy the 5 folders to the storage container. Choose Block Blob as Blob type



Should look like this..  


# Provision an Azure SQL database in the resource group you created

Use a database name and server name that is easy to type up and remember.

For example:

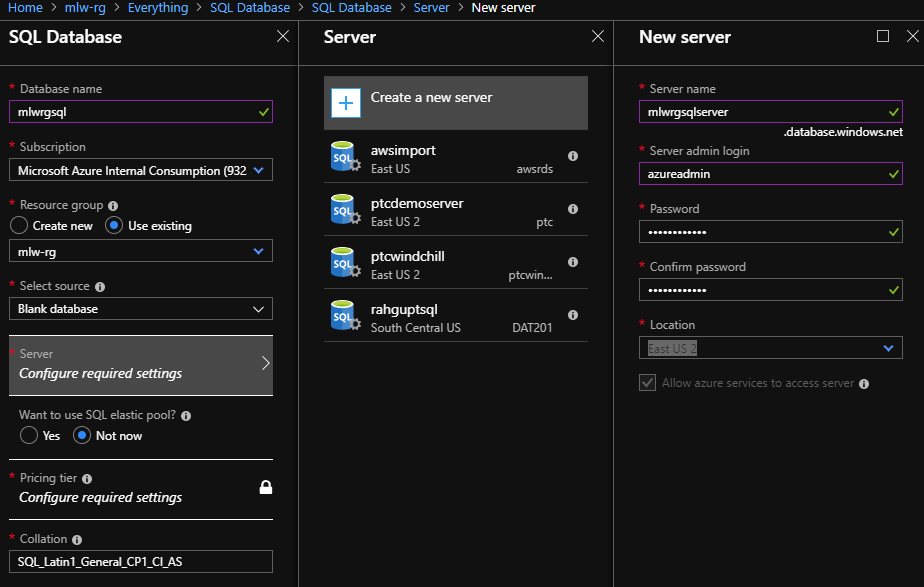
DB Name: mlw<yourinitials>sql

Server Name: mlw<yourinitials>sqlserver

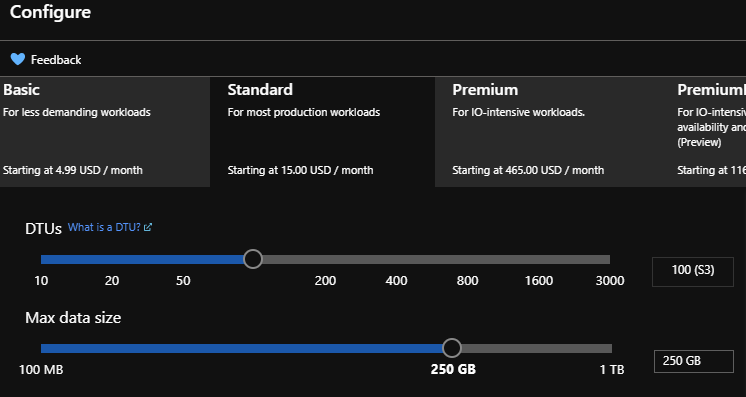
Server admin login: azureadmin

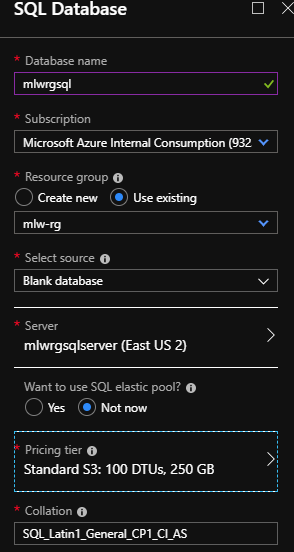
Server password: <your choice>

Make a note of the credentials



Change pricing tier to Standard 100 DTUs, 250GB

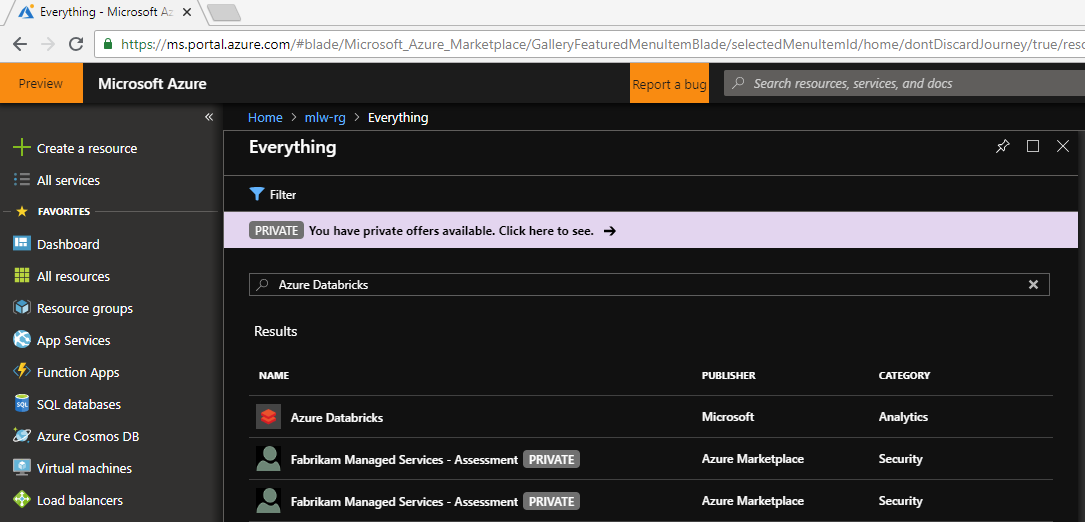




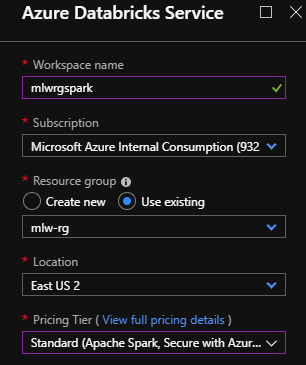
# Provision Azure Databricks workspace in the resource group you created

**Note: (Use Google Chrome or Edge browser. Do not use IE). If you hit issues, use incognito/inprivate window.**

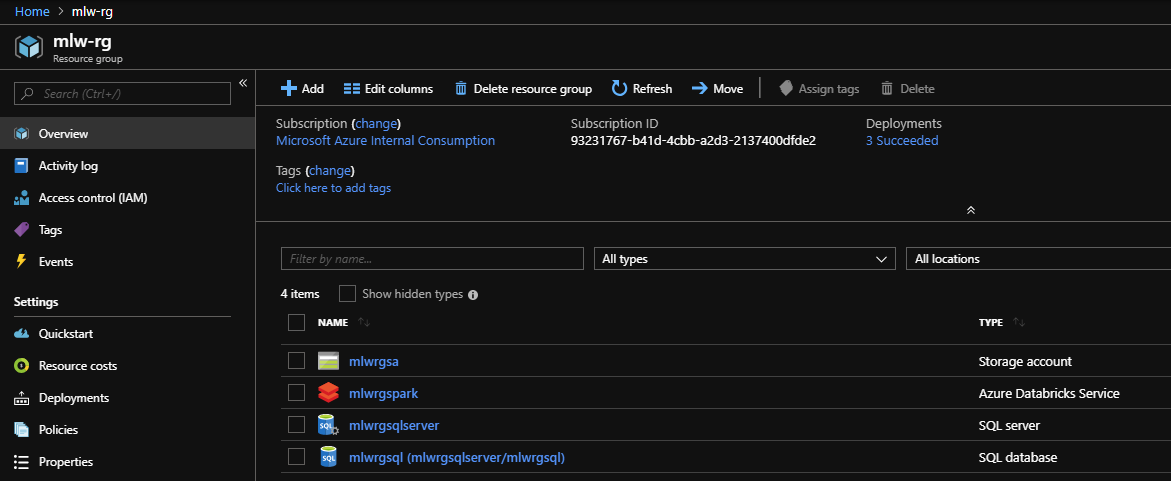
Go to Azure Portal -> Create a resource -> Search for Azure Databricks -> Select Azure Databricks -> Create

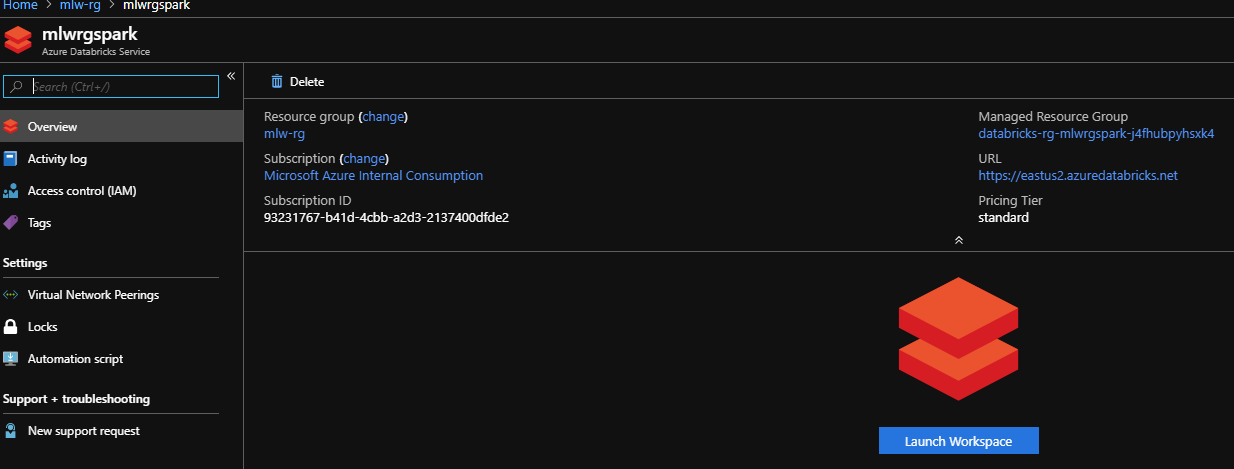


Put workspace name as: mlw<your initials>spark



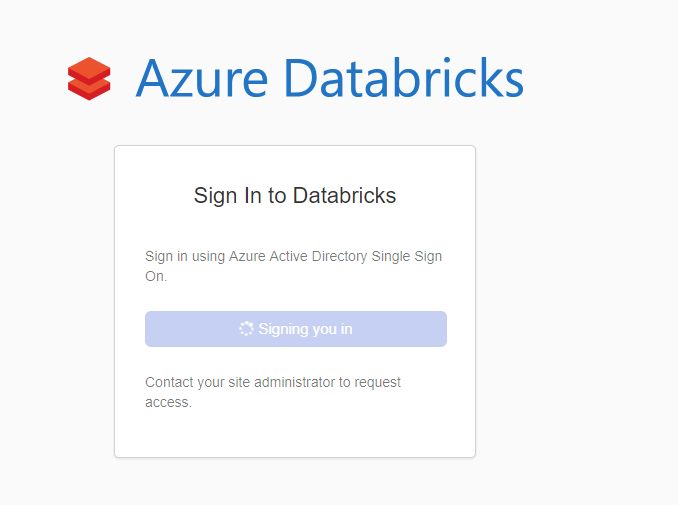
Go to Databricks workspace once created:





Click on Launch workspace:

It will open another tab in the browser and sign you in using Azure Active Directory SSO



Once successfully signed in you should see this screen:

