

# Quickstart Guide for Using Azure Machine Learning

version 1.0

Last updated April 28, 2020  
Created by Lee Kok How

This is a step-by-step guide from creating an Azure account to execute codes in Azure Notebooks that uses Azure ML SDK to interact with Azure Machine Learning workspace.

This guide complements the two Notebooks, predict-employee-retention-part1-training.ipynb and predict-employee-retention-part2-deploy.ipynb.

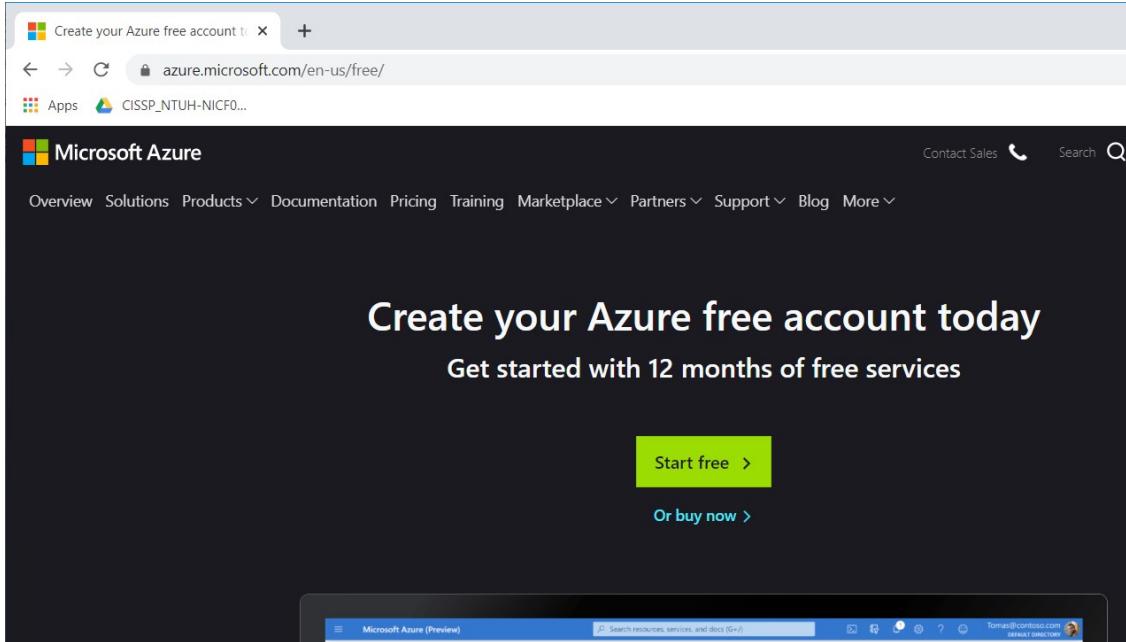
Please go through all the steps here to make sure the Azure Machine Learning setup is complete.

## Contents

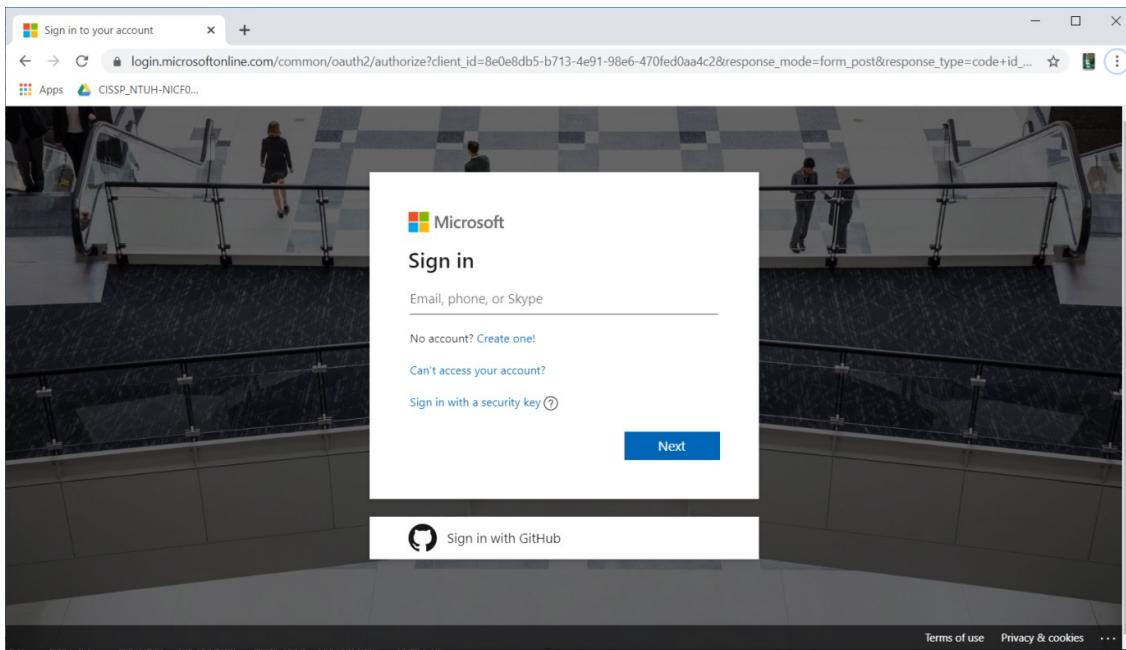
Step 1: Register Azure Account.....	3
Step 2: Create a Resource Group .....	8
Step 3: Create a Azure Machine Learning Workspace in your Resource Group.....	13
Step 4: Create Compute Resource .....	18
Step 5: Upload data into Datastore .....	24
Step 6: Create Microsoft Azure Notebooks account to run Jupyter Notebooks. ....	36
Step 7: Clone an Azure ML Notebook example to start with .....	38

## Step 1: Register Azure Account

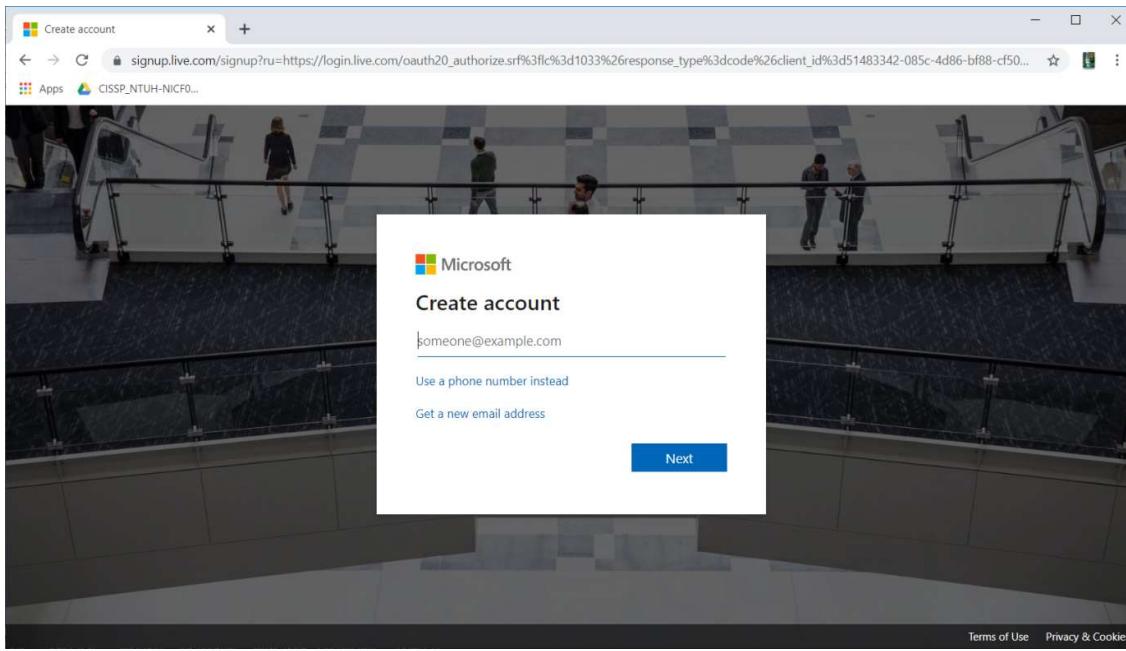
1. Go to <https://azure.microsoft.com/en-us/free>



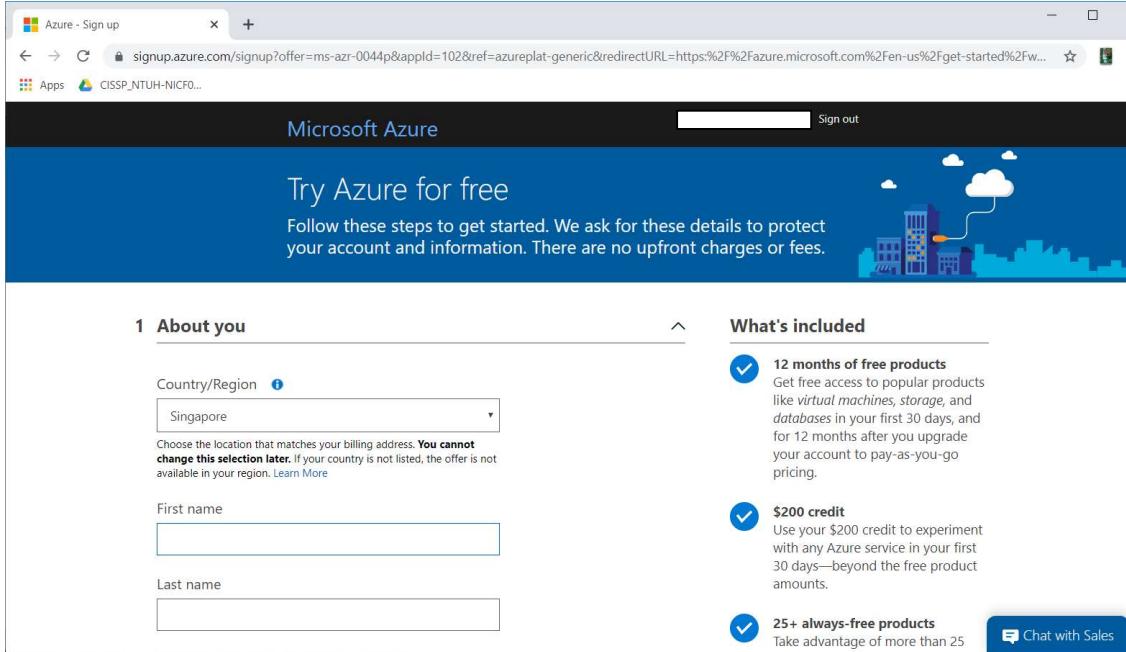
2. Click "Start free"



3. Click "Create one!"



4. Enter your existing email then click "Next". The next few steps will ask you to provide password and verify your email account. The next screen is displayed when your account is created.



- Fill up the particulars. Note you will be prompted to provide credit card details. But you won't be charged when you use the free account.

The screenshot shows the Azure sign-up process at step 3, titled "Identity verification by card". It asks for a credit card number to verify identity and keep out spam and bots. A note states "You won't be charged unless you upgrade." Below, fields are provided for Cardholder Name, Card number, and Expires (MM/YY). To the right, three promotional boxes are displayed:

- \$200 credit**: Use your \$200 credit to with any Azure service i 30 days—beyond the fi amounts.
- 25+ always-free prod**: Take advantage of mor products, including sen containers, and artificia that are always free. Ge your first 30 days, and once you choose to up
- No automatic charges**: You won't be charged u choose to upgrade. bef

- Once you filled up the details, click on "Sign up".

The screenshot shows the Azure sign-up process at step 4, titled "Agreement". It contains two checkboxes for accepting the customer agreement and privacy agreement, and for receiving information from Microsoft. Below, a "Sign up" button is shown, followed by a progress bar indicating "Setting up your account...". To the right, three promotional boxes are displayed:

- \$200 credit**: Use your \$200 credit to with any Azure service i 30 days—beyond the fi amounts.
- 25+ always-free prod**: Take advantage of mor products, including sen containers, and artificia that are always free. Ge your first 30 days, and once you choose to up
- No automatic charges**: You won't be charged u choose to upgrade. bef

7. Once the sign up is successful, you will see this screen.

The screenshot shows a Microsoft Edge browser window with the URL [azure.microsoft.com/en-us/get-started/welcome-to-azure/?subscriptionId=e99102b6-42f3-497b-8071-47c081d9cda5](https://azure.microsoft.com/en-us/get-started/welcome-to-azure/?subscriptionId=e99102b6-42f3-497b-8071-47c081d9cda5). The page title is "Welcome to Azure - You're Ready". The main heading is "You're ready to start with Azure". Below it is a blue button labeled "Go to the portal >". To the left, there's a section titled "Join the demo to see Azure in action" with a list of items: "Building a virtual machine.", "Creating a web app.", "Deploying a SQL database.", and "Customizing your DevOps dashboard.". To the right, there's a "Register to schedule a live demo" section with a "Select a time" dropdown showing "4/29/2020, 12:00 AM GMT+8 (Change time zone)" selected. There are also two radio buttons for "5/6/2020, 12:00 AM GMT+8" and a "Chat with Sales" button.

8. Click "Go to the portal". You will be re-directed to the Azure Portal screen.

The screenshot shows a Microsoft Edge browser window with the URL [portal.azure.com/?quickstart=true#blade/Microsoft\\_Azure\\_Resources/QuickstartCenterBlade](https://portal.azure.com/?quickstart=true#blade/Microsoft_Azure_Resources/QuickstartCenterBlade). The page title is "Quickstart Center - Microsoft Azure". The main heading is "Quickstart Center". Below it is a "Get started" section with a link "Take an online course". There's a "Setup guides" section with three cards: "Azure setup guide", "Azure migration guide", and "Azure innovation guide", each with an "Open >" button. There's also a "Start a project" section with four cards: "Create a web app", "Deploy a virtual machine", "Set up a database", and "Start a data analytics project", each with a "Start >" button. A "DEFAULT DIRECTORY" button is visible in the top right corner.

9. Bookmark this screen so that you can access it easily.

10. Click on the "Home" breadcrumb to go to the Azure home page. Alternatively, you can click



(on top left) to expand the menu.

A screenshot of the Microsoft Azure portal's home page. The left sidebar contains a navigation menu with sections like 'Create a resource', 'Home', 'Dashboard', 'All services', 'FAVORITES' (which is currently selected), 'All resources', 'Resource groups', 'Quickstart Center', 'App Services', 'Function App', 'SQL databases', 'Azure Cosmos DB', 'Virtual machines', 'Load balancers', 'Storage accounts', 'Virtual networks', 'Azure Active Directory', 'Monitor', and 'Advisor'. A 'with free online...' button is also visible. The main content area features a search bar at the top and a grid of service icons below it. The grid includes 'Resource groups', 'Virtual machines', 'App Services', 'Storage accounts', 'SQL databases', 'Azure Database for PostgreSQL...', 'Azure Cosmos DB', 'Kubernetes services', and a 'More services' button. Below the grid, there is a table showing 'Last Viewed' items: 'Machine Learning' (Machine Learning, last viewed an hour ago) and 'Resource group' (Resource group, last viewed 4 hours ago). At the bottom, there are links for 'Resource groups', 'All resources', 'Dashboard', 'Azure Monitor' (Monitor your apps and infrastructure), 'Security Center' (Secure your apps and infrastructure), and 'Cost Management' (Analyze and optimize your costs).

## Step 2: Create a Resource Group

1. Go to <http://portal.azure.com/> and login with the Azure account you created.
2. Once you login, you will see the following "Quickstart Center" screen.

The screenshot shows the Microsoft Azure Quickstart Center blade. At the top, there's a header bar with the title 'Quickstart Center - Microsoft Azure' and a search bar. Below the header, a navigation bar includes 'Home > Quickstart Center' and a 'Quickstart Center' logo. A notification bar on the right indicates '\$200.00 credit remaining' and 'Subscription 'Azure subscription 1' has a remaining credit of \$200.00'. Below the notification, a link says 'Click here to upgrade to a Pay-As-You-Go subscription.' The main content area is divided into sections: 'Get started' (with a 'Take an online course' link), 'Setup guides' (listing 'Azure setup guide', 'Azure migration guide', and 'Azure innovation guide' with 'Open >' links), 'Start a project' (listing 'Create a web app', 'Deploy a virtual machine', 'Set up a database', and 'Start a data analytics project' with 'Start >' links), and a 'DEFAULT DIRECTORY' dropdown.

3. Click "Home".

The screenshot shows the Microsoft Azure Home blade. At the top, there's a header bar with the title 'Home - Microsoft Azure' and a search bar. Below the header, a navigation bar includes 'Home > Quickstart Center <...'. The main content area is titled 'Azure services' and features a 'Create a resource' button (which is highlighted with a blue dashed border). Other service icons include 'Resource groups', 'Virtual machines', 'App Services', 'Storage accounts', 'SQL databases', 'Azure Database for PostgreSQL...', 'Azure Cosmos DB', 'Kubernetes services', and a 'More services' link. Below this, there's a 'Navigate' section with links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. Under the 'Tools' section, there are links for 'Microsoft Learn', 'Azure Monitor', 'Security Center', and 'Cost Management'. At the bottom, there's a 'Useful links' section with a link to 'https://portal.azure.com/?quickstart=true#blade/Microsoft\_Azure\_Monitoring/AzureMonitoringBrowseBlade' and an 'Azure mobile app' section.

4. Click "Resource groups". For the purpose of learning or exploring, you just need to create one Resource group.

5. Click "Add" (or the "Create resource group" button below). Enter a name for the Resource group and select Southeast Asia for Region.

6. Click "Review + create".

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/?quickstart=true#create/Microsoft.ResourceGroup>. The page title is "Create a resource group". The "Validation passed." message is displayed in a green bar. The "Review + create" tab is selected. Under the "Basics" section, the following details are shown:

Subscription	Azure subscription 1
Resource group	LKH-ResourceGroup
Region	Southeast Asia

At the bottom, there are "Create", "< Previous", and "Next >" buttons.

7. Click "Create". You can click on the bell icon to look at "Notifications". You can click on the "Go to resource group" button to access to the resource group you just created.

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/?quickstart=true#blade/HubsExtension/BrowseResourceGroups>. The page title is "Resource groups". A notification on the right side says "Resource group created" with the message "Creating resource group 'LKH-ResourceGroup' in subscription 'Azure subscription 1' succeeded.". Below the notification, there is a "Go to resource group" button. The main area shows a message "No resource groups to display" and a "Create resource group" button.

8. Alternatively, you can go to "Home", then click on the "Resource Group" button.

The screenshot shows the Microsoft Azure Home page. At the top, there's a search bar and a 'DEFAULT DIRECTORY' dropdown. Below the header, there's a section titled 'Azure services' with various service icons like Create a resource, Resource groups, Virtual machines, App Services, Storage accounts, SQL databases, Azure Database for PostgreSQL, Azure Cosmos DB, Kubernetes services, and More services. Under 'Navigate', the 'Resource groups' button is highlighted with a blue border. In the 'Tools' section, there are links to Microsoft Learn, Azure Monitor, Security Center, and Cost Management. At the bottom left, there's a 'Useful links' section with a link to the Azure mobile app. The URL in the address bar is <https://portal.azure.com/?quickstart=true#blade/HubsExtension/BrowseResourceGroupBlade/resourceType/Microsoft.Resources%2Fsubscriptions%2FresourceGroups>.

9. In Resource groups, you should be able to find the group that you created.

The screenshot shows the Microsoft Azure Resource groups page. The URL in the address bar is <https://portal.azure.com/#blade/HubsExtension/BrowseResourceGroups>. The page title is 'Resource groups'. It shows a table with one record: 'LKH-ResourceGroup' under 'Name', 'Azure subscription 1' under 'Subscription', and 'Southeast Asia' under 'Location'. There are filter options at the top and pagination at the bottom.

Name	Subscription	Location
LKH-ResourceGroup	Azure subscription 1	Southeast Asia

10. Click on the link of the resource group you created to see its details.

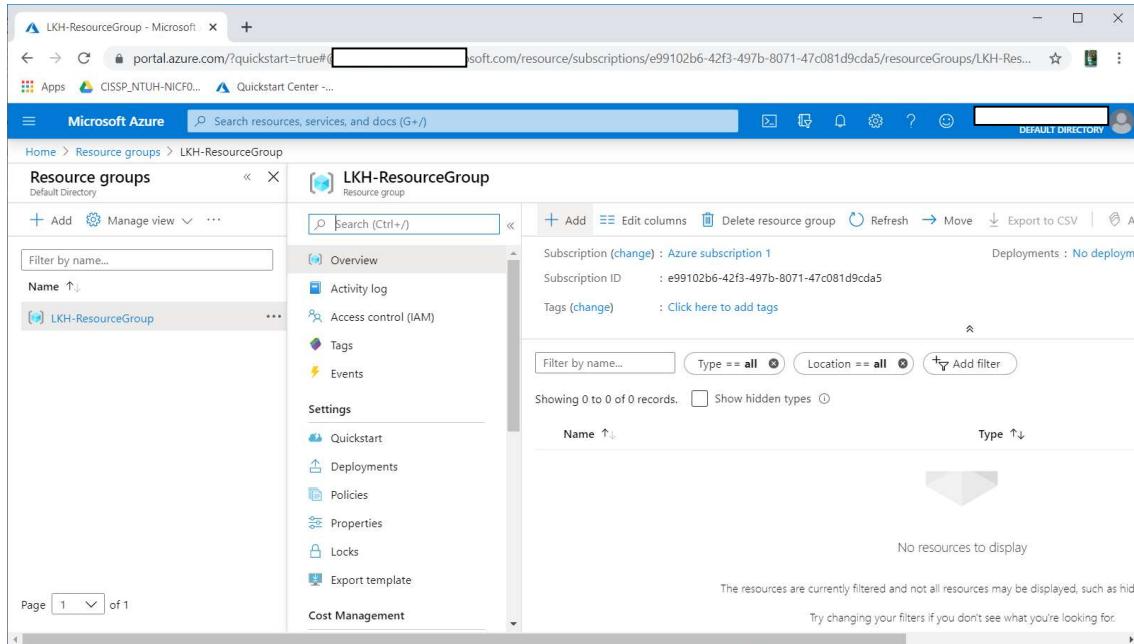
The screenshot shows the Microsoft Azure portal interface. The left sidebar displays 'Resource groups' with a single item, 'LKH-ResourceGroup'. The main content area is titled 'LKH-ResourceGroup' and contains the following information:

- Subscription (change) :** Azure subscription 1
- Subscription ID :** e99102b6-42f3-497b-8071-47c081d9cda5
- Tags (change) :** Click here to add tags

Below this, there is a table header with columns 'Name ↑↓' and 'Type ↑↓'. The message 'No resources to display' is shown below the table. At the bottom, a note says 'The resources are currently filtered and not all resources may be displayed, such as hidden resources.'

## Step 3: Create a Azure Machine Learning Workspace in your Resource Group

1. Go to your Resource Group. Click "Add" on the right pane (see screen below highlight in grey)



The screenshot shows the Microsoft Azure portal interface. The left sidebar displays the 'Resource groups' section with a list containing 'LKH-ResourceGroup'. The main content area is titled 'LKH-ResourceGroup' and shows the following details:

- Subscription (change) : Azure subscription 1
- Subscription ID : e99102b6-42f3-497b-8071-47c081d9cda5
- Tags (change) : Click here to add tags

The right pane lists resources with the following filters:

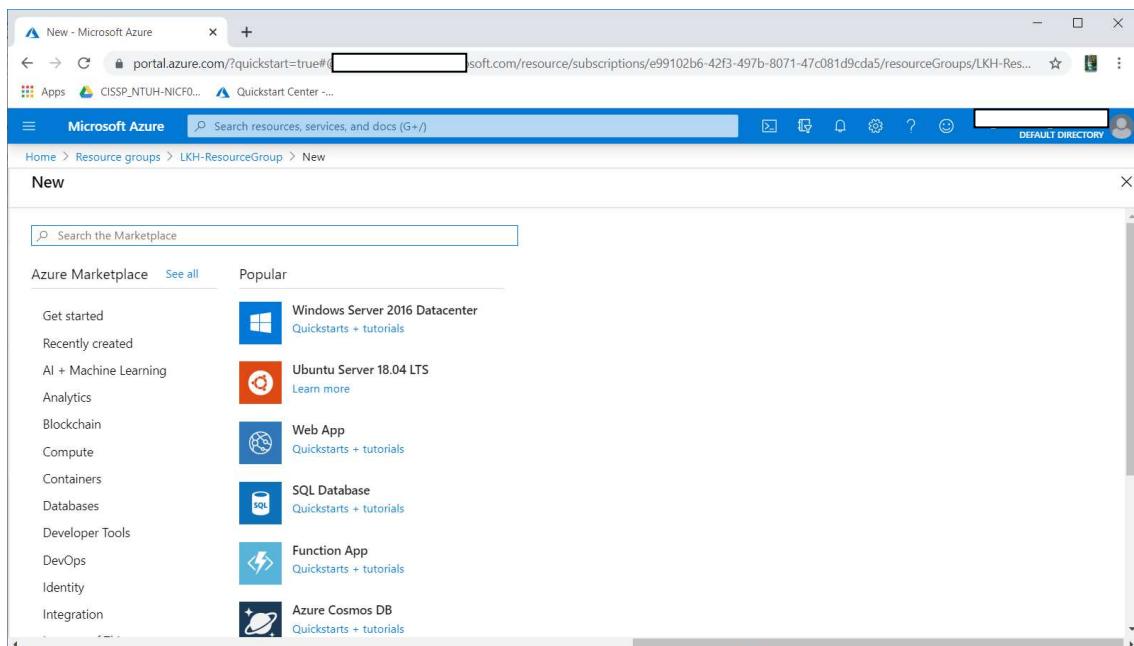
- Filter by name...
- Type == all
- Location == all
- Add filter

The results table shows:

Name	Type
No resources to display	

A message at the bottom states: "The resources are currently filtered and not all resources may be displayed, such as hidden resources. Try changing your filters if you don't see what you're looking for."

2. You will see the Azure Marketplace screen.



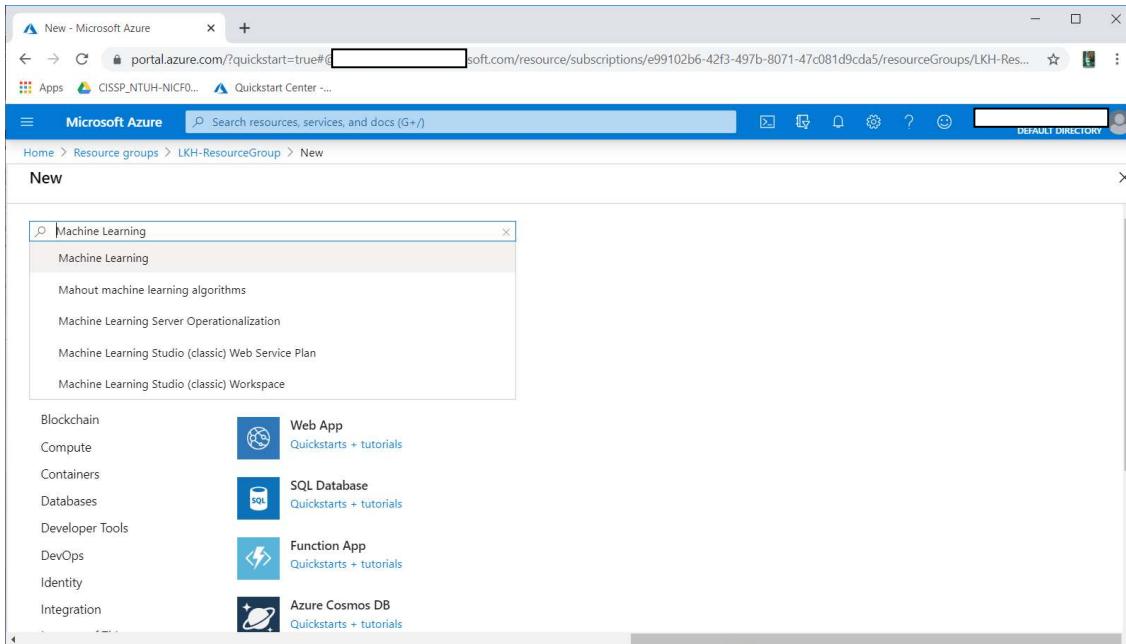
The screenshot shows the Microsoft Azure portal interface with a 'New' tab selected. The main content area is titled 'New' and features a search bar labeled 'Search the Marketplace'.

The 'Azure Marketplace' section includes:

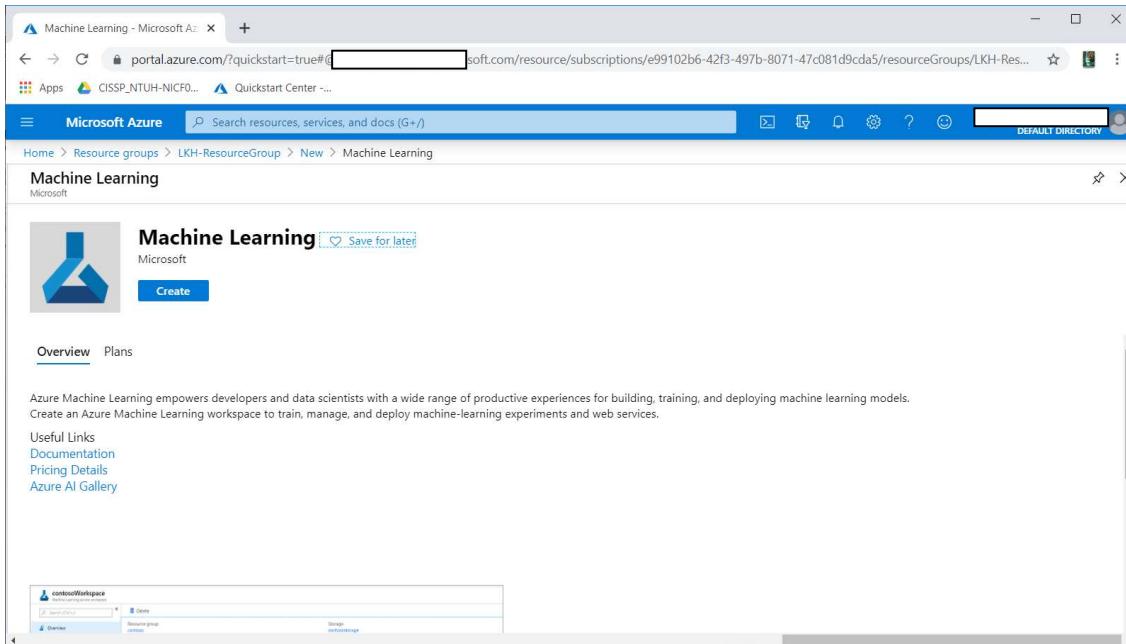
- A 'Popular' tab.
- A 'Get started' link.
- A 'Recently created' link.
- A 'AI + Machine Learning' link.
- A 'Analytics' link.
- A 'Blockchain' link.
- A 'Compute' link.
- A 'Containers' link.
- A 'Databases' link.
- A 'Developer Tools' link.
- A 'DevOps' link.
- A 'Identity' link.
- A 'Integration' link.

Next to each link is a small icon and a brief description.

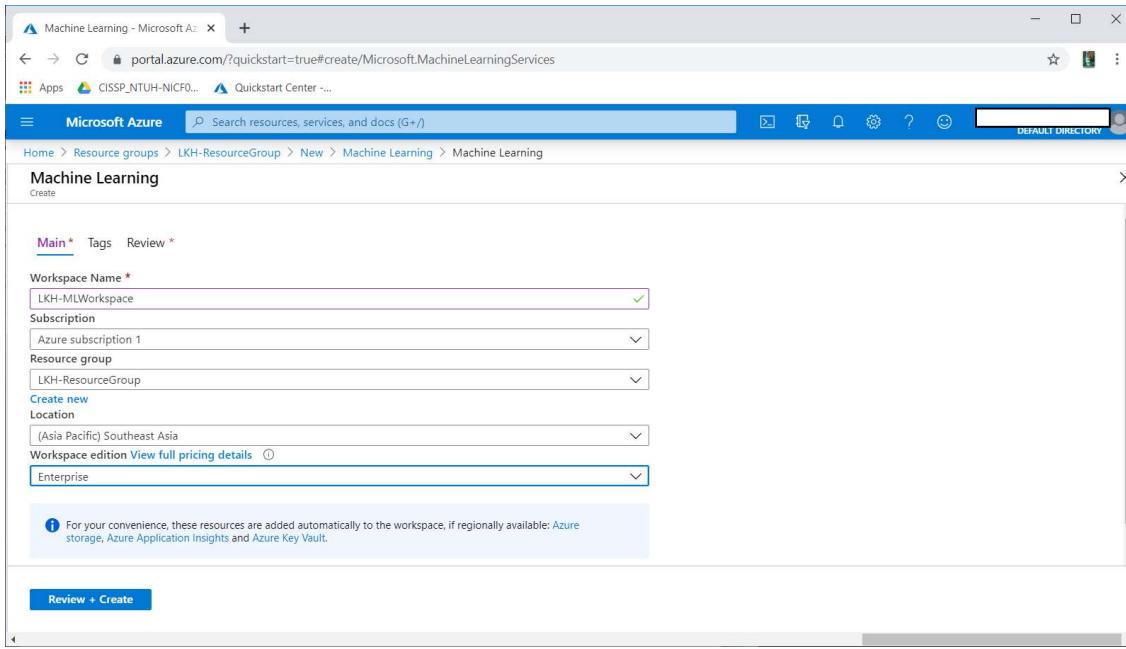
3. In the search box, enter "Machine Learning" then press Enter key, or click it when it appears in the auto-complete list.



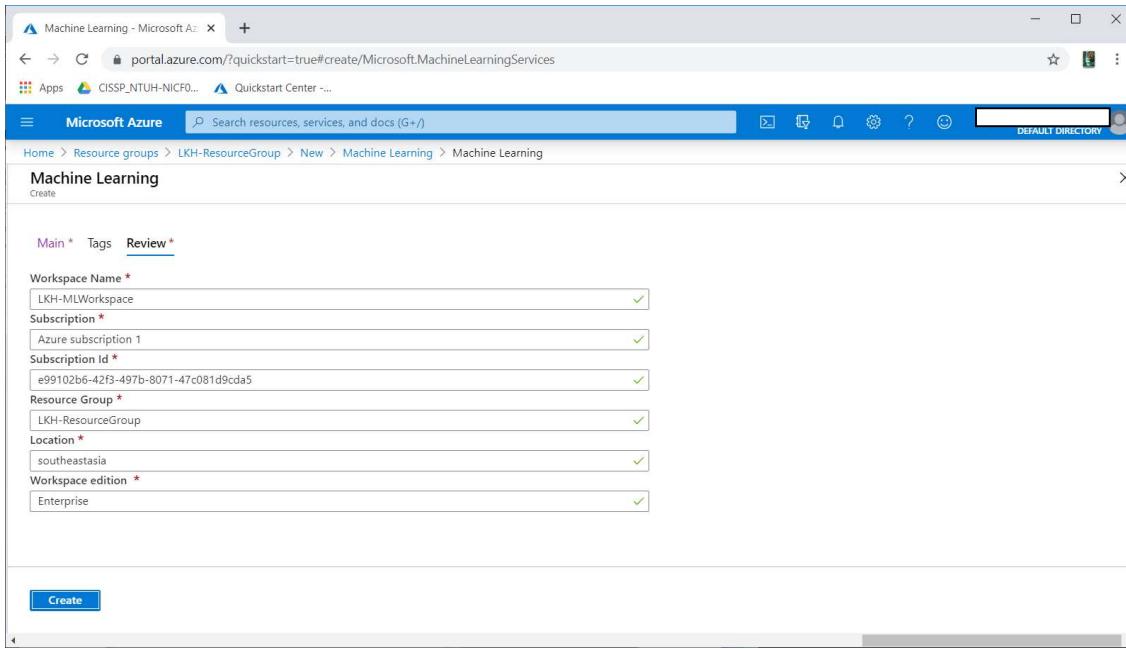
4. You will see the Azure Machine Learning service, click "Create".



5. Enter a name for your machine learning workspace, and select the Workspace edition as "Enterprise".



6. Click "Review + Create". Check the details.



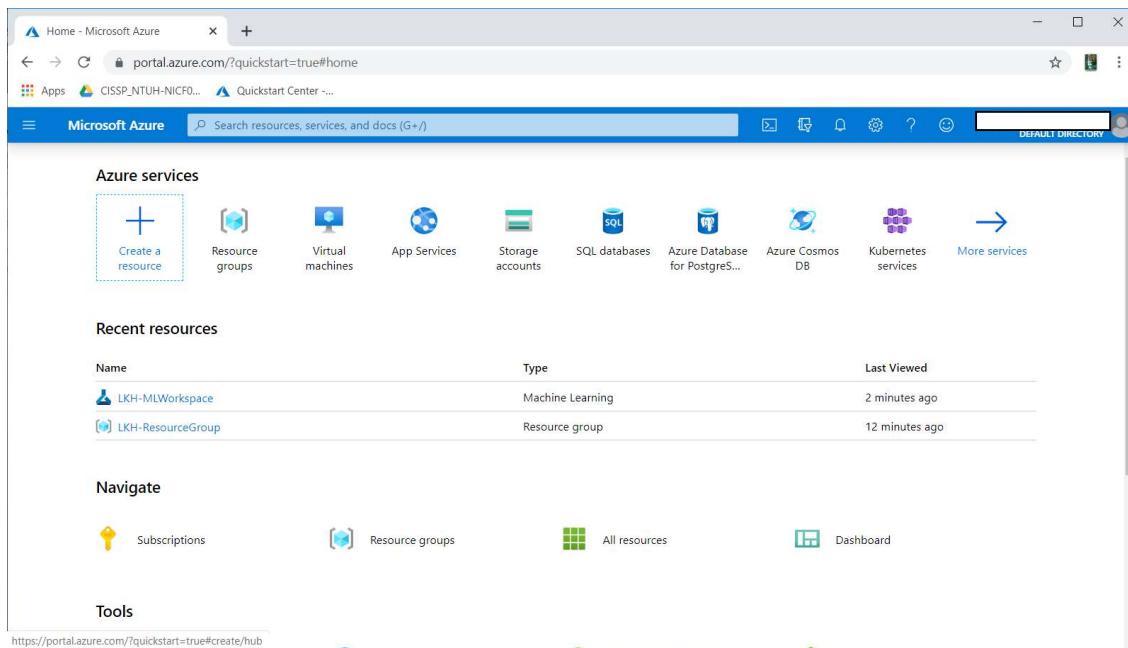
7. Click "Create". You will see the notification when your workspace is created.

The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with 'Overview', 'Inputs', 'Outputs', and 'Template'. The main content area displays a deployment status message: 'Your deployment is complete' with a green checkmark icon. Below it, deployment details are listed: Deployment name: Microsoft.MachineLearningServices, Subscription: Azure subscription 1, Resource group: LKH-ResourceGroup. There are two expandable sections: 'Deployment details' and 'Next steps', each with a 'Go to resource' button. To the right, a 'Notifications' panel is open, showing three items: 1) 'Deployment succeeded' (green checkmark, 2 minutes ago), 2) 'Resource group created' (green checkmark, an hour ago), and 3) '\$200.00 credit remaining' (info icon, an hour ago). Each notification has a 'Go to resource group' and 'Pin to dashboard' button.

8. Click "Go to resource" in the notification to access to your workspace.

The screenshot shows the Microsoft Azure portal with the URL 'LKH-MLWorkspace - Microsoft A...'. The left sidebar includes 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Events', 'Assets' (with sub-options like Experiments, Pipelines, Compute, Models, Images, Deployments, Activities), and 'Getting Started'. The main content area displays workspace details: Edition: Enterprise, Resource group: LKH-ResourceGroup, Location: Southeast Asia, Subscription: Azure subscription 1, and Subscription ID: e99102b6-42f3-497b-8071-47c081d9cda5. Below this is a large promotional box for the 'Azure Machine Learning studio': 'Try the new Azure Machine Learning studio' with a 3D cityscape graphic, 'Introducing a new immersive experience (preview) for managing the end-to-end machine learning lifecycle.', 'Launch now' button, and 'Learn more' link.

9. Alternatively, you can launch the workspace at the Home page.



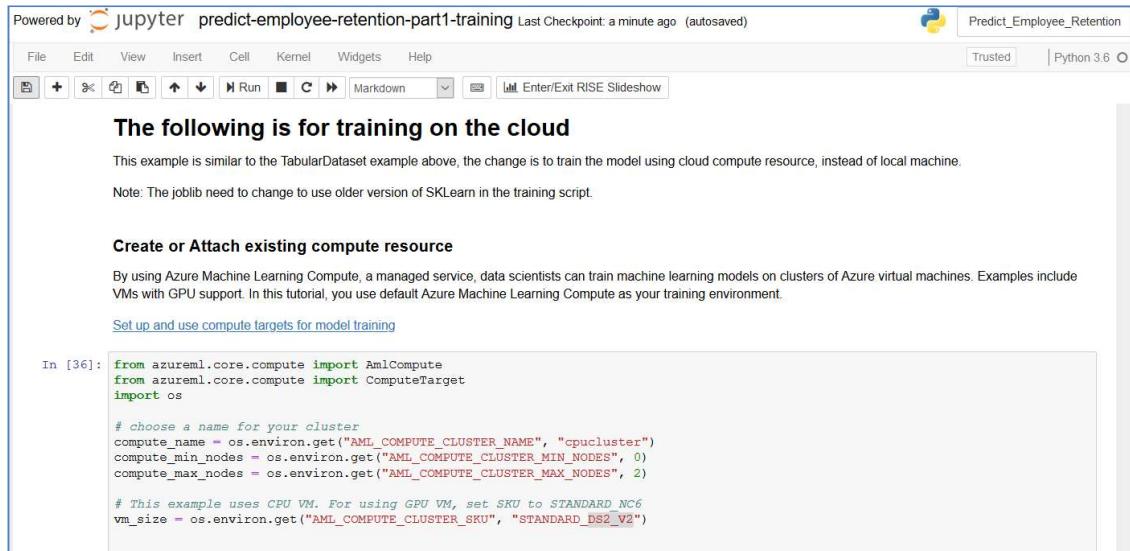
The screenshot shows the Microsoft Azure Home page. At the top, there's a navigation bar with icons for back, forward, refresh, and search, followed by the URL 'portal.azure.com/?quickstart=true#home'. Below the navigation bar is a 'Microsoft Azure' header with a search bar and various status indicators. The main content area is titled 'Azure services' and features a 'Create a resource' button (which is highlighted with a blue dashed box). Other service icons include Resource groups, Virtual machines, App Services, Storage accounts, SQL databases, Azure Database for PostgreSQL, Azure Cosmos DB, Kubernetes services, and a 'More services' link. Below this is a section titled 'Recent resources' with a table:

Name	Type	Last Viewed
LKH-MLWorkspace	Machine Learning	2 minutes ago
LKH-ResourceGroup	Resource group	12 minutes ago

Further down, there's a 'Navigate' section with links for Subscriptions, Resource groups, All resources, and Dashboard. At the bottom, there's a 'Tools' section with a link to 'https://portal.azure.com/?quickstart=true#create/hub'.

## Step 4: Create Compute Resource

This step is to create a compute resource that will be used for model training on the cloud. It is also possible to create this compute resource using the Azure ML SDK:



The following is for training on the cloud

This example is similar to the TabularDataset example above, the change is to train the model using cloud compute resource, instead of local machine.

Note: The joblib need to change to use older version of SKLearn in the training script.

**Create or Attach existing compute resource**

By using Azure Machine Learning Compute, a managed service, data scientists can train machine learning models on clusters of Azure virtual machines. Examples include VMs with GPU support. In this tutorial, you use default Azure Machine Learning Compute as your training environment.

[Set up and use compute targets for model training](#)

```
In [36]: from azureml.core.compute import AmlCompute
from azureml.core.compute import ComputeTargetException
import os

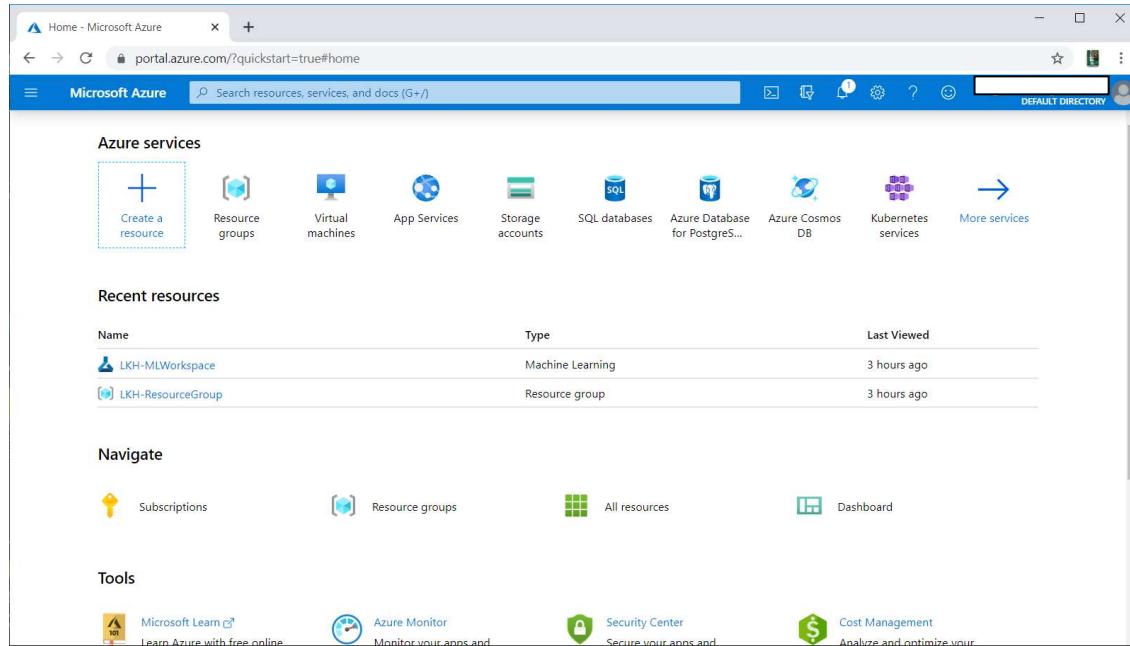
# choose a name for your cluster
compute_name = os.environ.get("AML_COMPUTE_CLUSTER_NAME", "cpucluster")
compute_min_nodes = os.environ.get("AML_COMPUTE_CLUSTER_MIN_NODES", 0)
compute_max_nodes = os.environ.get("AML_COMPUTE_CLUSTER_MAX_NODES", 2)

# This example uses CPU VM. For using GPU VM, set SKU to STANDARD_NC6
vm_size = os.environ.get("AML_COMPUTE_CLUSTER_SKU", "STANDARD_DS2_V2")
```

You can find the sample code in predict-employee-retention-part1-training.ipynb, under "Create or Attach existing compute resource".

We will create the same compute resource from the Azure Portal, so that we can confirm your account can create compute resource.

1. From Home, click on your Azure ML Workspace.



Azure services

- Create a resource
- Resource groups
- Virtual machines
- App Services
- Storage accounts
- SQL databases
- Azure Database for PostgreSQL
- Azure Cosmos DB
- Kubernetes services
- More services

Recent resources

Name	Type	Last Viewed
LKH-MLWorkspace	Machine Learning	3 hours ago
LKH-ResourceGroup	Resource group	3 hours ago

Navigate

- Subscriptions
- Resource groups
- All resources
- Dashboard

Tools

- Microsoft Learn
- Azure Monitor
- Security Center
- Cost Management

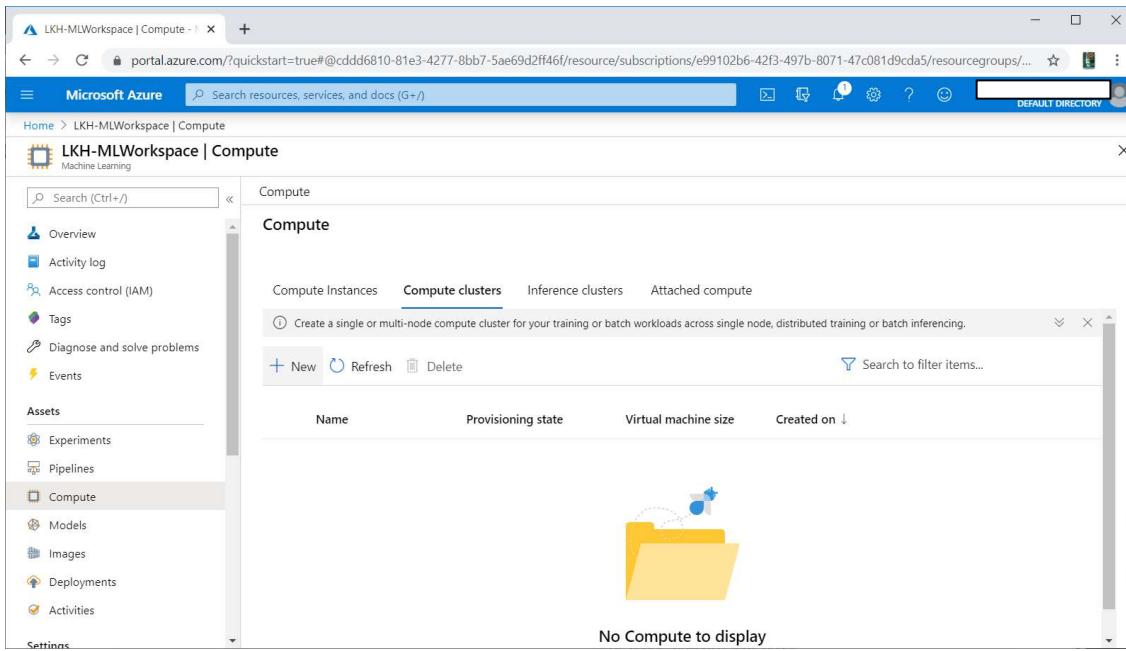
2. Click "Compute" on the left menu.

The screenshot shows the Microsoft Azure portal interface for the LKH-MLWorkspace. The left sidebar has a 'Compute' icon selected under 'Assets'. The main content area displays workspace details: Edition: Enterprise, Resource group: LKH-ResourceGroup, Location: Southeast Asia, Subscription: Azure subscription 1, and Subscription ID: e99102b6-42f3-497b-8071-47c081d9cda5. A large callout box in the center promotes the 'Azure Machine Learning studio' with a 'Launch now' button and a 'Learn more' link. Below the callout, there's a 'Getting Started' section with a 3D model icon.

3. You will see the Compute screen. Click on the "Compute clusters" tab.

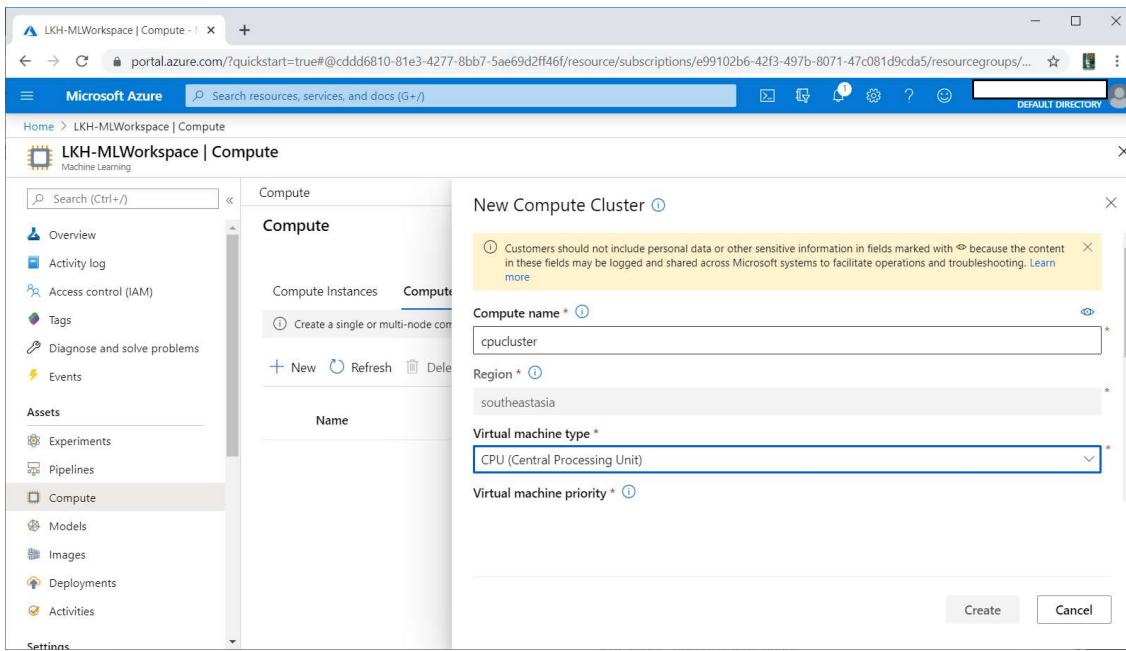
The screenshot shows the Microsoft Azure portal interface for the LKH-MLWorkspace, specifically the 'Compute' screen. The left sidebar has a 'Compute' icon selected under 'Assets'. The main content area is titled 'Compute' and contains tabs for 'Compute Instances', 'Compute clusters', 'Inference clusters', and 'Attached compute'. The 'Compute clusters' tab is active. It includes a help section with two items: 'Get started with AzureML notebooks by creating a new compute instance. Choose from a selection of CPU or GPU instances pre-installed with the most com...' and 'Compute instance is replacing the Notebook VM. You cannot create new Notebook VMs, but you can still use existing Notebook VMs. Learn More.' Below this is a toolbar with buttons for '+ New', 'Refresh', 'Start', 'Stop', 'Restart', 'Delete', and '...'. A search bar 'Search to filter items...' is also present. A table below lists compute instances with columns: Name, Status, Application URI, Virtual Machi..., and Created on. An illustration of a yellow folder with a blue gear icon is displayed at the bottom.

4. In the Compute clusters, click "New" to create a computing resource.



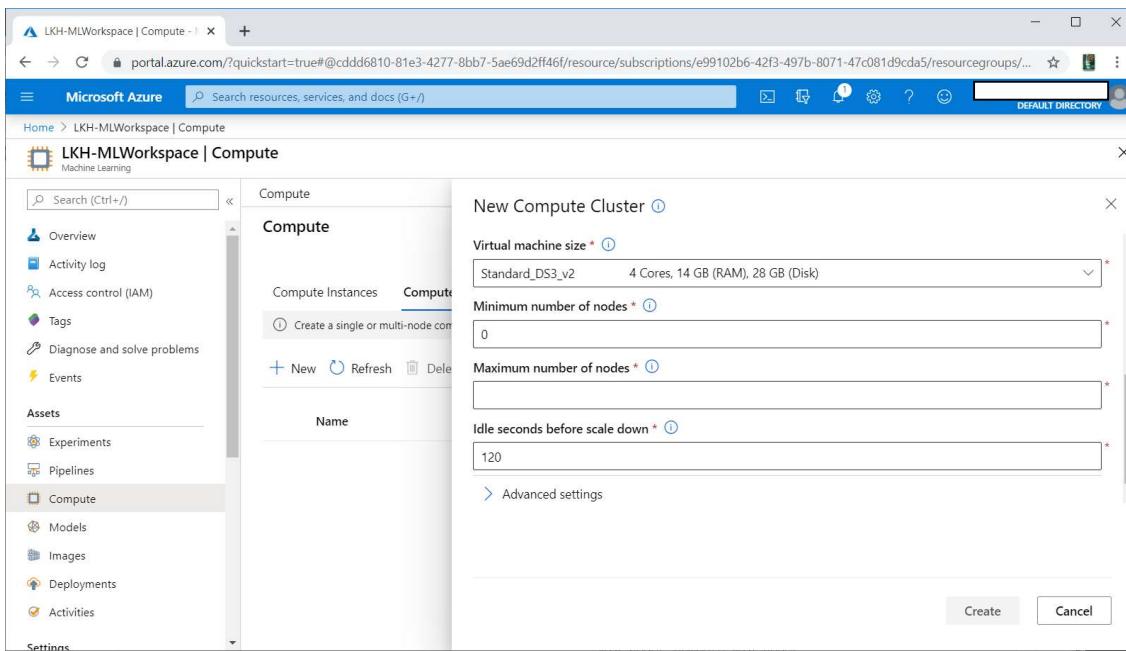
The screenshot shows the Microsoft Azure portal interface for a Machine Learning workspace named 'LKH-MLWorkspace'. The left sidebar has 'Compute' selected under 'Machine Learning'. The main area is titled 'Compute' and shows the 'Compute clusters' tab. A sub-header says 'Create a single or multi-node compute cluster for your training or batch workloads across single node, distributed training or batch inferencing.' Below it are buttons for '+ New', 'Refresh', and 'Delete', and a search bar 'Search to filter items...'. A table lists columns: Name, Provisioning state, Virtual machine size, and Created on. A large yellow folder icon with a blue plus sign is centered, and the message 'No Compute to display' is at the bottom.

5. Enter "cpucluster" for the Computer name. Select "CPU (Central Processing Unit)" in the Virtual machine type.

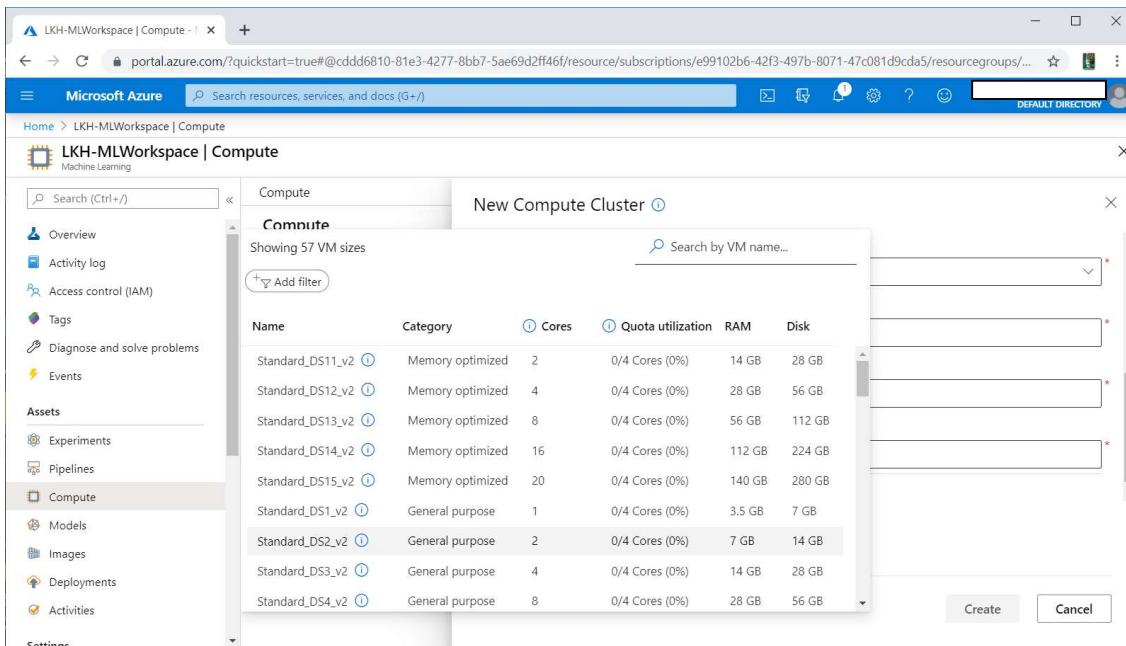


The screenshot shows the 'New Compute Cluster' dialog box overlaid on the Azure Compute blade. The dialog has a yellow header 'New Compute Cluster'. It contains a warning message: 'Customers should not include personal data or other sensitive information in fields marked with because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting. [Learn more](#)' with a close button 'X'. The form fields are: 'Compute name \*' with value 'cpucluster', 'Region \*' with value 'southeastasia', 'Virtual machine type \*' with value 'CPU (Central Processing Unit)' (selected from a dropdown), and 'Virtual machine priority \*' (unfilled). At the bottom are 'Create' and 'Cancel' buttons.

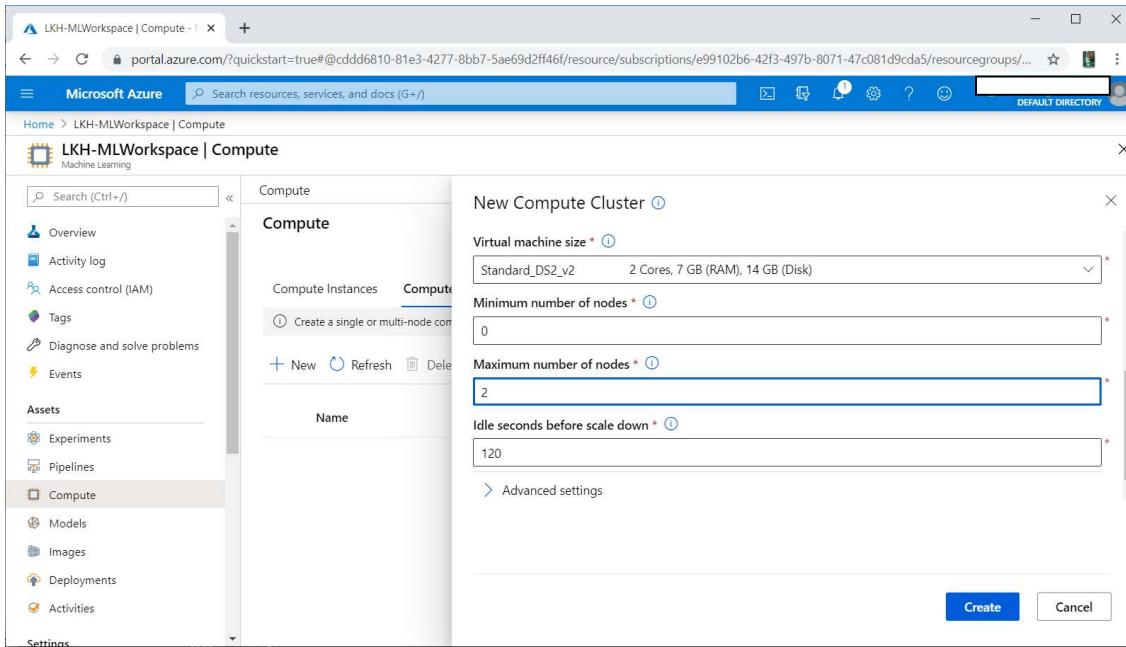
6. Scroll down further to enter more details.



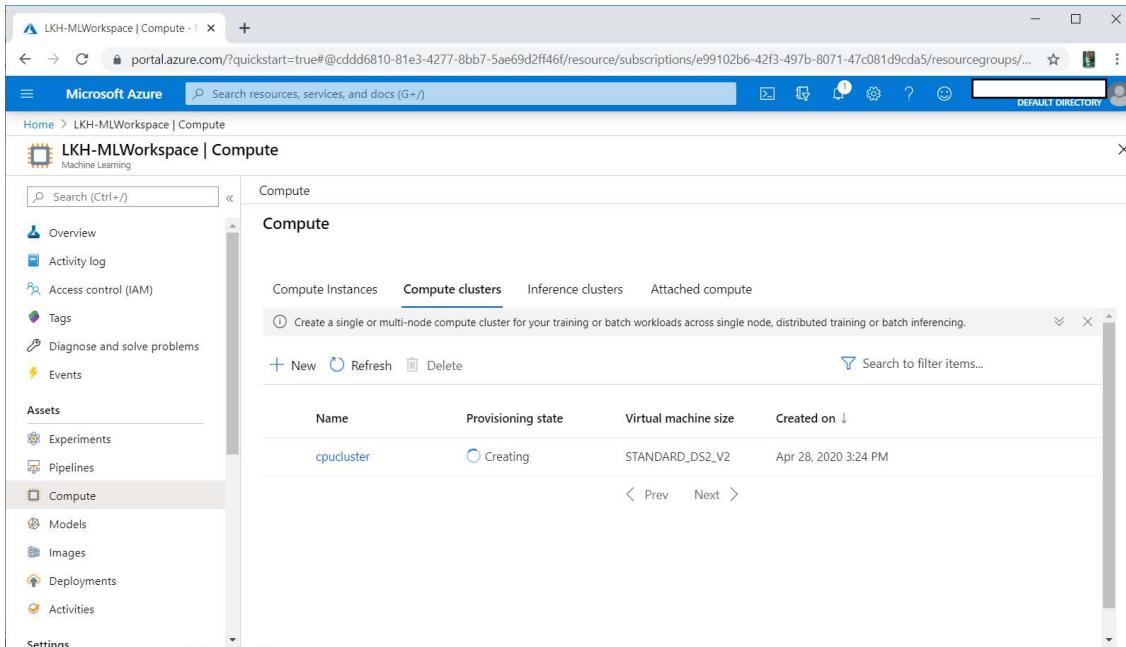
7. Select Virtual machine size "Standard\_DS2\_V2".



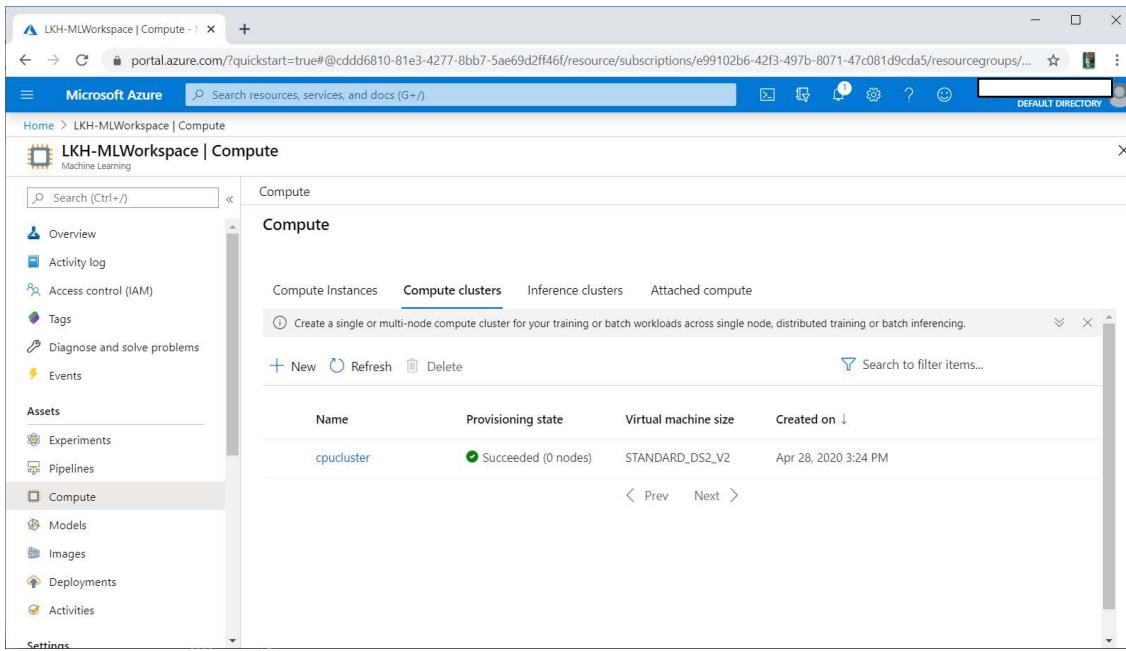
8. Enter 0 for minimum number of nodes, 2 for maximum number of nodes. Then click "Create".



9. Wait for the resource to be created...

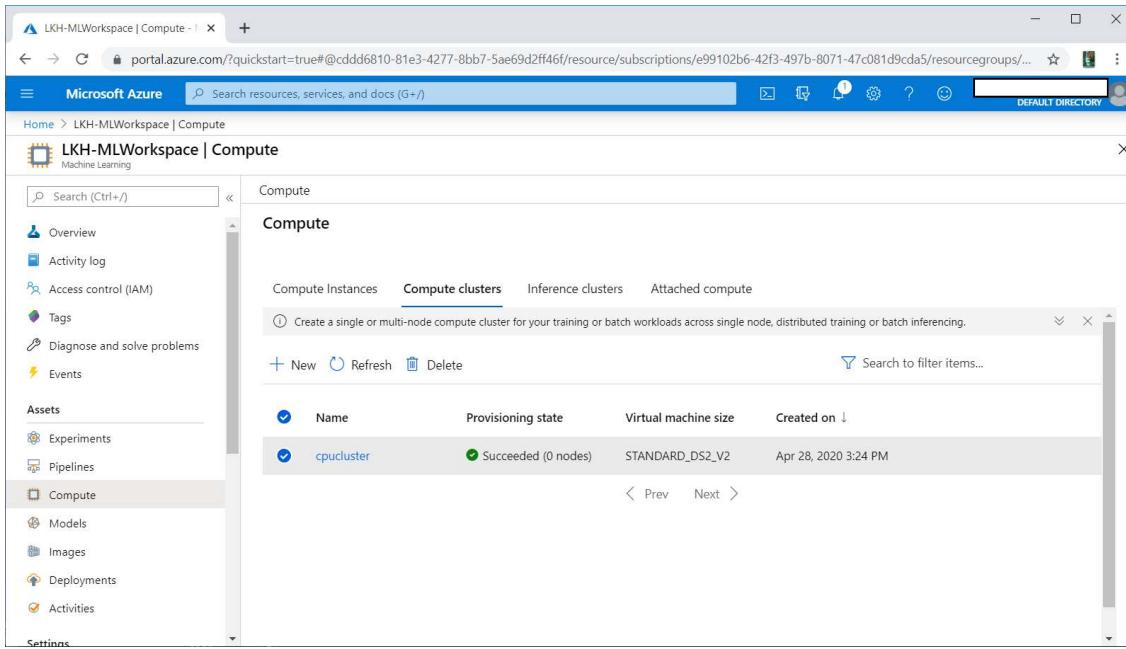


10. Once the resource is created, you should see its status as "Succeeded (0 nodes)"



The screenshot shows the Microsoft Azure Compute blade for the LKH-MLWorkspace. The left sidebar has 'Compute' selected under 'Assets'. The main area is titled 'Compute' and shows the 'Compute clusters' tab. A table lists one cluster named 'cpucluster' with a provisioning state of 'Succeeded (0 nodes)', a virtual machine size of 'STANDARD\_DS2\_V2', and a creation date of 'Apr 28, 2020 3:24 PM'. There are buttons for '+ New', 'Refresh', and 'Delete' at the top of the table.

11. If you want to delete this resource, select it then click "Delete".

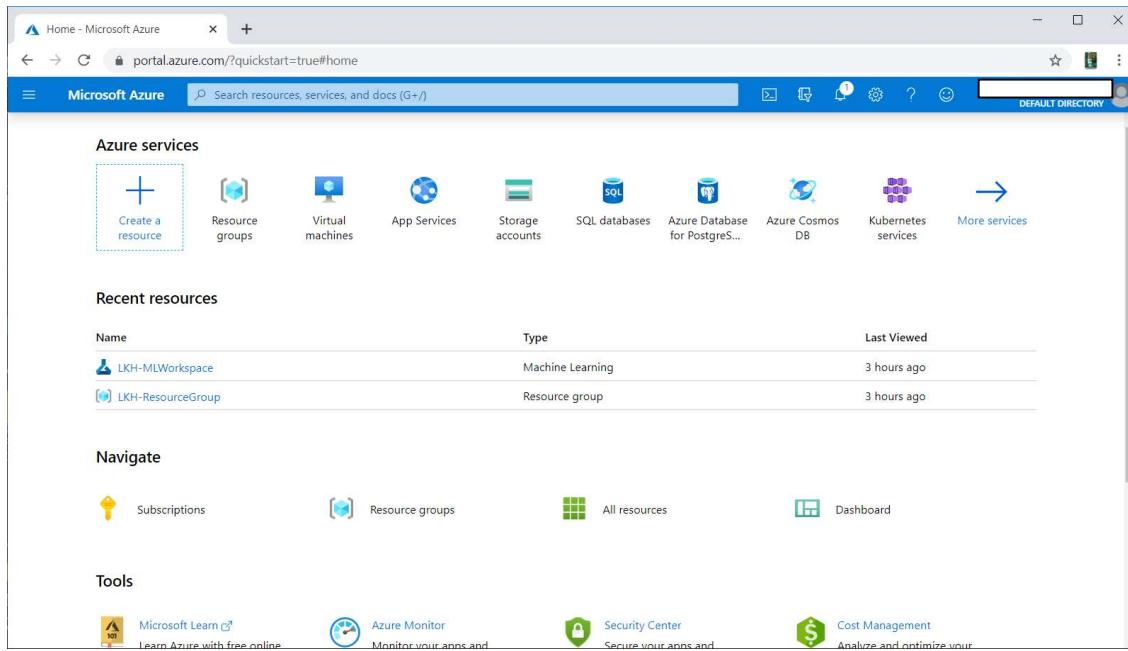


This screenshot is identical to the previous one, showing the Microsoft Azure Compute blade for the LKH-MLWorkspace. The 'Compute clusters' tab is selected, and the table shows the 'cpucluster' entry. However, the 'Delete' button is now highlighted in blue, indicating it is the active action for the selected item.

## Step 5: Upload data into Datastore

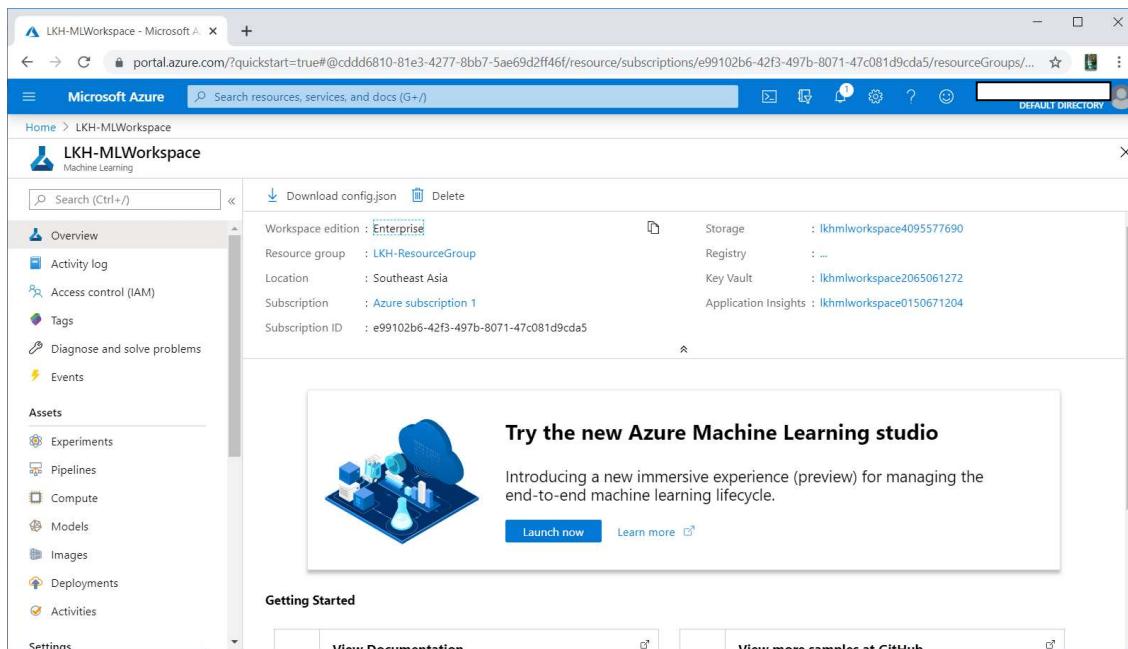
This step is to show how to use the data store in the Azure Machine Learning studio. The predict-employee-retention-part1-training.ipynb will illustrate how to use FileDataset and TabularDataset that is read from your datastore.

1. From Home, click on your Azure ML Workspace.



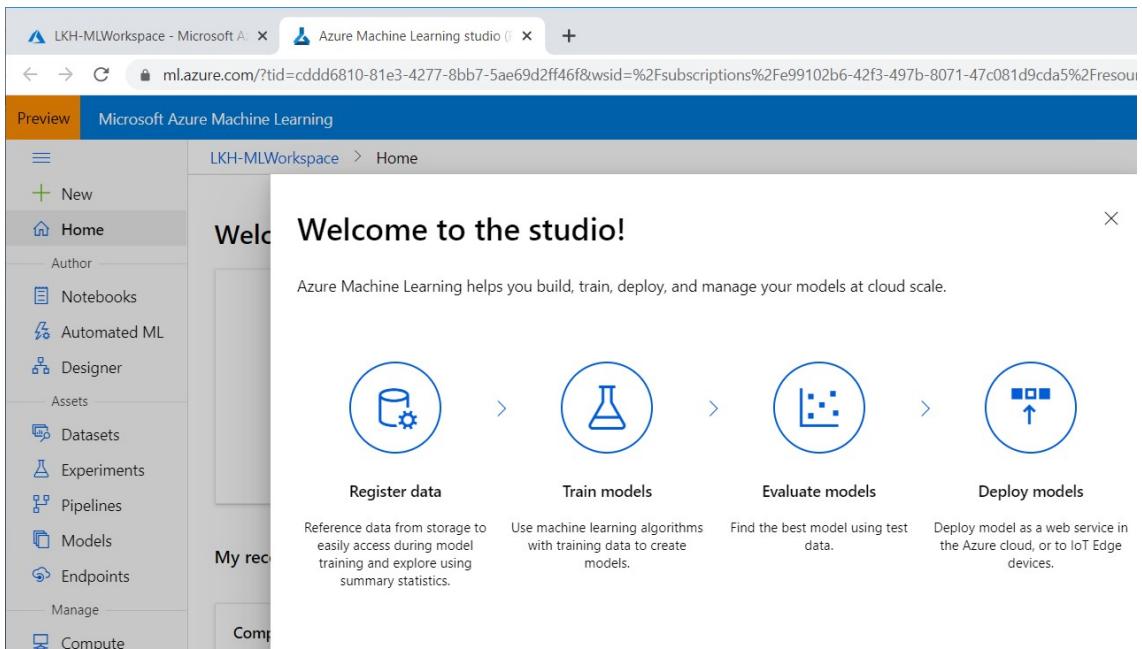
The screenshot shows the Microsoft Azure portal's Home page. At the top, there's a search bar and a 'DEFAULT DIRECTORY' dropdown. Below the header, there's a section titled 'Azure services' with various icons and links: 'Create a resource' (highlighted with a dashed box), 'Resource groups', 'Virtual machines', 'App Services', 'Storage accounts', 'SQL databases', 'Azure Database for PostgreSQL...', 'Azure Cosmos DB', 'Kubernetes services', and 'More services'. Under 'Recent resources', there are two entries: 'LKH-MLWorkspace' (Machine Learning) and 'LKH-ResourceGroup' (Resource group). Both were last viewed 3 hours ago. Below this is a 'Navigate' section with links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. At the bottom, there's a 'Tools' section with links for 'Microsoft Learn', 'Azure Monitor', 'Security Center', and 'Cost Management'.

2. Once you are in the Workspace, you will see the "Try the new Azure Machine Learning studio" message. Click on the "Launch now" button.

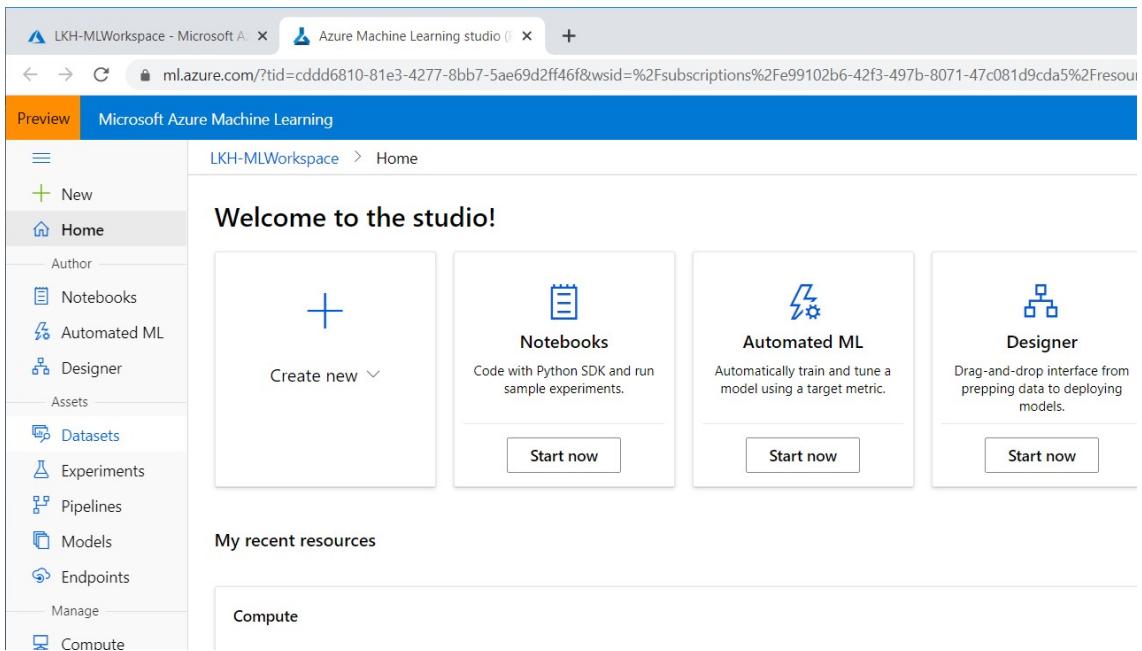


The screenshot shows the 'LKH-MLWorkspace' page within the Azure portal. The left sidebar includes 'Overview', 'Activity log', 'Access control (IAM)', 'Tags', 'Diagnose and solve problems', 'Events', 'Assets' (with sub-options like 'Experiments', 'Pipelines', 'Compute', 'Models', 'Images', 'Deployments', and 'Activities'), and 'Settings'. The main content area displays workspace details: Edition (Enterprise), Resource group (LKH-ResourceGroup), Location (Southeast Asia), Subscription (Azure subscription 1), and Subscription ID (e99102b6-42f3-497b-8071-47c081d9cda5). It also lists Storage (lkhmlworkspace4095577690), Registry (...), Key Vault (lkhmlworkspace2065061272), and Application Insights (lkhmlworkspace0150671204). A prominent message box says 'Try the new Azure Machine Learning studio' with a 'Launch now' button and a 'Learn more' link. Below this is a 'Getting Started' section with 'View Documentation' and 'View more complete examples' buttons.

3. You can "Start the tour" to learn how to navigate in Azure Machine Learning Studio or close it.

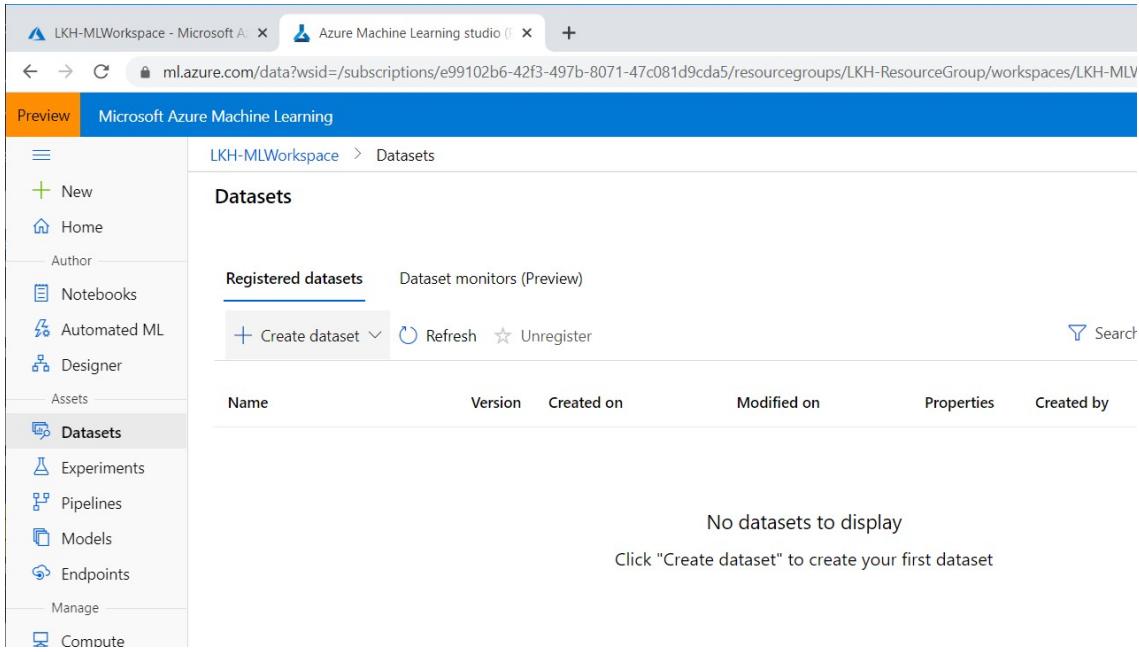


4. Click on the "Datasets" on the left menu.



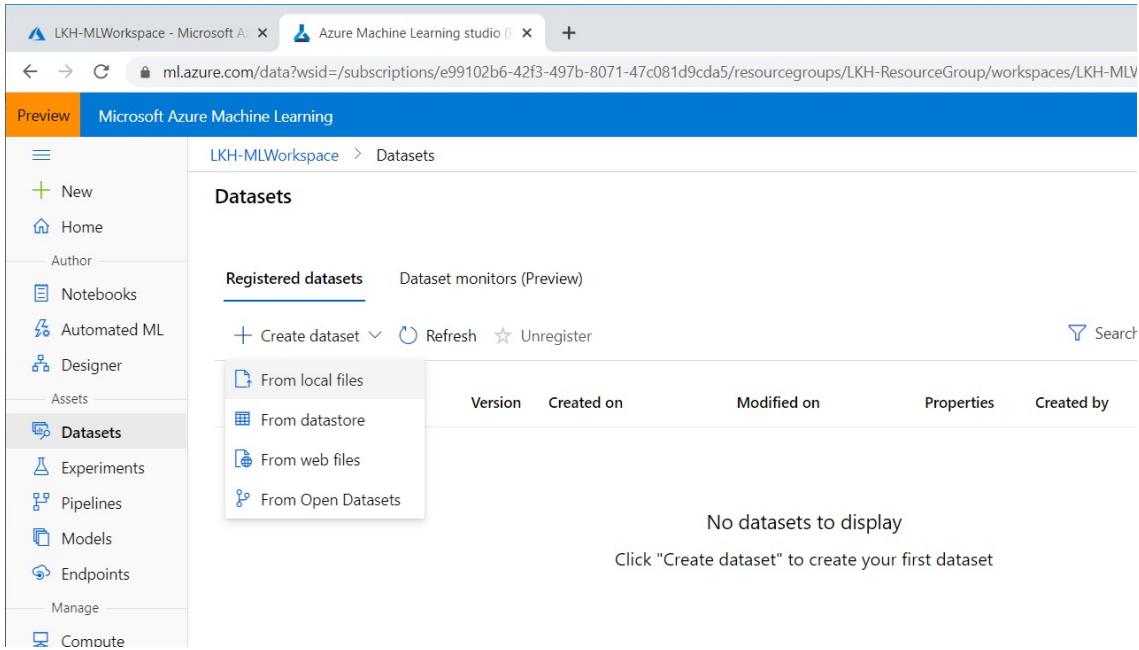
## Create FileDataset

5. Click "Create dataset".



The screenshot shows the Microsoft Azure Machine Learning studio interface. The left sidebar has a 'Datasets' section selected. The main area is titled 'Datasets' and shows a table header for 'Registered datasets'. Below the table, it says 'No datasets to display' and 'Click "Create dataset" to create your first dataset'. A prominent blue button labeled '+ Create dataset' is located at the top right of the table area.

6. This example will use the training-data.csv. Choose "From local files".



This screenshot is similar to the previous one, showing the 'Datasets' page in the Azure Machine Learning studio. The 'Create dataset' button is open, revealing four options: 'From local files', 'From datastore', 'From web files', and 'From Open Datasets'. The 'From local files' option is highlighted with a light gray background.

7. Enter the following details for creating a FileDataset, then click "Next".

Name : predict-employee-retention-training-data

Dataset type : File

Description : <can leave it blank>

The screenshot shows the 'Create dataset from local files' wizard in Azure Machine Learning studio. The left sidebar has icons for Home, Datasets, Models, Experiments, and more. The main area is titled 'Basic info'. It includes fields for 'Name' (predict-employee-retention-training-data), 'Dataset version' (1), 'Dataset type' (File), and 'Description' (Dataset description). A note at the top right says: 'Customers should not include personal data or other sensitive information in fields marked with because the content in these fields may be logged and shared across Microsoft systems to facilitate operations and troubleshooting.' A 'Learn more' link is provided.

8. By default, the "workspaceblobstore" is already created. Select it and scroll down further.

The screenshot shows the 'Datastore and file selection' step of the wizard. The left sidebar now shows 'Basic info' checked. The main area is titled 'Datastore and file selection' and includes a section for 'Select a datastore'. It shows a list with one item: 'Previously created datastore' (workspaceblobstore) selected. Below the list are 'Refresh' and 'Search to filter items...' buttons. At the bottom, there are navigation buttons: '< Prev' and 'Next >'.

9. Click "Browse" to select training-data.csv. Enter "predict-employee-retention" in the Upload path, then click "Next".

The screenshot shows the 'Create dataset from local files' wizard in the Azure Machine Learning studio. The left sidebar has icons for Home, Datasets, Experiments, and Models. The main panel title is 'Datastore and file selection'. On the left, a vertical navigation bar shows 'Basic info' (checked), 'Datastore and file selection' (checked), and 'Confirm details'. On the right, there's a section titled 'Select files for your dataset \*' with a note about supported file types. A 'Browse' button is shown, and below it is a table with one row for 'training-data.csv'. At the bottom are navigation arrows for 'Prev' and 'Next'.

File name	Size (MiB)	Upload %	Status
training-data.csv	0.5405		

10. Confirm the details then click "Create".

The screenshot shows the 'Create dataset from local files' wizard in the Azure Machine Learning studio. The left sidebar has icons for Home, Datasets, Experiments, and Models. The main panel title is 'Confirm details'. On the left, a vertical navigation bar shows 'Basic info' (checked), 'Datastore and file selection' (checked), and 'Confirm details' (checked). On the right, there are two side-by-side tables: 'Basic info' and 'Datastore and file selection'. The 'Basic info' table shows 'Name: predict-employee-retention-training-data', 'Dataset version: 1', and 'Dataset type: File'. The 'Datastore and file selection' table shows 'Datastore: workspaceblobstore', 'Selected files (1): training-data.csv', and 'Path: predict-employee-retention-training-data/2020\_085943.UTC/training'.

Basic info
Name predict-employee-retention-training-data
Dataset version 1
Dataset type File

Datastore and file selection
Datastore workspaceblobstore
Selected files (1) training-data.csv
Path predict-employee-retention-training-data/2020_085943.UTC/training

11. The created dataset will be shown in the Datasets screen.

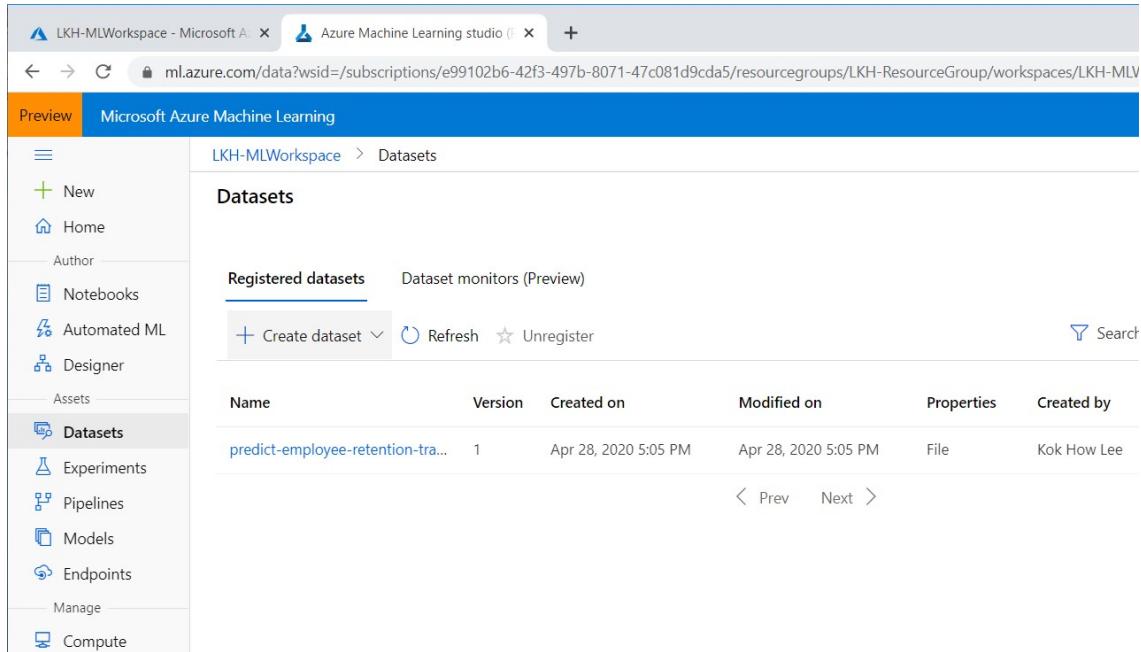
The screenshot shows the Microsoft Azure Machine Learning studio interface. The left sidebar has a 'Preview' tab selected, showing navigation links like 'New', 'Home', 'Author', 'Notebooks', 'Automated ML', 'Designer', 'Assets', 'Datasets' (which is highlighted), 'Experiments', 'Pipelines', 'Models', 'Endpoints', 'Manage', and 'Compute'. The main content area is titled 'Datasets' under 'LKH-MLWorkspace > Datasets'. It displays a table of 'Registered datasets' with one entry:

Name	Version	Created on	Modified on	Properties	Created by
predict-employee-retention-tra...	1	Apr 28, 2020 5:05 PM	Apr 28, 2020 5:05 PM	File	Kok How Lee

Below the table are buttons for 'Create dataset', 'Refresh', 'Unregister', and a 'Search' bar. Navigation arrows for 'Prev' and 'Next' are also present.

## Create TabularDataset

12. Click "Create dataset".

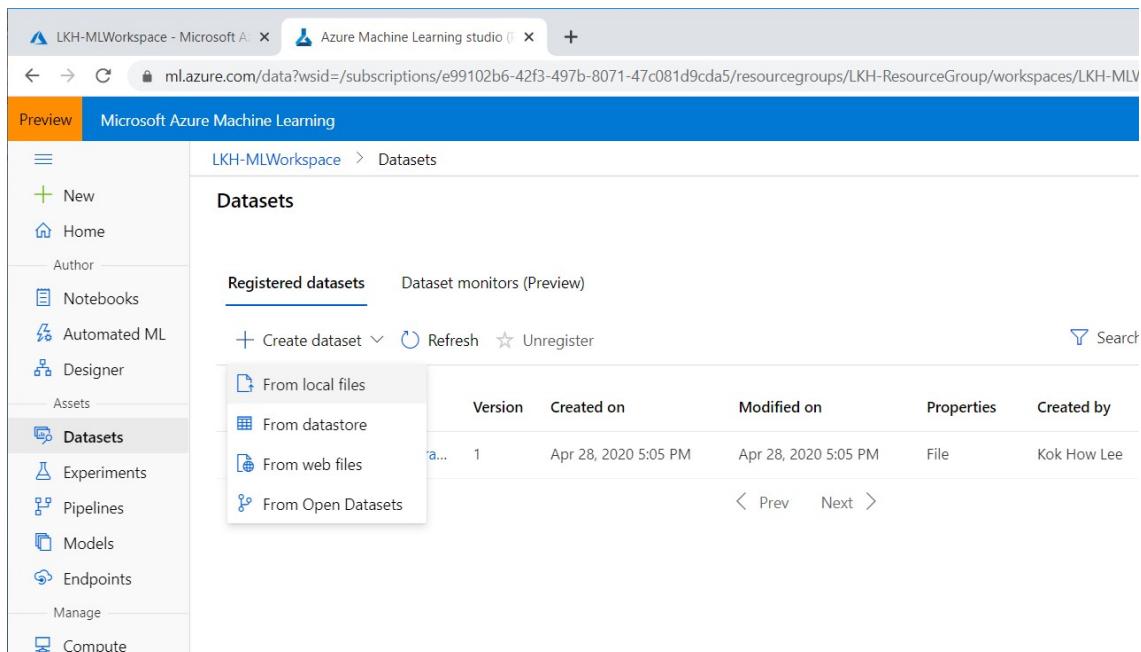


The screenshot shows the Azure Machine Learning studio interface. The left sidebar has a 'Datasets' section selected. The main area displays a table of registered datasets. One row is visible:

Name	Version	Created on	Modified on	Properties	Created by
predict-employee-retention-tra...	1	Apr 28, 2020 5:05 PM	Apr 28, 2020 5:05 PM	File	Kok How Lee

At the top right, there is a search bar and navigation links for 'Prev' and 'Next'.

13. Choose "From local files".



The screenshot shows the same Azure Machine Learning studio interface. The 'Create dataset' button in the top right of the main area is highlighted with a red box. A dropdown menu is open, showing four options:

- From local files
- From datastore
- From web files
- From Open Datasets

14. Enter the following details for creating a TabularDataset, then click "Next".

Name : predict-employee-retention-tabular

Dataset type : Tabular

Description : <can leave it blank>

The screenshot shows the 'Create dataset from local files' wizard in the Azure Machine Learning studio. The left sidebar lists steps: Basic info (selected), Datastore and file selection, Settings and preview, Schema, and Confirm details. The main area is titled 'Basic info' and contains fields: Name \* (predict-employee-retention-tabular), Dataset version (1), Dataset type \* (Tabular), and Description (Dataset description). A yellow callout box provides a note about not including sensitive information in certain fields.

15. Select "workspaceblobstore" and scroll down further.

The screenshot shows the 'Datastore and file selection' step in the wizard. The left sidebar shows 'Basic info' (checked) and 'Datastore and file selection' (selected). The main area shows a section 'Select a datastore \*' with a radio button for 'Previously created datastore'. Below it is a table listing datastores:

Name	Type	Stora...
workspaceblobstore	Azure Blob Storage	lkhml...

Buttons at the bottom include 'Refresh' and 'Search to filter items...', and navigation arrows '< Prev' and 'Next >'.

16. Click "Browse" to select training-data.csv. Enter "predict-employee-retention-tabular" in the Upload path, uncheck "Skip data validation", then click "Next".

The screenshot shows the 'Create dataset from local files' wizard in Microsoft Azure Machine Learning studio. The left sidebar has tabs for 'Basic info', 'Datastore and file selection' (which is selected), 'Settings and preview', 'Schema', and 'Confirm details'. The main area is titled 'Datastore and file selection' and contains a section for 'Select files for your dataset'. A 'Browse' button is shown with the message '1 files selected. Total size 0.5405 MiB. 0/1 files uploaded'. Below this is a table with one row for 'training-data.csv'. At the bottom is an 'Upload path' field containing 'predict-employee-retention-tabular'.

17. Select "All files have same headers" under Column Headers, then click "Next".

The screenshot shows the 'Create dataset from local files' wizard in Microsoft Azure Machine Learning studio. The left sidebar has tabs for 'Basic info', 'Datastore and file selection' (selected), 'Settings and preview' (selected), 'Schema', and 'Confirm details'. The main area is titled 'Settings and preview' and contains sections for 'File format' (Delimited), 'Delimiter' (Comma), 'Encoding' (UTF-8), 'Column headers' (set to 'All files have same headers'), and 'Skip rows' (None). Below these settings is a preview table showing columns: satisfaction\_level (0.0), last\_evaluation (0.0), number\_project (1), and average\_mont

18. The data types will be detected automatically (because previously uncheck "Skip data validation"). No changes required, click "Next".

The screenshot shows the 'Create dataset from local files' interface in Microsoft Azure Machine Learning studio. On the left, a vertical navigation bar lists various workspace items like Datasets, Experiments, and Models. The main panel has a title 'Create dataset from local files' and a sidebar with steps: 'Basic info', 'Datastore and file selection', 'Settings and preview', 'Schema', and 'Confirm details'. The 'Schema' step is currently selected. To its right, a table titled 'Schema' lists columns with their properties:

Include	Column name	Properties
<input checked="" type="checkbox"/>	Path	Not applicable to select... ▾
<input checked="" type="checkbox"/>	satisfaction_level	Not applicable to select... ▾
<input checked="" type="checkbox"/>	last_evaluation	Not applicable to select... ▾
<input checked="" type="checkbox"/>	number_project	Not applicable to select... ▾
<input checked="" type="checkbox"/>	average_montly_hours	Not applicable to select... ▾
<input checked="" type="checkbox"/>	time_spend_company	Not applicable to select... ▾
...	...	...

19. Confirm the details then click "Create".

The screenshot shows the 'Create dataset from local files' interface in Microsoft Azure Machine Learning studio, specifically the 'Confirm details' step. The left sidebar remains the same. The main panel has a title 'Confirm details' and displays the following configuration:

1	Dataset type Tabular
File settings	
File format	Delimited
Delimiter	Comma
Encoding	UTF-8
Column headers	All files have same headers
Skip rows	

20. The created dataset will be shown in the Datasets screen.

The screenshot shows the Microsoft Azure Machine Learning studio interface. The left sidebar has a 'Datasets' section selected. The main area displays a table of registered datasets:

Name	Version	Created on	Modified on	Properties	Created by
predict-employee-retention-ta...	1	Apr 28, 2020 5:26 PM	Apr 28, 2020 5:26 PM	Tabular	Kok How Lee
predict-employee-retention-tra...	1	Apr 28, 2020 5:05 PM	Apr 28, 2020 5:05 PM	File	Kok How Lee

A green success message at the top right says: "Success: predict-employee-retention-tabular dataset created successfully".

21. To view the dataset, click on the hyperlink.

The screenshot shows the Microsoft Azure Machine Learning studio interface, specifically the details page for the 'predict-employee-retention-tabular' dataset. The left sidebar has a 'Datasets' section selected. The main area shows the dataset details:

**predict-employee-retention-tabular** Version 1 (latest)

**Details**   [Consume](#)   [Explore](#)   [Models](#)

[Refresh](#)   [Generate profile](#)   [Unregister](#)   [New version](#)

Attributes	Tags
<b>Properties</b> Tabular	No data
<b>Description</b> --	
<b>Created by</b> Kok How Lee	
<b>Datastore</b> workspaceblobstore	
<b>Relative path</b>	

22. To see the data, click "Explore" tab.

The screenshot shows the Azure Machine Learning studio interface. The left sidebar has a 'Datasets' section selected. The main area shows the 'predict-employee-retention-tabular' dataset details. The 'Explore' tab is active, displaying a preview of the data. The preview table has 10 columns and 5 rows of sample data. The columns are labeled: 0.0 satisfaction\_level, 0.0 last\_evaluation, number\_project, average\_monthly\_hours, time\_spend\_company, and others partially visible. The data shows values like 0.38, 0.8, 0.11, 0.72, etc.

	0.0 satisfaction_level	0.0 last_evaluation	number_project	average_monthly_hours	time_spend_company	...
1	0.38	0.53	2	157	3	0
2	0.8	0.86	5	262	6	0
3	0.11	0.88	7	272	4	0
4	0.72	0.87	5	223	5	0

23. To see sample codes on how to retrieve this dataset, click "Consume" tab. This code is used in predict-employee-retention-part1-training.ipynb.

The screenshot shows the Azure Machine Learning studio interface. The left sidebar has a 'Datasets' section selected. The main area shows the 'predict-employee-retention-tabular' dataset details. The 'Consume' tab is active, displaying sample usage code. The code uses the azureml-core and azureml-datatools libraries to interact with the workspace and dataset.

```
# azureml-core of version 1.0.72 or higher is required
# azureml-datatools of version 1.1.34 or higher is required
from azureml.core import Workspace, Dataset

subscription_id = 'e99102b6-42f3-497b-8071-47c081d9cda5'
resource_group = 'LKH-ResourceGroup'
workspace_name = 'LKH-MLWorkspace'

workspace = Workspace(subscription_id, resource_group, workspace_name)

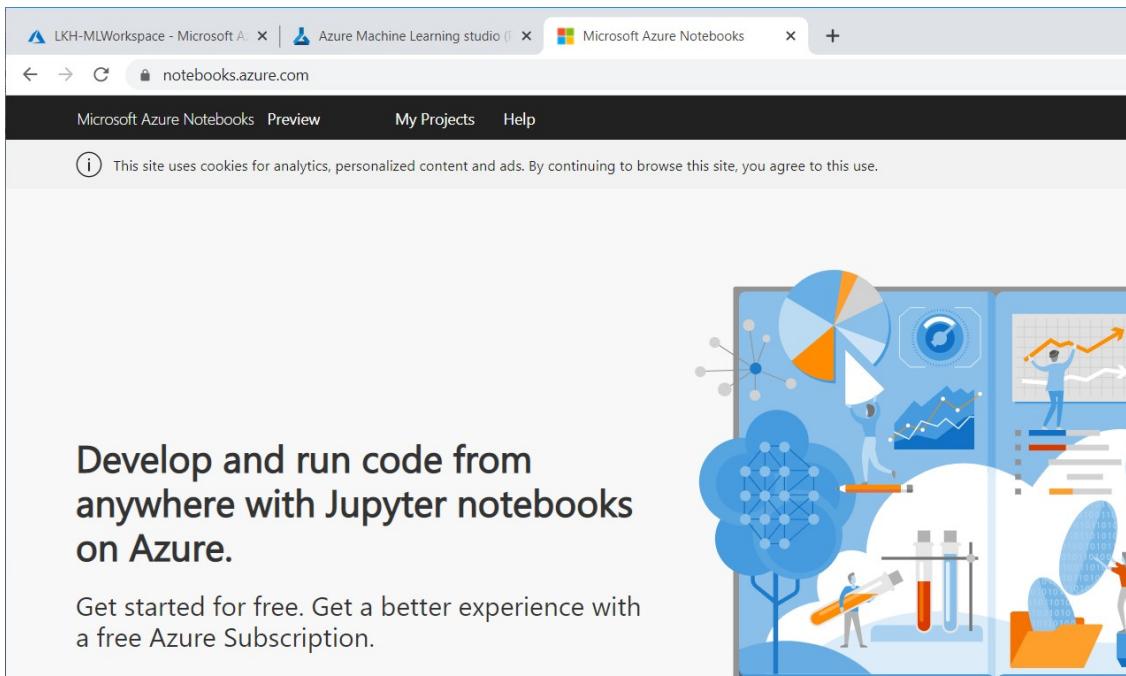
dataset = Dataset.get_by_name(workspace, name='predict-employee-retention-tabular')
dataset.to_pandas_dataframe()
```

## Step 6: Create Microsoft Azure Notebooks account to run Jupyter Notebooks.

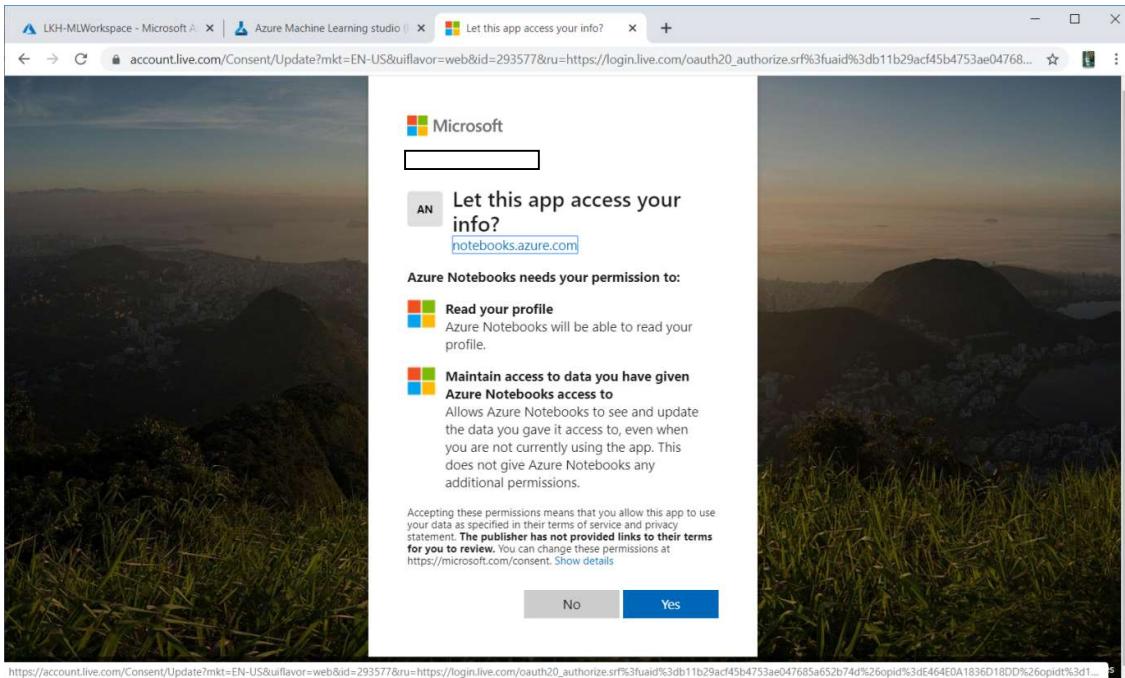
To run the Jupyter Notebooks, you can use Anaconda on your local machine or use the free Microsoft Azure Notebooks. Another way is to upload the Notebooks into Azure Machine Learning Studio, but you need to pay for computing resource to run the Notebooks. So for exploration purpose, you can choose to use the free Microsoft Azure Notebooks.

The Microsoft Azure Notebooks is pre-installed with Azure ML SDK in Python. You can use it directly to interact with your Azure ML Workspace that you have created.

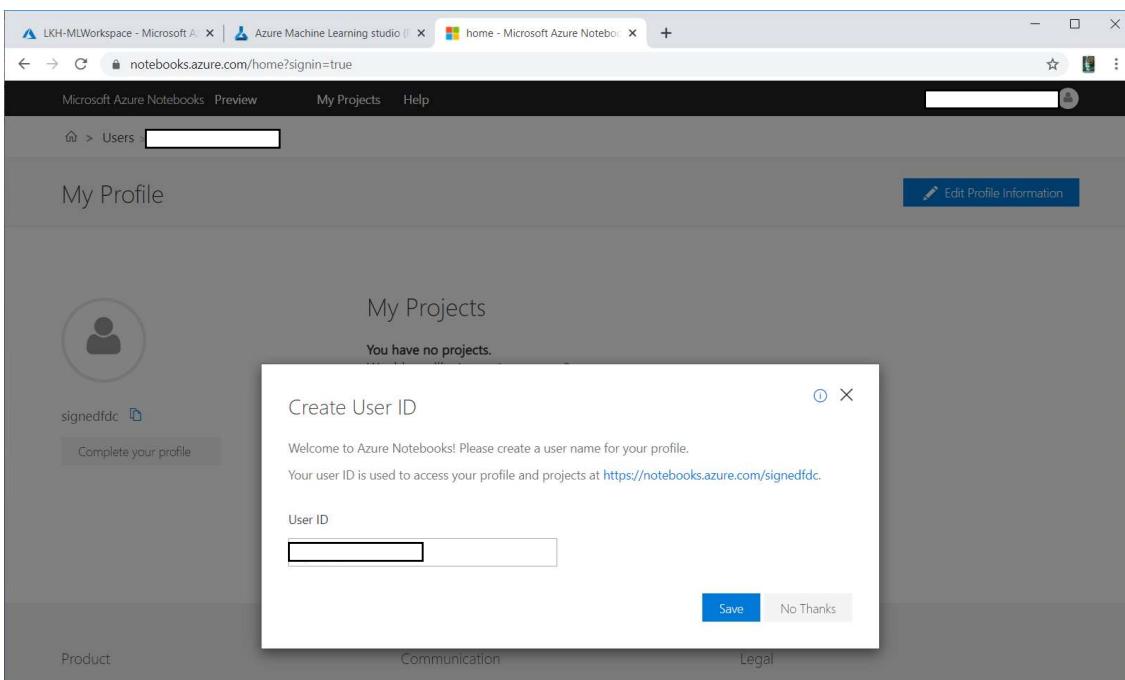
1. Go to <https://notebooks.azure.com/>



2. Click on "Sign In" on top right of the screen.
3. If you already login to the Azure Portal, it can auto sign in. Click "Yes" if you allow this.



4. Use the default User ID or change it. Click "Save".



## Step 7: Clone an Azure ML Notebook example to start with

The quickest way to explore is to clone an Azure ML sample to start with.

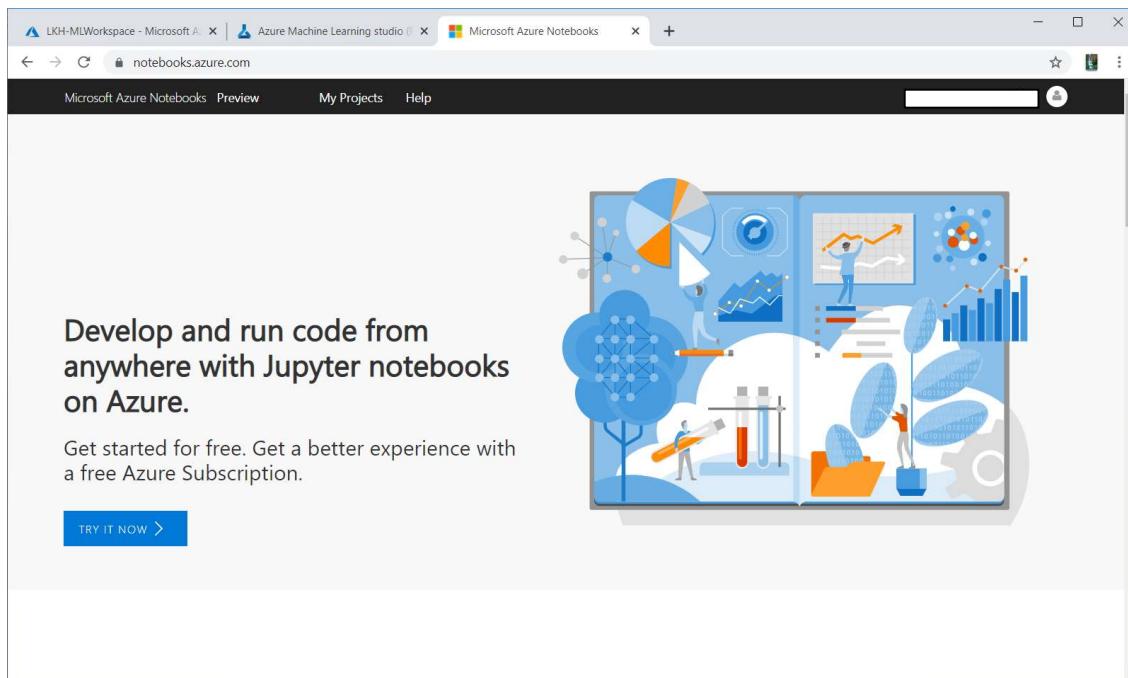
The Notebooks, predict-employee-retention-part1-training.ipynb and predict-employee-retention-part2-deploy.ipynb, are based on img-classification-part1-training.ipynb and img-classification-part2-deploy.ipynb respectively, but have modified to use a different use case, plus more details on how to use Azure SDK, and fixes some code errors due to difference in packages used.

Note:

Please run the codes in predict-employee-retention-part1-training.ipynb sequentially from top to bottom.

You can proceed to predict-employee-retention-part2-deploy.ipynb after complete part1, but it will use the last model version (created using cloud compute) after you completed "Option 3: Mount data file to a remote compute target for training script to access" example.

1. Go to <https://notebooks.azure.com/>



2. Scroll the page and you will see samples available. Click on the "Azure ML Getting Sta...".

The screenshot shows a web browser window with the URL [notebooks.azure.com](https://notebooks.azure.com). The page displays a grid of six project cards under the heading "Featured Projects".

- Azure ML Getting Sta...** (blue card): Azure Machine Learning is an Azure cloud service that you can use to develop and deploy machine learning models. Try it out. 19704 clones, 90 stars.
- Access Data in Noteb...** (purple card): This notebook provides examples of different ways to import data, all in a format that you can run and consume directly. 9479 clones, 101 stars.
- Python at Cambridge...** (orange card): An introduction to computing with Python for engineering and scientific applications that has been developed at Cambridge University. 2918 clones, 114 stars.
- Python for Data Anal...** (grey card): Materials and notebooks for "Python for Data Analysis" by Wes McKinney, the creator of Pandas. Published by O'Reilly Media.
- Data Science Handbo...** (blue card): This project contains the entire "Python Data Science Handbook" by Jake VanderPlas in the form of Azure Notebooks. 2005 clones, 101 stars.
- Linear Regression Tut...** (purple card): Learn about linear regression through a simple walkthrough that relates the rate of cricket chirps to temperature from The Song of

3. You will see the project files listed. Click on "Clone" to clone this project into your Notebook account.

The screenshot shows a web browser window with the URL [notebooks.azure.com/azureml/projects/azureml-getting-started](https://notebooks.azure.com/azureml/projects/azureml-getting-started). The page displays the contents of the "Getting Started With Azure Machine Learning Service" project.

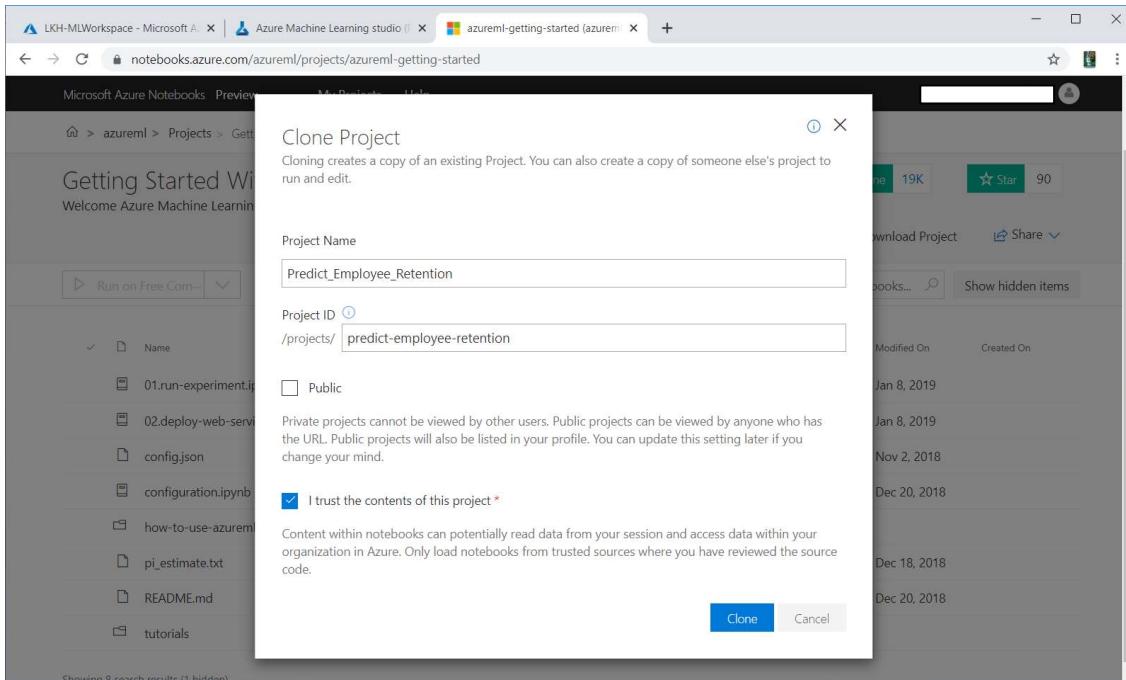
Project details:  
Getting Started With Azure Machine Learning Service  
Welcome Azure Machine Learning service through Azure Notebooks

Project stats:  
Clone 19K | Star 90 | Download Project | Share

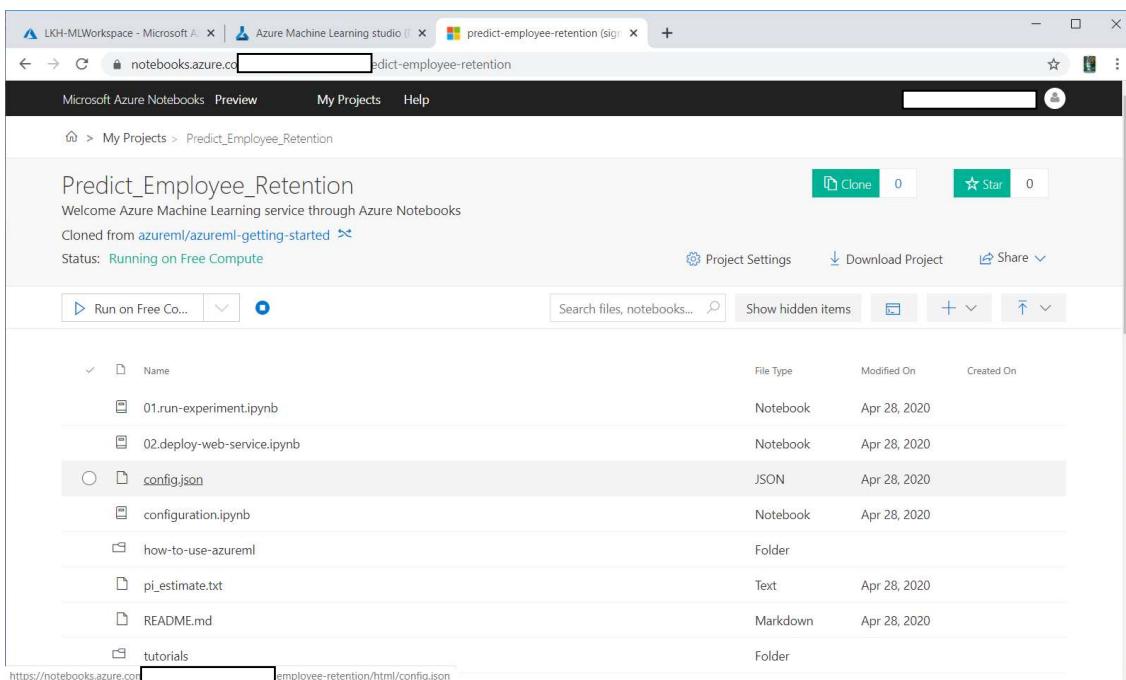
File list:  
Run on Free Compute ▾  
Name | File Type | Modified On | Created On  
01.run-experiment.ipynb | Notebook | Jan 8, 2019 |  
02.deploy-web-service.ipynb | Notebook | Jan 8, 2019 |  
config.json | JSON | Nov 2, 2018 |  
configuration.ipynb | Notebook | Dec 20, 2018 |  
how-to-use-azureml | Folder | Dec 20, 2018 |  
pi\_estimate.txt | Text | Dec 18, 2018 |  
README.md | Markdown | Dec 20, 2018 |  
tutorials | Folder |

Showing 8 search results (1 hidden)

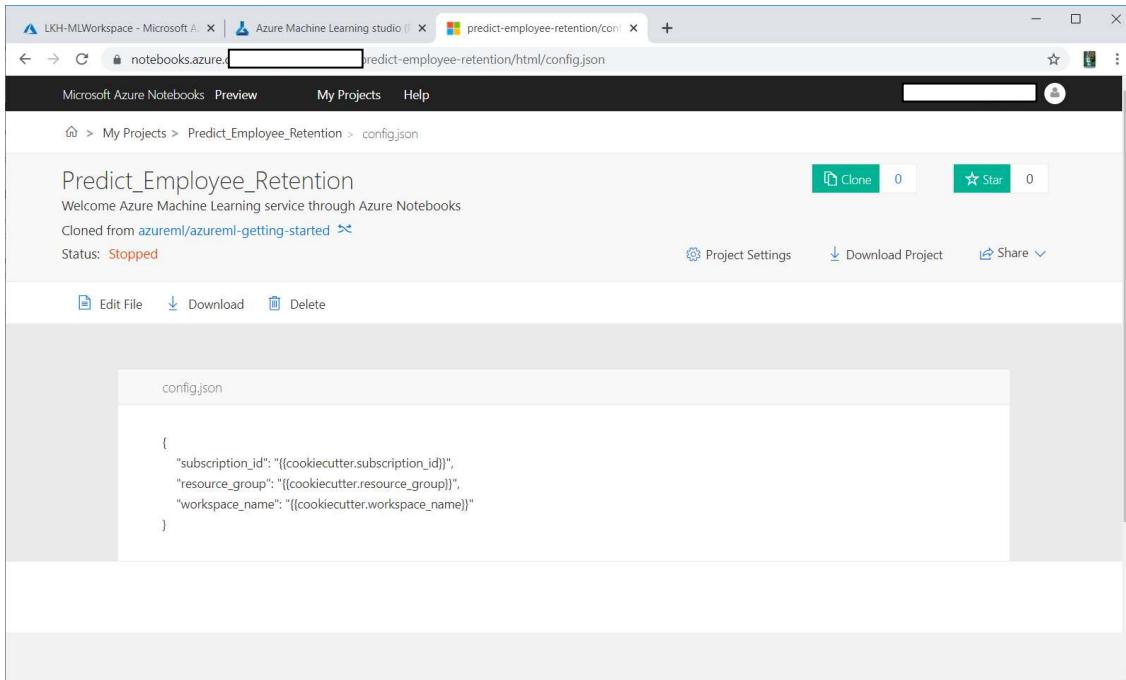
4. Change the default project name (e.g. Predict\_Employee\_Retention) and project ID (e.g predict-employee-retention), check "I trust the contents of these files". Click "Clone".



5. Notice the project is created under "My Projects" folder.



6. Click to view the config.json file. Click "Edit File", you need to provide the subscription\_id, resource\_group, workspace\_name. You can get the info from the Azure Portal.



7. Go to your Azure ML workspace to get the info: Your workspace name, Resource group, Subscription ID.

The screenshot shows the Azure portal's 'LKH-MLWorkspace' details page. The left sidebar lists workspace assets like Experiments, Pipelines, Compute, Models, Images, Deployments, and Activities. The main content area displays workspace properties:

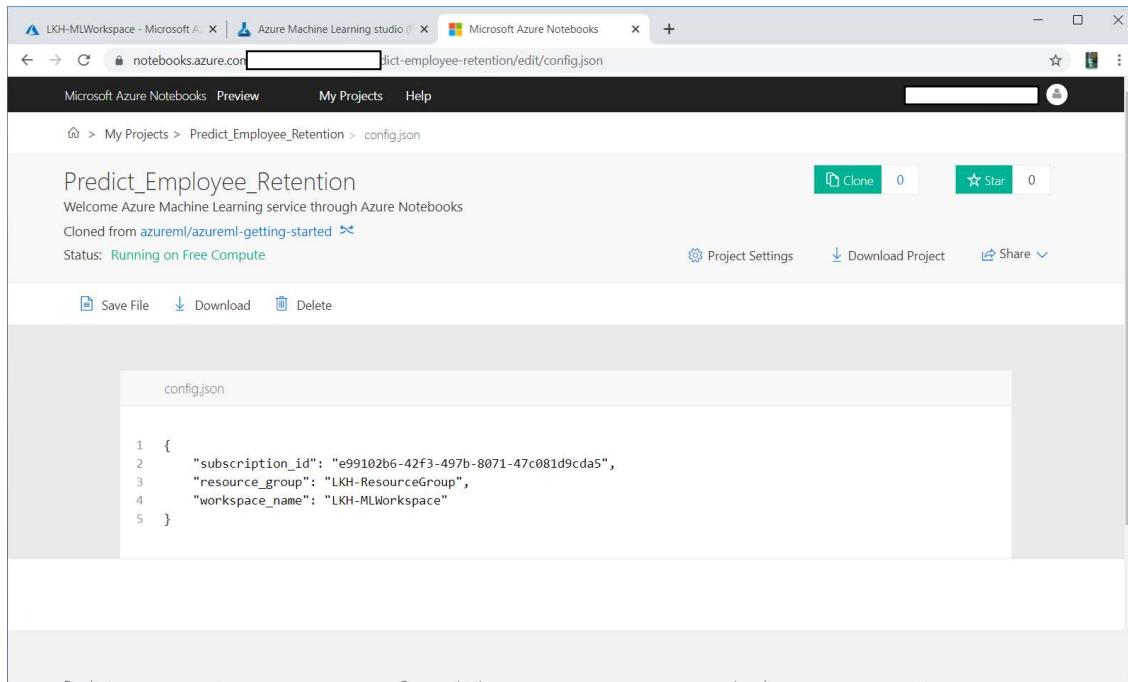
Workspace edition :	Enterprise	Storage :	lkhmlworkspace4095577690
Resource group :	LKH-ResourceGroup	Registry :	...
Location :	Southeast Asia	Key Vault :	lkhmlworkspace2065061272
Subscription :	Azure subscription 1	Application Insights :	lkhmlworkspace0150671204
Subscription ID :	e99102b6-42f3-497b-8071-47c081d9cda5		

Below the properties, there is a promotional section for the new Azure Machine Learning studio:

**Try the new Azure Machine Learning studio**  
Introducing a new immersive experience (preview) for managing the end-to-end machine learning lifecycle.  
[Launch now](#) [Learn more](#)

At the bottom, there are links for 'View Documentation' and 'View more samples at GitHub'.

8. Once your copied the info into the JSON file, click "Save File".

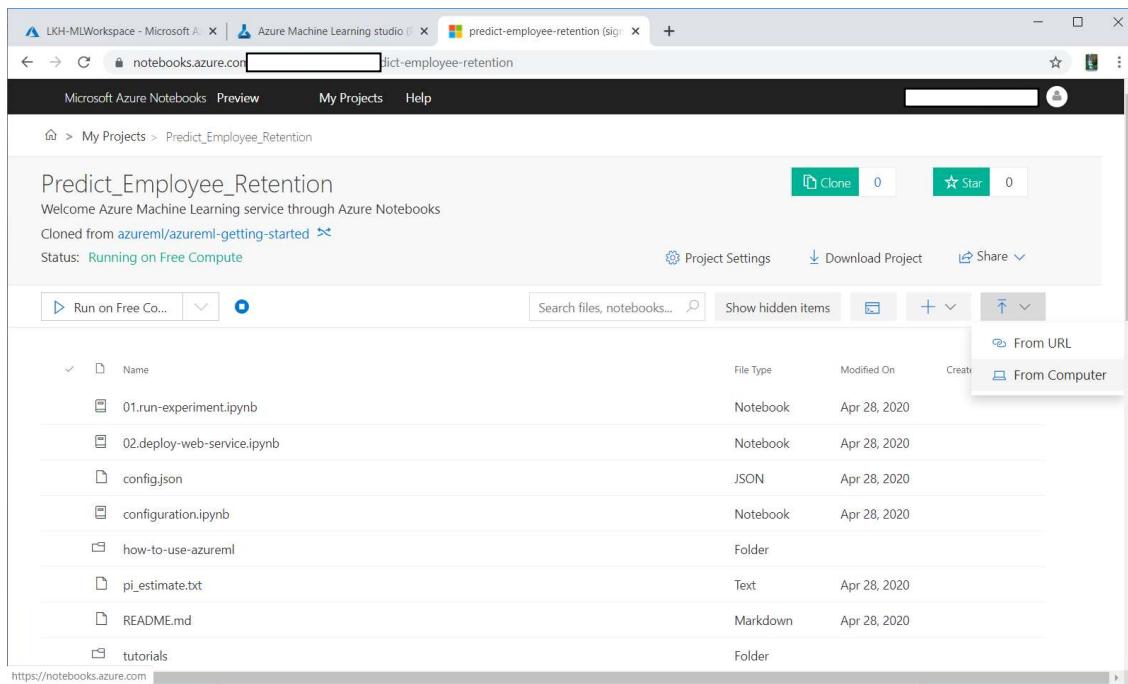


The screenshot shows a Microsoft Azure Notebooks interface. The URL in the address bar is `notebooks.azure.com/predict-employee-retention/edit/config.json`. The page title is "Predict\_Employee\_Retention". The content area displays the "config.json" file with the following JSON code:

```
1 {  
2     "subscription_id": "e99102b6-42f3-497b-8071-47c081d9cda5",  
3     "resource_group": "LKH-ResourceGroup",  
4     "workspace_name": "LKH-MLWorkspace"  
5 }
```

Below the code, there are buttons for "Save File", "Download", and "Delete".

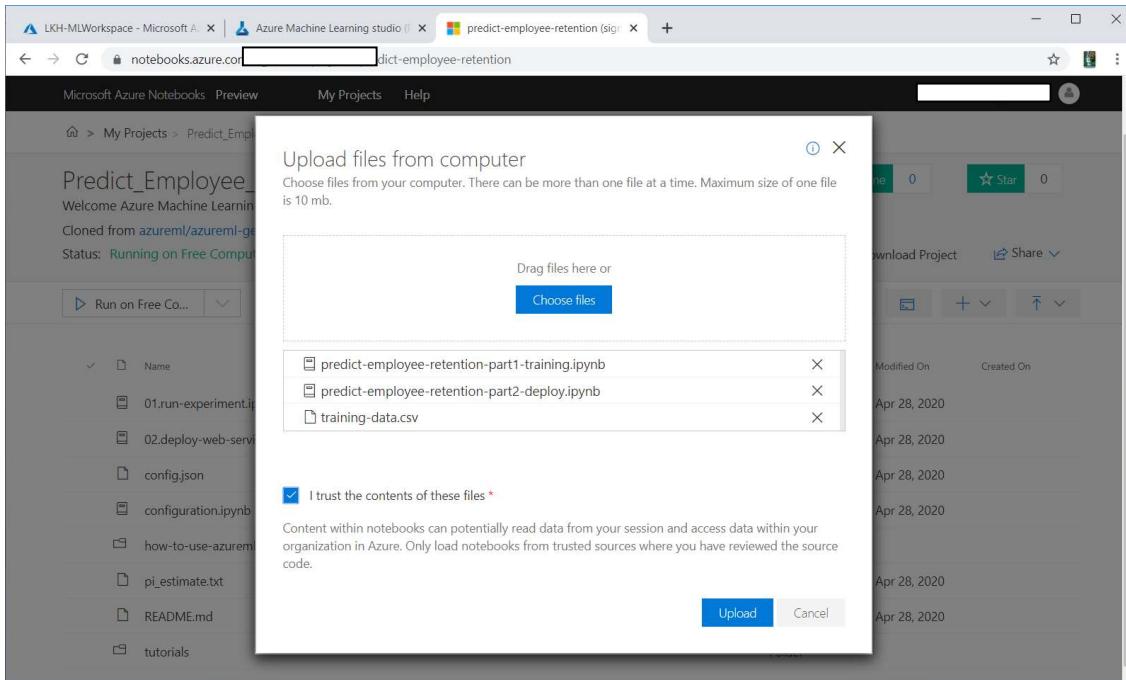
9. Now to upload the tutorials. Click on  and select "From Computer" to upload file.



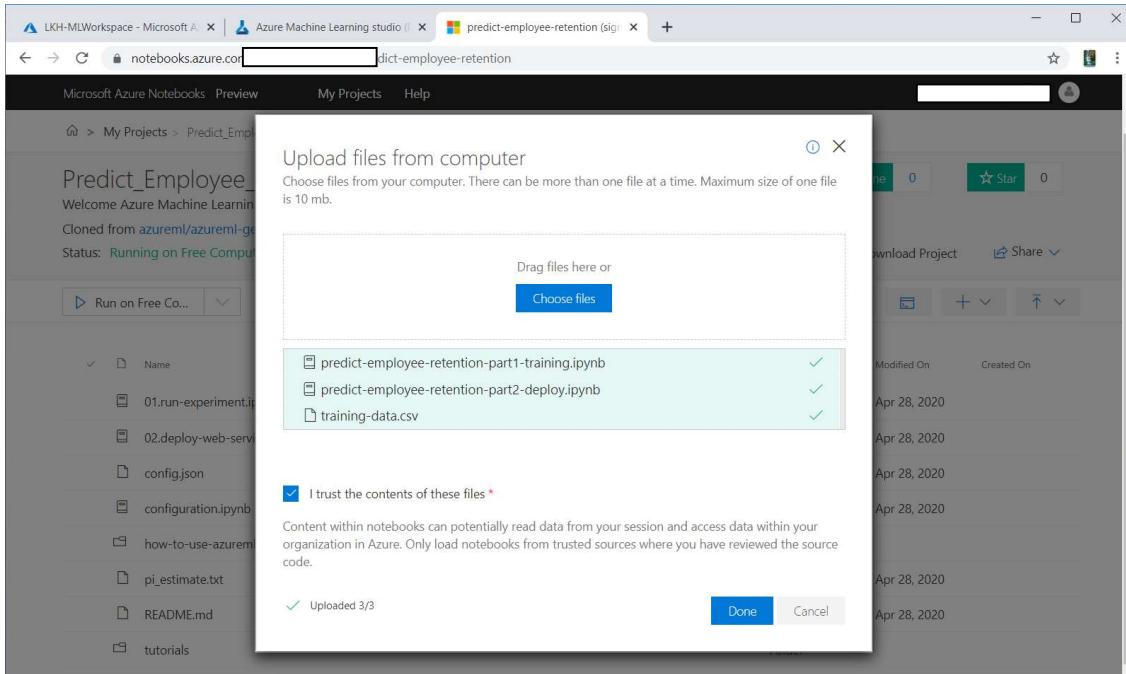
The screenshot shows the same Microsoft Azure Notebooks interface as the previous one, but now with a focus on the "tutorials" folder. The URL in the address bar is `notebooks.azure.com/predict-employee-retention`. The "tutorials" folder is selected in the left sidebar. The top navigation bar includes "Run on Free Compute", "Search files, notebooks...", "Show hidden items", and an "Upload" button (). A context menu is open over the "tutorials" folder, with options "From URL" and "From Computer" visible.

Name	Type	Modified On	Created On
01.run-experiment.ipynb	Notebook	Apr 28, 2020	
02.deploy-web-service.ipynb	Notebook	Apr 28, 2020	
config.json	JSON	Apr 28, 2020	
configuration.ipynb	Notebook	Apr 28, 2020	
how-to-use-azurerm	Folder		
pi_estimate.txt	Text	Apr 28, 2020	
README.md	Markdown	Apr 28, 2020	
tutorials	Folder		

10. Click "Choose files" to select files to upload. Check "I trust the contents of these files" then click "Upload".



11. Click "Done" to close the upload files dialog.



## 12. Click to launch predict-employee-retention-part1-training.ipynb

The screenshot shows the Microsoft Azure Notebooks interface. At the top, there are tabs for 'LKH-MLWorkspace - Microsoft A.', 'Azure Machine Learning studio ()', and 'predict-employee-retention (signed)'. The main area displays the 'Predict\_Employee\_Retention' project. The notebook 'predict-employee-retention-part1-training.ipynb' is highlighted with a blue border. The interface includes a search bar, a file list, and various project management buttons like 'Clone', 'Star', 'Project Settings', 'Download Project', and 'Share'.

## 13. Click Run to step through each code cell.

Note: the Kernel is selected as "Python 3.6" - this version comes with Azure ML SDK. If not selected as default, you can change the kernel from Kernel > Change kernel

The screenshot shows a Jupyter notebook interface. The title bar indicates it's running on 'predict-employee-retention-part1-training.ipynb'. The notebook content starts with a section titled 'Tutorial #1: Model training with Azure Machine Learning'. It provides instructions for creating a workspace in the Azure portal and explores training a machine learning model using scikit-learn and Azure Machine Learning. The interface includes a toolbar with various icons for file operations, a code editor, and a help section.

14. Notice the SDK version is displayed in this code cell. You should see the current version used in Azure Notebooks is 1.0.72.

Note: if install the Azure ML SDK on your local machine, the latest available version is currently 1.3.0.

The screenshot shows a Microsoft Azure Notebooks Jupyter interface. The title bar includes tabs for 'LKH-MLWorkspace - Microsoft A...', 'Azure Machine Learning studio (...', and 'predict-employee-retention (sig...'. The main area has a header 'Microsoft Azure Notebooks Preview My Projects Help' and a toolbar with 'File Edit View Insert Cell Kernel Widgets Help' buttons. Below the toolbar is a 'Code' button. The notebook content starts with a note about installing the Azure ML SDK:

- pip install azureml-sdk[notebooks,automl]
- pip install azureml-databricks[pandas]
- conda update conda

OR you can use a [free Microsoft Azure Notebooks](#) to run this notebook if you don't have Anaconda.

### Import Azure Machine Learning SDK for Python

This step is to test you have installed Azure Machine Learning SDK for Python. Most of the coding will required the use of the Azure ML SDK.

Display the Azure Machine Learning SDK version.

```
In [1]: import azureml.core  
# check core SDK version number (need Python 3.6 kernel if you run this in Microsoft Azure Notebooks)  
print("Azure ML SDK Version: ", azureml.core.VERSION)
```

Azure ML SDK Version: 1.0.72

### Connect to workspace

Create a workspace object from the existing workspace. `workspace.from_config()` reads the file `config.json` and loads the details into an object named `workspace`.

If you see this message: "Performing interactive authentication. Please follow the instructions on the terminal. To sign in, use a web browser to open the page <https://microsoft.com/device/login> and enter the code <token> to authenticate."

15. Run the code cell under "Connect to workspace" to make sure you are connected to your Azure ML workspace.

The screenshot shows a Microsoft Azure Notebooks Jupyter interface. The title bar includes tabs for 'LKH-MLWorkspace - Microsoft A...', 'Azure Machine Learning studio (...', and 'predict-employee-retention (sig...'. The main area has a header 'Microsoft Azure Notebooks Preview My Projects Help' and a toolbar with 'File Edit View Insert Cell Kernel Widgets Help' buttons. Below the toolbar is a 'Code' button. The notebook content starts with the previous message about the Azure ML SDK version:

Azure ML SDK Version: 1.0.72

### Connect to workspace

Create a workspace object from the existing workspace. `Workspace.from_config()` reads the file `config.json` and loads the details into an object named `workspace`.

If you see this message: "Performing interactive authentication. Please follow the instructions on the terminal. To sign in, use a web browser to open the page <https://microsoft.com/device/login> and enter the code <token> to authenticate."

Click on the link and use the <token> given to authenticate. After authenticated, run this script again to get load the Workspace.</token></token>

```
In [3]: # Load workspace configuration from the config.json file in the current folder.  
from azureml.core import Workspace  
workspace = Workspace.from_config()  
print(workspace.name, workspace.location, workspace.resource_group, workspace.location, sep='\t')  
LKH-MLWorkspace southeastasia LKH-ResourceGroup southeastasia
```

### Create experiment

Create an experiment to track the runs in your workspace. A workspace can have multiple experiments.

```
In [ ]: from azureml.core import Experiment  
experiment_name = 'predict-employee-retention'
```

16. Run the code cell under "Create experiment".

```
from azureml.core import Experiment
experiment_name = 'predict-employee-retention'
exp = Experiment(workspace=workspace, name=experiment_name)
```

**Create experiment**  
Create an experiment to track the runs in your workspace. A workspace can have multiple experiments.

In [7]:

```
# Connect to datastore
from azureml.core import Datastore
datastore = Datastore.get(workspace, datastore_name='workspaceblobstore')
```

**Connect to datastore**  
From your machine learning Workspace, launch Azure Machine Learning studio. Click "Datastores" (under "Manage" section) on the left menu.  
This tutorial will use the default datastore "workspaceblobstore" to store the datasets.

**Upload data files into a Datastore**  
This tutorial will use datasets created in the "workspaceblobstore" datastore.

17. You will be able to find the Experiment created in the Azure Machine Learning Studio.

Experiment	Latest run	Last submitted	Created	Created by	Run types
predict-employee-retention	--	Apr 28, 2020 7:33...	Apr 28, 2020 7:33...	Kok How Lee	

18. Run the code cell under "Connect to datastore".

```
In [9]: # Connect to datastore
from azureml.core import Datastore
datastore = Datastore.get(workspace, datastore_name='workspaceblobstore')
```

The notebook interface shows the following sections:

- Connect to datastore**: Instructions to launch Azure Machine Learning studio and click "Datastores" under "Manage".
- Upload data files into a Datastore**: Instructions to use training-data.csv to create datasets.
- Explore ways to download data from datastore**: A link to explore download methods.

19. Under the "Explore ways to download data from datastore", you need to make sure the file path matches your uploaded file path. You can check the path from the Datasets in Azure Machine Learning Studio.

```
In [13]: from azureml.core import Dataset
import os

# 3 possible ways to get data from datastore.
# 1. Download data file from datastore using "Relative path"
# Notice the downloaded file will be stored in a folder that has structure similar to the relative path.
temp_folder = os.path.join(os.getcwd(), "temp")
training_data_file = 'predict_employee-retention/04-28-2020_085943_UTC/training-data.csv' # You can download file using the exact
os.makedirs(temp_folder, exist_ok=True)
datastore.download(target_path=temp_folder,
                   prefix=training_data_file,
                   show_progress=True,
                   overwrite=True)

# 2. Download data file from datastore as FileDataset
# Below is generated sample code found in "Microsoft Azure Machine Learning > Datasets > (choose your dataset) > Consume". You can
# azureml-core of version 1.0.72 or higher is required
from azureml.core import Workspace, Dataset

subscription_id = 'XXX'
resource_group = 'XXX'
workspace_name = 'XXX'
```

The screenshot shows the Azure Machine Learning studio interface. On the left, there is a navigation sidebar with options like New, Home, Author, Notebooks, Automated ML, Designer, Assets, Datasets, Experiments, Pipelines, Models, Endpoints, Compute, Datastores, and Data Labeling. The 'Datasets' option is selected. The main panel displays a dataset named 'predict-employee-retention-training-data'. It shows the following details:

- Properties:** Version 1 (latest)
- Description:** --
- Created by:** Kok How Lee
- Datastore:** workspaceblobstore
- Relative path:** predict-employee-retention/04-28-2020\_085943.UTC/training-data.csv
- Files in dataset:** 1
- Total size of files in dataset:** 553.5 KIB
- Current version:** 1
- Latest version:** 1
- Created time:** Apr 28, 2020 5:05 PM
- Modified time:** Apr 28, 2020 5:05 PM

20. Once you have made the changes, run the code cell to see its output.

The screenshot shows the Microsoft Azure Notebooks interface. A Jupyter notebook cell is running, titled 'Predict\_Employee\_Retention'. The code cell contains the following Python code:

```
# 3. Download data from datastore as TabularDataset.
# Below is generated sample code found in "Microsoft Azure Machine Learning > Datasets > (choose your dataset) > Consume". You can copy and paste this code into a new cell and run it.
# azureml-core of version 1.0.72 or higher is required
# azureml-databricks[pandas] of version 1.1.34 or higher is required
from azureml.core import Workspace, Dataset

subscription_id = 'XXX'
resource_group = 'XXX'
workspace_name = 'XXX'
workspace = Workspace(subscription_id, resource_group, workspace_name)

dataset = Dataset.get_by_name(workspace, name='predict-employee-retention-tabular')
df = dataset.to_pandas_dataframe()

display(df.head())
```

The output of the code cell shows the downloaded data:

```
Downloading predict-employee-retention/04-28-2020_085943.UTC/training-data.csv
Downloaded predict-employee-retention/04-28-2020_085943.UTC/training-data.csv, 1 files out of an estimated total of 1
```

	satisfaction_level	last_evaluation	number_project	average_monthly_hours	time_spend_company	Work_accident	left	promotion_last_5years	sales	salary
0	0.38	0.53	2	157	3	0	1	0	sales	low
1	0.80	0.86	5	262	6	0	1	0	sales	medium
2	0.11	0.88	7	272	4	0	1	0	sales	medium
3	0.72	0.87	5	223	5	0	1	0	sales	low
4	0.37	0.52	2	159	3	0	1	0	sales	low

21. You should be able to run through all the remaining code cells in the tutorial. Do note some cells may take a while to execute.

22. After you have registered the model, you can find it under "Models" in the Azure Machine Learning Studio.

**Register Model**

The last step in the training script wrote the file outputs/sklearn\_mnist\_model.pkl in a directory named outputs in the VM of the cluster where the job is run. outputs is a special directory in that all content in this directory is automatically uploaded to your workspace. This content appears in the run record in the experiment under your workspace. So the model file is now also available in your workspace.

You can see files associated with that run:

```
In [32]: print(run.get_file_names())
['azureml-logs/60_control_log.txt', 'azureml-logs/70_driver_log.txt', 'logs/azureml/1055_azureml.log', 'outputs/predict-employee-retention-model.pkl']
```

Register the model in the workspace, so that you or other collaborators can later query, examine, and deploy this model:

```
In [33]: # register model
model = run.register_model(model_name='predict-employee-retention',
                           model_path='outputs/predict-employee-retention-model.pkl')
print(model.name, model.id, model.version, sep='\t')
predict-employee-retention predict-employee-retention:1 1
```

**See next tutorial: Deploy the model**

**Model List**

Name	Version	Experiment	Run ID	Created on	Tags
predict-employee-retention	1	predict-employee-retention	predict-employee-retention_15...	Apr 28, 2020 8:17 P...	Kok H...

23. When you run the code under "Create or Attach existing compute resource", it should be able to detect an existing compute resource "cpucluster" which you have already created, thus it will not create a new one.

```

In [43]: from azureml.core.compute import AmlCompute
from azureml.core.compute import ComputeTargetException
import os

# choose a name for your cluster
compute_name = os.environ.get("AML_COMPUTE_CLUSTER_NAME", "cpucluster")
compute_min_nodes = os.environ.get("AML_COMPUTE_CLUSTER_MIN_NODES", 0)
compute_max_nodes = os.environ.get("AML_COMPUTE_CLUSTER_MAX_NODES", 2)

# This example uses CPU VM. For using GPU VM, set SKU to STANDARD_NC6
vm_size = os.environ.get("AML_COMPUTE_CLUSTER_SKU", "STANDARD_DS2_V2")

if compute_name in workspace.compute_targets:
    compute_target = workspace.compute_targets[compute_name]
    if compute_target and type(compute_target) is AmlCompute:
        print('found compute target. just use it. ' + compute_name)
    else:
        print('creating a new compute target...')
        provisioning_config = AmlCompute.provisioning_configuration(vm_size,
                                                                    min_nodes=compute_min_nodes,
                                                                    max_nodes=compute_max_nodes)

# create the cluster
compute_target = ComputeTargetException.create(
    workspace, compute_name, provisioning_config)

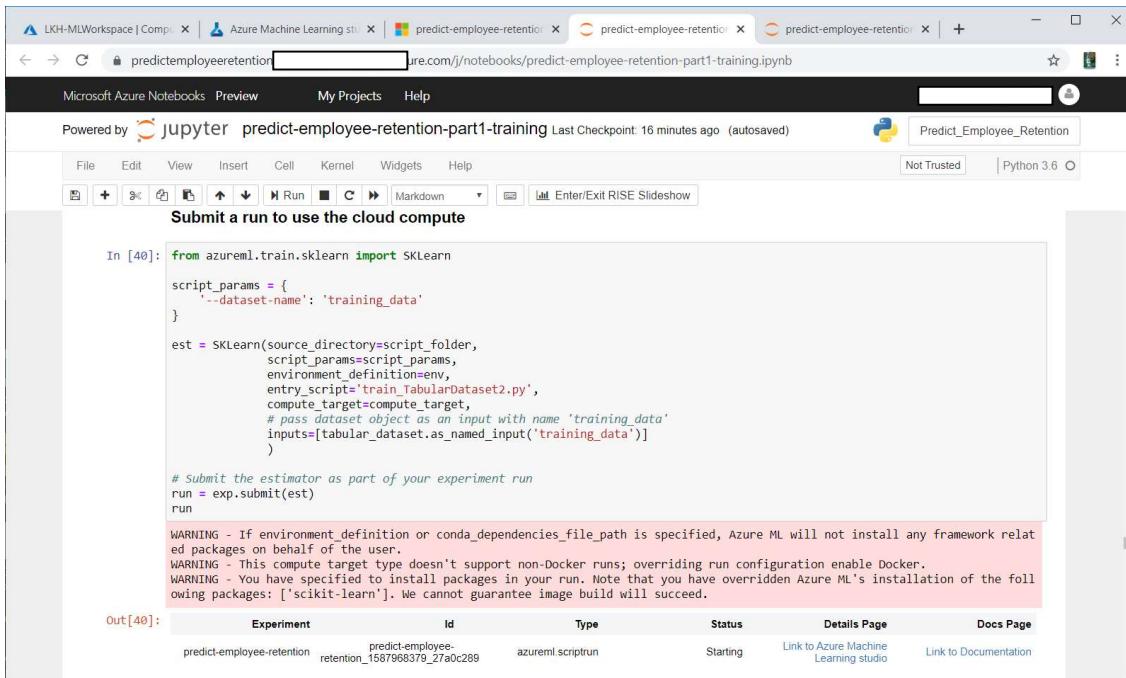
# poll for a minimum number of nodes and for a specific timeout.
# If no min node count is provided it will use the scale settings for the cluster
compute_target.wait_for_completion(
    show_output=True, min_node_count=None, timeout_in_minutes=20)

# For a more detailed view of current AmlCompute status, use get_status()
print(compute_target.get_status().serialize())

```

Found compute target. just use it. cpucluster

24. Run the code under "Submit a run to use the cloud compute". This may take a while to run.



The screenshot shows a Microsoft Azure Notebooks Jupyter interface. The title bar indicates the workspace is 'LKH-MLWorkspace' and the notebook is 'predict-employee-retention.ipynb'. The menu bar includes 'File', 'Edit', 'View', 'Insert', 'Cell', 'Kernel', 'Widgets', and 'Help'. The toolbar below the menu has icons for file operations like Open, Save, and Run, along with tabs for 'Preview', 'My Projects', and 'Help'. A Python 3.6 kernel is selected. The main area displays a code cell titled 'Submit a run to use the cloud compute'. The code imports 'azureml.train.sklearn' and uses it to define a 'SKLearn' estimator with specific parameters. It then calls 'exp.submit' to start the run. A warning message is displayed in a red box at the bottom of the cell: 'WARNING - If environment\_definition or conda\_dependencies\_file\_path is specified, Azure ML will not install any framework related packages on behalf of the user.' and 'WARNING - This compute target type doesn't support non-Docker runs; overriding run configuration enable Docker.' Below the code cell is an 'Out[40]' section showing a table with one row of data. The table columns are 'Experiment', 'Id', 'Type', 'Status', 'Details Page', and 'Docs Page'. The data row shows 'predict-employee-retention' as the experiment name, 'retention\_158798379\_27a0c289' as the id, 'azureml.scriptrun' as the type, 'Starting' as the status, a 'Link to Azure Machine Learning studio' as the details page, and a 'Link to Documentation' as the docs page.

Experiment	Id	Type	Status	Details Page	Docs Page
predict-employee-retention	retention_158798379_27a0c289	azureml.scriptrun	Starting	<a href="#">Link to Azure Machine Learning studio</a>	<a href="#">Link to Documentation</a>

25. You can delete the "cpucluster" in the Azure ML Workspace (Compute > Compute clusters) when you run the under "Option 3: Mount data file to a remote compute target for training script to access" which also uses cloud compute.

**Well done! You have setup your Azure Machine Learning successfully and learnt some basic concepts!**