



DXC REALTIME PROJECTS

AZ-900, DP - 203



[DATE]
[COMPANY NAME]
[Company address]

Name:Bandi Praveen Kumar

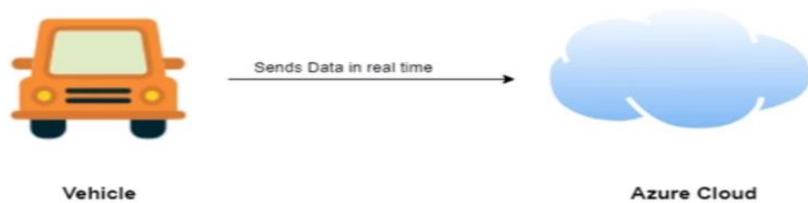
Reg No: DXC262ab1216

Project1 Name: Smart Vehicles

Date:10-06-2022

Project 1 : Connected Vehicles

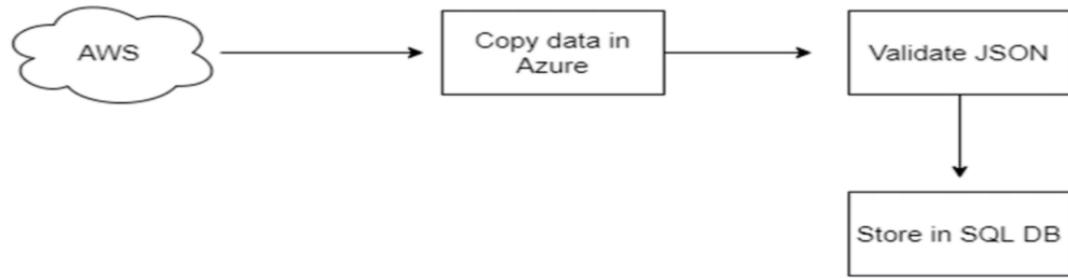
- General Motors is one of the leading heavy vehicle manufacture company. To improve their service they are planning to rollout lot new features based on IoT.



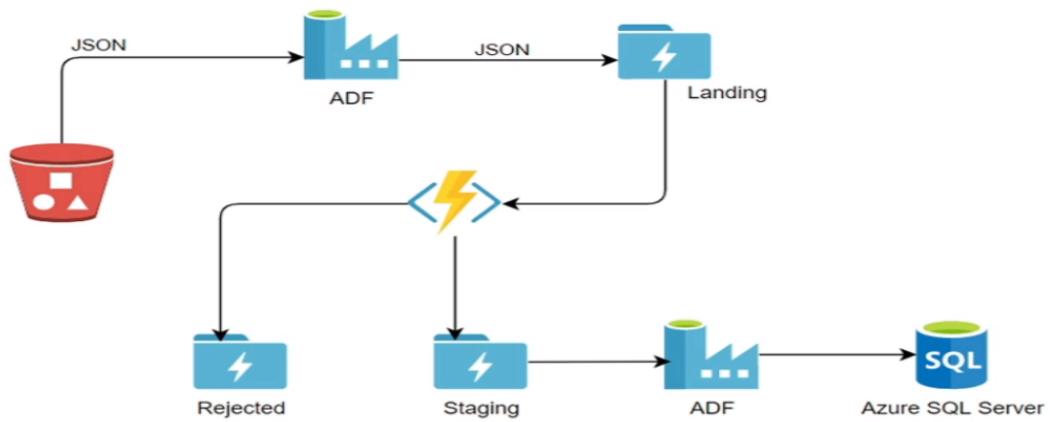
Project 1 : Connected Vehicles

- Vehicle has third party IoT device which will send the telemetry data (in JSON format) over the AWS cloud.
- You need to move data from third party AWS to General Motors Azure cloud.
- You need to validate the JSON sometime it could be incomplete or wrong JSON which need to be rejected.
- Once JSON got validated this data would be stored in the SQL database which will be further utilized by data science team.

Project 1 : Connected Vehicles



Project 1 : Connected Vehicles



Architecture Diagram for Connected Vehicle Project

Practical Lab: Create **Azure Data Factory** Account For Data pipelines

The screenshot shows the Microsoft Azure portal interface. At the top, there are several tabs: 'Subscription Details | Nuvepro', 'Home - Microsoft Azure', 'Session expired', 'Create Cluster - Databricks', 'Compute - Databricks', and a '+' button. The main content area is titled 'Azure services' and shows a grid of icons for various services: Create a resource, Monitor, Azure Databricks, Azure Synapse Analytics, Resource groups, SQL databases, Storage accounts, and Data factories. The 'Data factories' icon is highlighted with a tooltip. Below this is a section titled 'Resources' with tabs 'Recent' and 'Favorite'. Under 'Recent', there is a table listing four resources: 'morgodata' (Azure Databricks Service), 'apmorgan' (Resource group), 'dxcdata2' (Azure Databricks Service), and 'dxcrig2' (Resource group). The table includes columns for 'Name', 'Type', and 'Last Viewed'. A 'See all' link is also present. To the right of the table is a 'Navigate' section with links to 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. At the bottom of the main content area, there is a message: 'No data factories to display. Try changing or clearing your filters.' with a 'Create data factory' button. The status bar at the bottom shows the weather (34°C, Partly sunny), system icons, and the date (10-06-2022).

Practical Lab: Create ADF Pipeline End to end pipeline with triggers enabled

The screenshot shows the Microsoft Azure portal interface. The top navigation bar includes tabs for 'Subscription Details | Nuvepro', 'Home - Microsoft Azure', 'Resource groups - Microsoft Azure', and 'New Tab'. The main content area is titled 'Resource groups' and shows a list of resource groups. The first item is 'dxcdata2', which contains 'Azure Databricks Service' (8 hours ago). Below it is 'dxcrg2', which is a 'Resource group' (8 hours ago). A search bar at the top says 'subs'. A sidebar on the left lists 'Services' like Subscriptions, Event Hubs, and Web PubSub Service. A 'Navigate' section at the bottom has links for 'Subscriptions', 'Resource groups', 'All resources', and 'Dashboard'. A 'Tools' bar at the bottom includes icons for weather (35°C, Partly sunny), search, and other utilities.

This screenshot shows the 'Subscriptions' blade in the Microsoft Azure portal. It displays a single subscription named 'Azure-DXC262AB12Lab' with the ID '406c09c5-382f-44be-b843-b50a69eadc92'. The subscription status is 'Specified access'. It is part of the 'Tenant Root Group' and has an 'Active' status. The page includes a search bar and filter options for 'Subscriptions', 'My role', and 'Status'. Navigation links include 'Add', 'Manage Policies', and 'View Requests'. The bottom of the screen shows a standard Windows taskbar with icons for file operations, search, and system status.

Practical Lab: Create Azure blob trigger logic

The screenshot shows the Microsoft Azure Storage account interface for 'storage1011'. On the left, there's a navigation sidebar with links like Overview, Activity log, Tags, Diagnose and solve problems, Access Control (IAM), Data migration, Events, and Storage browser (preview). Under Data storage, the 'Containers' section is selected, showing a list of existing containers: 'Logs' (Last modified: 6/10/2022, 5:23:53 PM, Public access level: Private). A search bar at the top allows searching by prefix. On the right, a 'New container' dialog box is open, asking for a 'Name' (set to 'source') and 'Public access level' (set to 'Private (no anonymous access)'). Below the dialog are 'Advanced' settings and 'Create' and 'Discard' buttons.

Practical Lab: Create Azure SQL Server and Database

The screenshot shows the Microsoft Azure portal homepage. In the center, the 'SQL databases' blade is open, featuring a 'Create' button and a 'View' button. To the right, there's a 'Free training from Microsoft' section with three course cards: 'Provision an Azure SQL database to store appli...', 'Secure your Azure SQL Database', and 'Scale multiple Azure SQL Databases with SQL e...'. Below these cards is a 'Useful links' section. On the left, there's a 'Create a resource' button and sections for 'Resources' (Recent and Favorite) and 'Navigate' (Subscriptions, Resource groups, All resources, Dashboard). At the bottom, there's a toolbar with various icons and a status bar showing the URL, weather (35°C Partly sunny), and system information (ENG INTL, 17:22, 10-06-2022).

Subscription Details | Nuvelpro | SQL databases - Microsoft Azure | Resource groups - Microsoft Azure | +

portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.Sql%2Fservers%2fdatabases

Microsoft Azure

Search resources, services, and docs (G+)

Home > SQL databases

Manipal Pro Learn (manipalazure.onmicrosoft.com)

+ Create Reservations Manage view Refresh Export to CSV Open query Assign tags Delete

Filter for any field... Subscription == all Resource group == all Location == all Add filter

No grouping List view

Name ↑ Server ↓ Replica type ↓ Pricing tier ↓ Location ↓ Subscription ↓

No SQL databases to display

Try changing or clearing your filters.

Create SQL database

Learn more ↗

Subscription Details | Nuvelpro | Create SQL Database - Microsoft | Resource groups - Microsoft Azure | +

portal.azure.com/#create/Microsoft.SQLDatabase

Microsoft Azure

Search resources, services, and docs (G+)

Home > SQL databases > Create SQL Database

Microsoft

Basics Networking Security Additional settings Tags Review + create

Create a SQL database with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize. [Learn more ↗](#)

Did you know that new users in Azure can create a free Azure SQL Database and use it for 12 months using Azure free account? [Learn more ↗](#)

Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Azure-DXC262AB12Lab

Resource group * (New) dxcstorage Create new

Database details

Enter required settings for this database, including picking a logical server and configuring the compute and storage resources

Database name * Enter database name

Review + create Next : Networking >

https://portal.azure.com/#

35°C Partly sunny

ENG INTL 17:26 10-06-2022

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#view/SqIAzureExtension/SharedCreateServer/isNewServer~/true/subscriptionId/406c09c5-382f-44be-b843-b50a69eadc92/resourceGroupName/dxcstorage/infoFilters...>. The page title is "Create SQL Database Server".

Server details

Enter required settings for this server, including providing a name and location. This server will be created in the same subscription and resource group as your database.

Server name *: Enter server name: . Location:

Authentication

Select your preferred authentication methods for accessing this server. Create a server admin login and password to access your server with SQL authentication, select only Azure AD authentication [Learn more](#), or select both SQL and Azure AD authentication.

Authentication method: Use SQL authentication Use only Azure Active Directory (Azure AD) authentication Use both SQL and Azure AD authentication

Server admin login *: Enter server admin login:

Password *:

Confirm password *:

OK

The taskbar at the bottom shows the date as 10-06-2022 and the time as 17:28.

Practical Lab: Add another pipelines for moving data from Staging to SQL DB

The screenshot shows the Microsoft Azure portal with the URL <https://portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.DataFactory%2FdataFactories>. The page title is "Data factories".

Data factories

Manage Pro Learn (manipalazure.onmicrosoft.com)

+ Create Manage view Refresh Export to CSV Open query Assign tags

Filter for any field... Subscription == all Type == all Resource group == all Location == all Add filter

No grouping List view

Name ↑ Type ↑↓ Subscription ↑↓ Resource group ↑↓ Location ↑↓

No data factories to display
Try changing or clearing your filters.
[Create data factory](#) [Learn more](#)

The taskbar at the bottom shows the date as 10-06-2022 and the time as 17:56.

The screenshot shows the Microsoft Azure portal interface. On the left, there's a sidebar with various navigation options like Overview, Activity log, Tags, Diagnose and solve problems, Getting started, and Query editor (preview). The main area is titled "newsqldatabase (newserver101/newsqldatabase) | Query editor (preview)". It displays a table named "transaction.Azure Blob Storage file". The table has columns: Date, Account, Transaction, Amount, Merchant, PaymentChannel, and Type. The data in the table is as follows:

Date	Account	Transaction	Amount	Merchant	PaymentChannel	Type
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	DEPOSITED QR CAS...	-\$150.00			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$10,935.00			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	BANKCARD MTOT D...	\$109.64			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$616.63			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	\$702.15			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$590.92			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	\$905.78			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	T-Mobile	-\$442.87	T-Mobile		place
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	BUSINESS TO BUSIN...	-\$3,847.67		ACH	place
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$636.90			special
3/21/2022 12:00:00 ...	BUSINESS CHECKIN...	CHECK	-\$1,208.50			special
3/18/2022 12:00:00 ...	BUSINESS CHECKIN...	AMERICAN EXPRESS...	\$4,422.12			special

Result: Hence the creation of **ADF Pipeline End to end**

pipeline with triggers enabled and created, **ADF Pipeline End to end pipeline with triggers enabled,Azure blob trigger logic**
Azure SQL Server and Database

Conclusion: Finally ADE pipeline blob trigger ,SQL server and database

References: [Azure Blob storage trigger for Azure Functions - Microsoft Docs](#)

<https://docs.microsoft.com> > Azure > Functions

Name:Bandi Praveen Kumar

Reg No: DXC262ab1216

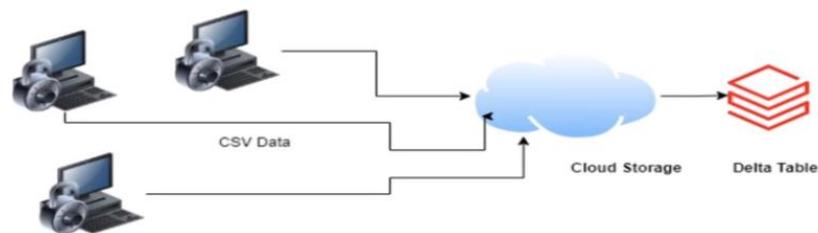
Project2 Name: AP Morgan Data Platform

Date:10-06-2022

Project 2: AP Morgan Data Platform

Project 2 : AP Morgan

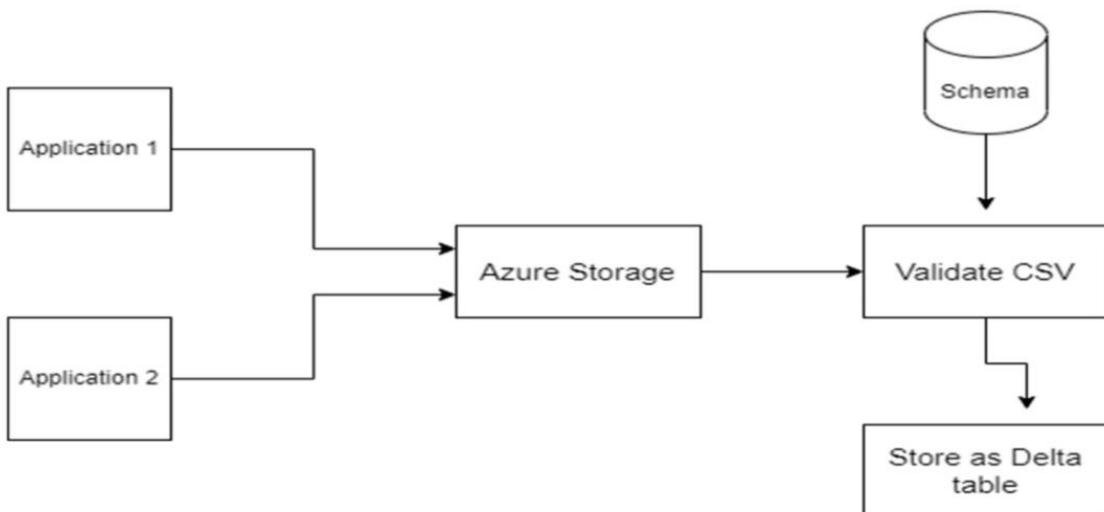
- Multiple Internal applications sends the data(huge size) in CSV format on daily basis in the cloud storage location. There are couple of Data/schema validation needed to be performed on this incoming data. Once everything is passed data to be persisted as Delta table in Databricks for downstream system.



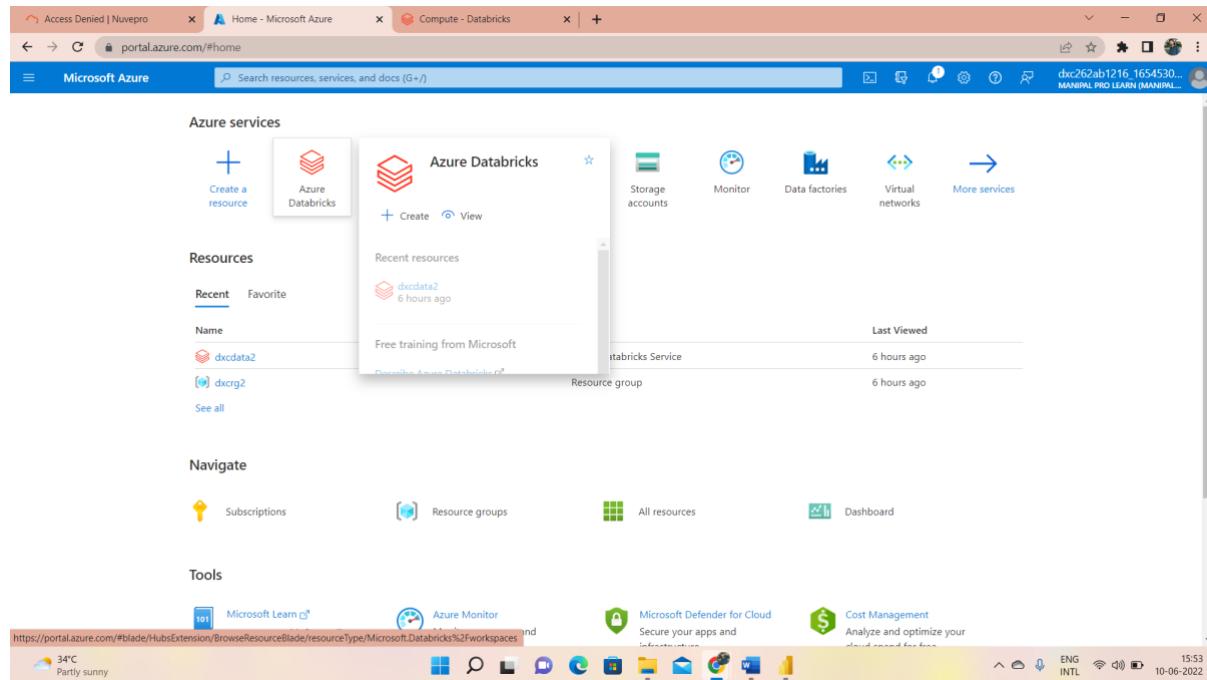
Project 2 : AP Morgan- High Level Detail

- Internal Application sends CSV file in Azure data lake storage.
- Validation needed to apply on this follows:
 - Check for duplicate rows. If it contains duplicate rows, file need to be rejected.
 - Need to validate the date format for all the date fields. Date column names and desired date format is stored in a Azure SQL server. If validation fails file will be rejected.
- Move all the rejected files to Reject folder.
- Move all the passed files to Staging folder.
- Write the passed files as the Delta table in the Azure Databricks

Project 2 : AP Morgan



Practical Lab: Create a Databricks



Access Denied | Nuvepro

Create an Azure Databricks workspace

Compute - Databricks

Microsoft Azure

Search resources, services, and docs (G+)

dxc262ab1216.1654530...
MANUAL PRO LEARN (MANUAL...)

Home > Azure Databricks >

Create an Azure Databricks workspace

Basics Networking Advanced Tags Review + create

Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription *

Resource group *

Instance Details

Workspace name *

Region *

Pricing Tier *

Review + create < Previous Next : Networking >

34°C Partly sunny

ENG INTL 15:53 10-06-2022

Access Denied | Nuvepro

Create an Azure Databricks workspace

Compute - Databricks

Microsoft Azure

Search resources, services, and docs (G+)

dxc262ab1216.1654530...
MANUAL PRO LEARN (MANUAL...)

Home > Azure Databricks >

Create an Azure Databricks workspace

Validation Succeeded

Basics Networking Advanced Tags Review + create

Summary

Workspace name: morgondata
Subscription: Azure-DXC262AB12Lab
Resource group: apmorgon
Region: UK South
Pricing Tier: standard

Networking

Deploy Azure Databricks workspace with Secure Cluster Connectivity (No Public IP): No
Deploy Azure Databricks workspace in your own Virtual Network (VNet): No

Advanced

Enable Infrastructure Encryption: No

Create < Previous Download a template for automation

34°C Partly sunny

ENG INTL 15:59 10-06-2022

The screenshot displays two Microsoft Azure portal windows side-by-side.

Top Window (Deployment Overview):

- Deployment Name:** apmargon_morgondata
- Subscription:** Azure-DXC262AB12Lab
- Resource group:** apmargon
- Start time:** 6/10/2022, 4:00:09 PM
- Correlation ID:** 009f5bcf-c7e5-4597-8451-8826541f3a70

Bottom Window (Databricks Service Overview):

- Name:** morgondata - Microsoft Azure
- Status:** Active
- Resource group:** apmargon
- Location:** UK South
- Subscription:** Azure-DXC262AB12Lab
- Subscription ID:** 406c09c5-382f-44be-b843-b50a69eadc92
- Tags:** Click here to add tags
- Managed Resource Group:** databricks-rg-morgondata-vhvoal3jz5wg4
- URL:** <https://adb-59242257778157055.azure.databricks.net>
- Pricing Tier:** standard

Common UI Elements:

- Search bar: Search (Ctrl+Shift+F)
- Header: Microsoft Azure, portal.azure.com, and user profile information.
- Footer: Weather (34°C, Partly sunny), system tray (ENG INTL, 16:02, 10-06-2022).

Practical Lab: Create Cluster in Azure Databricks

The screenshot shows the Microsoft Azure Databricks landing page. On the left, there's a sidebar with sections like 'Get started', 'Set up your workspace', and 'Next steps'. The main area is titled 'Data Science & Engineering' and contains cards for 'Notebook', 'Data import', and 'Guide: Quickstart tutorial'. Below these are sections for 'Recents', 'Documentation', 'Release notes', and 'Blog posts'. The documentation section includes links to 'Get started guide', 'Best practices', 'Data guide', and 'More documentation'. The release notes section lists 'Runtime release notes' and 'Platform release notes'. The blog posts section features articles like 'Building ETL pipelines for the cybersecurity lakehouse with Delta Live Tables' and 'Streaming Windows Event Logs into the Cybersecurity Lakehouse'. At the bottom, there's a weather widget showing '34°C Partly sunny'.

The screenshot shows the 'Create a cluster' page under the 'Compute' tab. It has tabs for 'All-purpose clusters', 'Job clusters', and 'Pools'. A search bar at the top allows filtering by 'Name', 'Runtime', 'Active memory', 'Active cores', 'Active DBU / h', 'Source', and 'Creator'. Below the search bar, a note says 'Depending on your workloads we recommend different cluster configurations.' and a link to 'this guide for best practices'. The page also includes a 'Create Cluster' button and a 'Don't show again' link at the bottom.

Access Denied | Nuvepro

morgondata - Microsoft Azure

Create Cluster - Databricks

Compute - Databricks

adb-5924225777815705.azuredatabricks.net/?o=5924225777815705#create/cluster

Microsoft Azure | Databricks

Create a cluster

You'll use compute resources (clusters) to run your commands.

Click 'Create cluster' and use our [best practices guide](#) to set up your cluster.

New Cluster

Cancel Create Cluster

DBU / hour: 2.25 - 6.75

2-8 Workers: 28-112 GB Memory, 8-32 Cores
1 Driver: 14 GB Memory, 4 Cores

apmorgon1

Cluster mode: Standard

Databricks runtime version: Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)

Promotional discount applied to Photon during preview

Use Photon Acceleration

Autopilot options: Enable autoscaling, Terminate after 30 minutes of inactivity

Worker type: Standard_DS3_v2 (14 GB Memory, 4 Cores), Min workers: 2, Max workers: 8, Spot instances

Driver type: Same as worker (14 GB Memory, 4 Cores)

DBU / hour: 2.25 - 6.75

Standard_DS3_v2

Advanced options

Don't show again

34°C Party sunny

16:13 ENG INTL 10-06-2022

This screenshot shows the 'Create a cluster' wizard in the Microsoft Azure Databricks portal. The 'New Cluster' tab is selected. The cluster name is 'apmorgon1'. The 'Cluster mode' is set to 'Standard'. The 'Databricks runtime version' is 'Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)'. Under 'Autopilot options', 'Enable autoscaling' and 'Terminate after 30 minutes of inactivity' are checked. The 'Worker type' is 'Standard_DS3_v2' with '14 GB Memory, 4 Cores', and the 'Min workers' is '2' and 'Max workers' is '8'. The 'Driver type' is 'Same as worker' with '14 GB Memory, 4 Cores'. The DBU / hour range is '2.25 - 6.75' and the driver type is 'Standard_DS3_v2'. There is also an 'Advanced options' section.

Access Denied | Nuvepro

Session expired

Create Cluster - Databricks

Compute - Databricks

adb-5924225777815705.azuredatabricks.net/?o=5924225777815705#setting/clusters/0610-104429-u0fwd41/configuration

Microsoft Azure | Databricks

Create a cluster

Completed

You'll use compute resources (clusters) to run your commands.

Click 'Create cluster' and use our [best practices guide](#) to set up your cluster.

Next step

Ingest data

Configuration Notebooks (0) Libraries Event log Spark UI Driver logs Metrics Apps Spark cluster UI - Master

More Edit Terminate

Cluster mode: Standard

Databricks Runtime Version: 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)

Use Photon Acceleration

Autopilot options: Enable autoscaling, Terminate after 30 minutes of inactivity

Worker type: Standard_DS3_v2 (14 GB Memory, 4 Cores), Min workers: 2, Max workers: 8, Current: 2, Spot instances

Driver type: Standard_DS3_v2 (14 GB Memory, 4 Cores)

DBU / hour: 2.25 - 6.75

Standard_DS3_v2

Advanced options

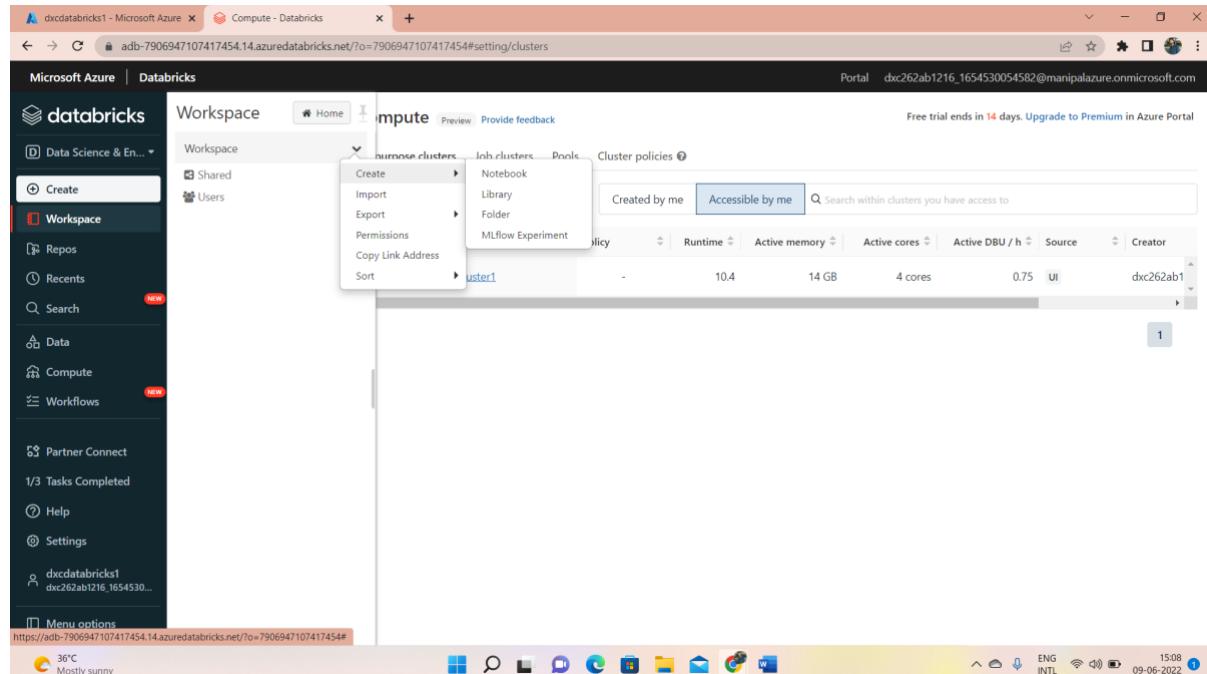
Don't show again

34°C Party sunny

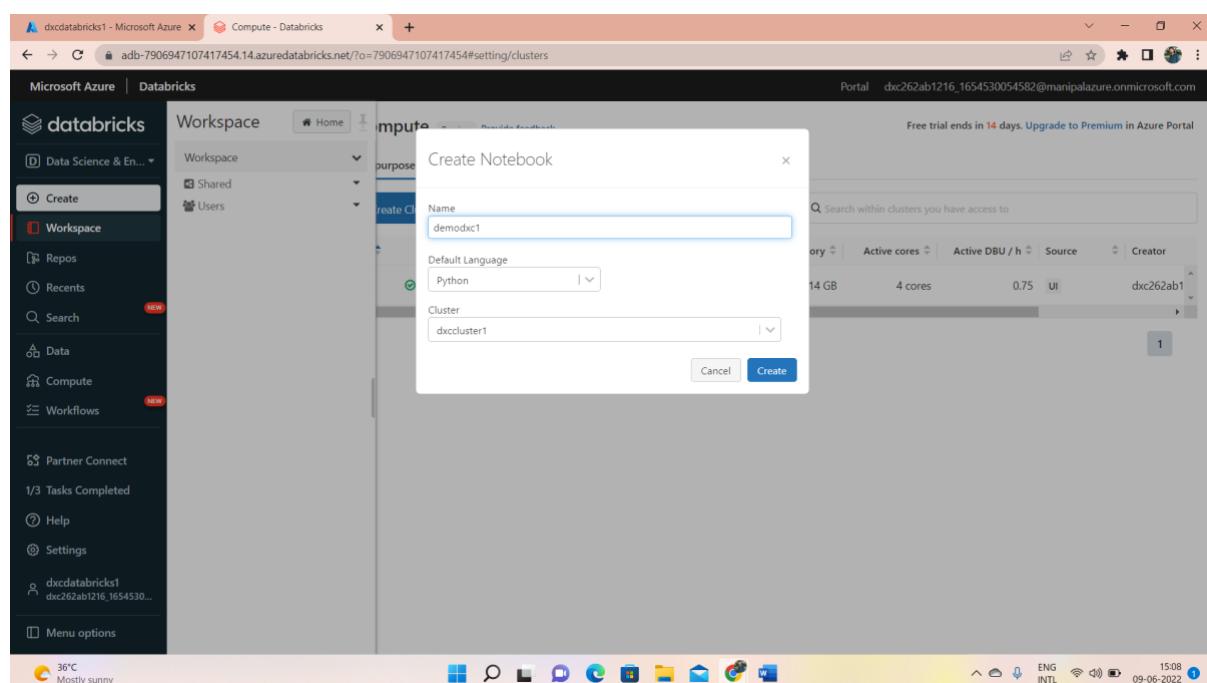
16:20 ENG INTL 10-06-2022

This screenshot shows the 'Create a cluster' wizard in the Microsoft Azure Databricks portal. The 'Completed' step is displayed. The cluster configuration is identical to the previous screenshot: 'Cluster mode: Standard', 'Databricks Runtime Version: 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)', 'Autopilot options: Enable autoscaling, Terminate after 30 minutes of inactivity', 'Worker type: Standard_DS3_v2 (14 GB Memory, 4 Cores)', 'Min workers: 2', 'Max workers: 8', 'Current: 2', and 'Driver type: Standard_DS3_v2 (14 GB Memory, 4 Cores)'. The DBU / hour range is '2.25 - 6.75'. The 'Advanced options' section is also present. Below the configuration, there is a 'Next step' section with a green button labeled 'Ingest data'.

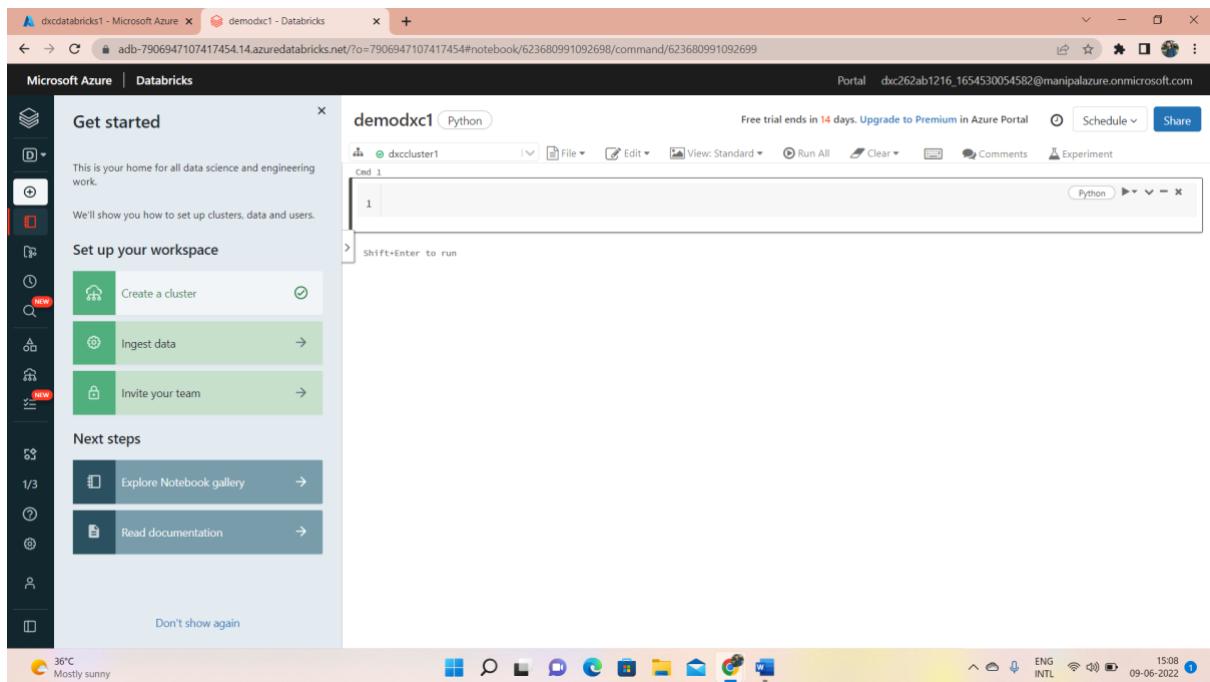
Practical Lab: Add notebook in Databricks and Implement the Business Logic



The screenshot shows the Microsoft Azure Databricks workspace interface. The 'Compute' tab is active. A context menu is open over a cluster named 'dxccluster1'. The 'Create' option is highlighted, and a submenu shows 'Notebook' as the selected item.



The screenshot shows the Microsoft Azure Databricks workspace interface. A 'Create Notebook' dialog box is open. The 'Name' field is set to 'demodxc1', 'Default Language' is set to 'Python', and the 'Cluster' dropdown is set to 'dxccluster1'. The 'Create' button is visible at the bottom right.



Practical Lab: Azure Data Factory For AP Morgan

Screenshot of the Microsoft Azure portal showing the Data factories service page.

The top navigation bar includes tabs for Subscription Details, Home - Microsoft Azure, Session expired, Create Cluster - Databricks, Compute - Databricks, and a search bar.

The main content area features the "Azure services" section with icons for Create a resource, Monitor, Azure Databricks, Azure Synapse Analytics, Resource groups, SQL databases, Storage accounts, and Data factories. The Data factories icon is highlighted with a tooltip showing "+ Create" and "View".

The "Resources" section displays a list of recent resources:

Name	Type	Last Viewed
morgodata	Azure Databricks Service	24 minutes ago
apmorgon	Resource group	25 minutes ago
dxcdata2	Azure Databricks Service	6 hours ago
dxcrq2	Resource group	7 hours ago

Below this is a "See all" link and a "Navigate" section with links for Subscriptions, Resource groups, All resources, and Dashboard.

The browser address bar shows the URL: https://portal.azure.com/#blade/HubsExtension/BrowseResourceBlade/resourceType/Microsoft.DataFactory%2FdataFactories. The system tray at the bottom indicates a temperature of 34°C and a weather condition of Party sunny. The date is 10-06-2022.

Screenshot of the Microsoft Azure portal showing the Data factories list page.

The top navigation bar includes tabs for Subscription Details, Data factories - Microsoft Azure, Session expired, Create Cluster - Databricks, Compute - Databricks, and a search bar.

The main content area shows the "Data factories" list with the following filters applied:

- + Create
- Manage view
- Refresh
- Export to CSV
- Open query
- Assign tags

Filter options include:

- Subscription == all
- Type == all
- Resource group == all
- Location == all
- Add filter

Sorting options include:

- Name ↑↓
- Type ↑↓
- Subscription ↑↓
- Resource group ↑↓
- Location ↑↓

The central message states: "No data factories to display. Try changing or clearing your filters." It includes a "Create data factory" button and a "Learn more" link.

The browser address bar shows the URL: https://portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.DataFactory%2FdataFactories. The system tray at the bottom indicates a temperature of 34°C and a weather condition of Party sunny. The date is 10-06-2022.

Subscription Details | Nuvelpro | Create Data Factory - Microsoft | Session expired | Create Cluster - Databricks | Compute - Databricks | + | - | X | portal.azure.com/#create/Microsoft.DataFactory | Microsoft Azure | Search resources, services, and docs (G+) | Home > Data factories > Create Data Factory | dxc262ab1216.1654530... | MANIFAL PRO LEARN (MANIFAL...)

Basics Git configuration Networking Advanced Tags Review + create

Project details

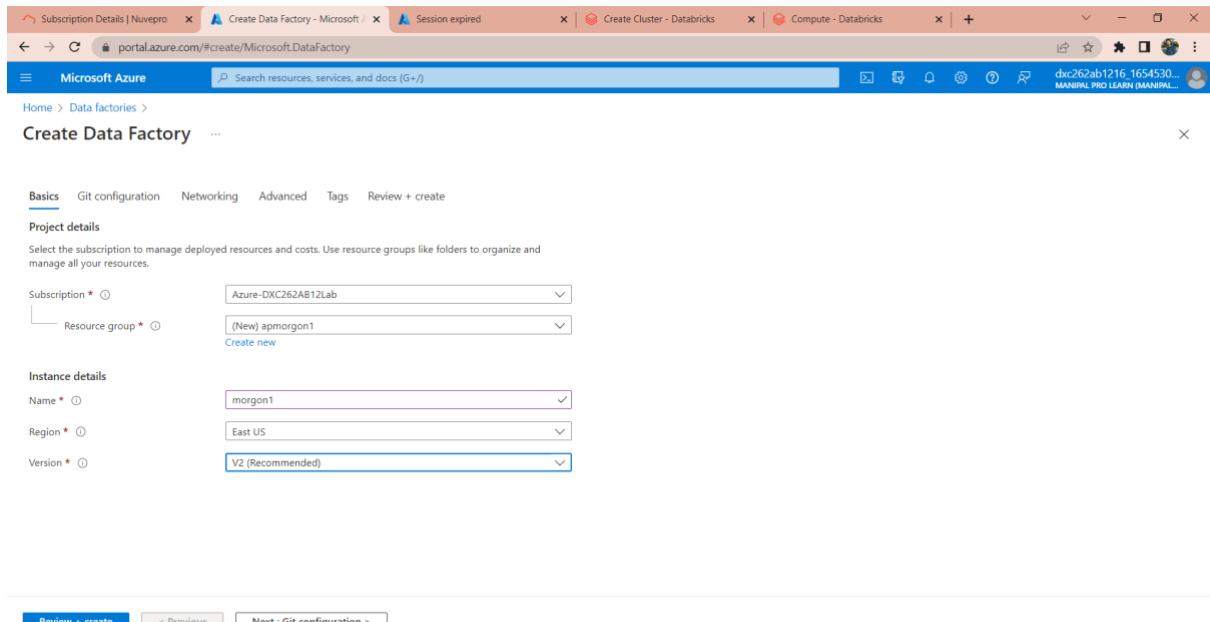
Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * Resource group * [Create new](#)

Instance details

Name * Region * Version *

[Review + create](#) < Previous Next : Git configuration >



Subscription Details | Nuvelpro | Create Data Factory - Microsoft | Session expired | Create Cluster - Databricks | Compute - Databricks | + | - | X | portal.azure.com/#create/Microsoft.DataFactory | Microsoft Azure | Search resources, services, and docs (G+) | Home > Data factories > Create Data Factory | dxc262ab1216.1654530... | MANIFAL PRO LEARN (MANIFAL...)

Basics Git configuration Networking Advanced Tags [Review + create](#)

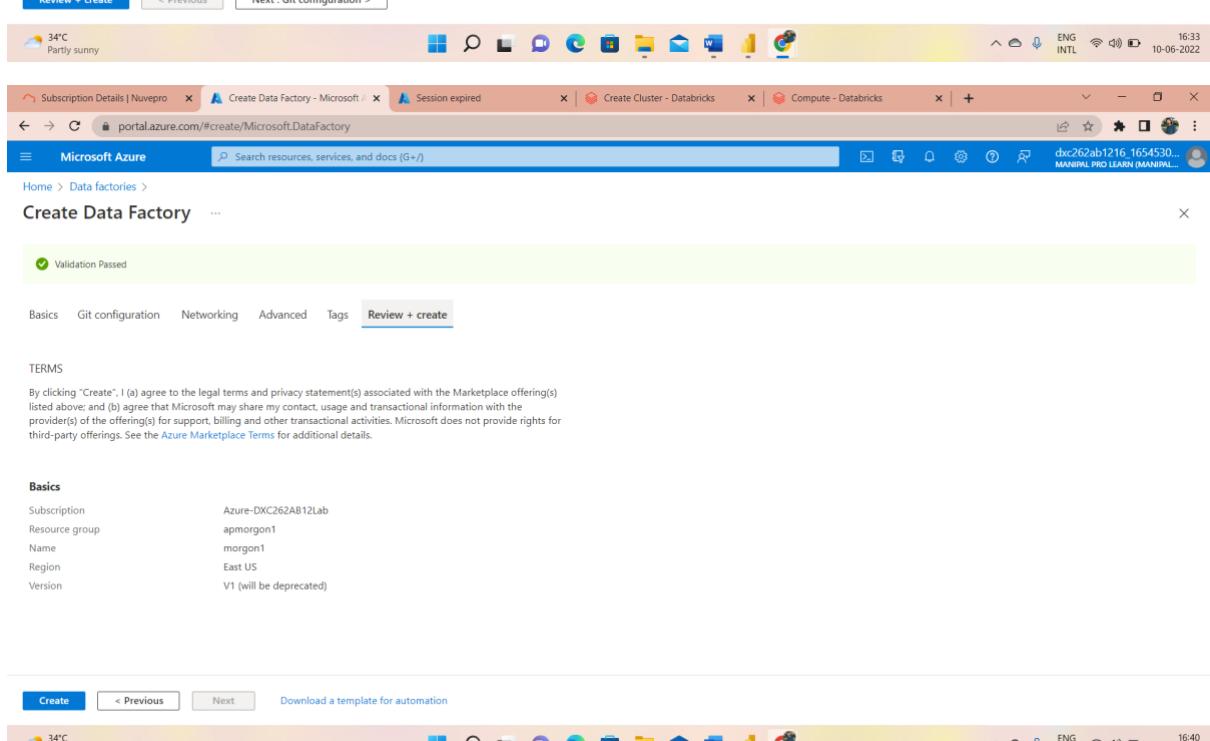
TERMS

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; and (b) agree that Microsoft may share my contact, usage and transactional information with the provider(s) of the offering(s) for support, billing and other transactional activities. Microsoft does not provide rights for third-party offerings. See the [Azure Marketplace Terms](#) for additional details.

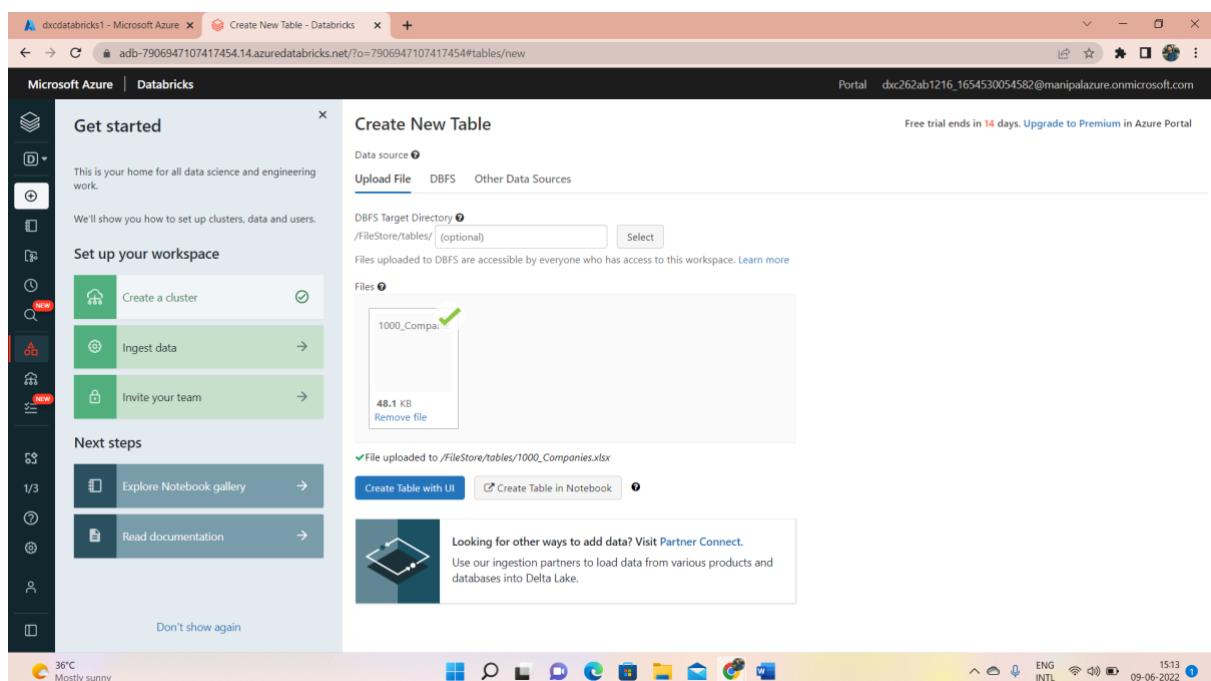
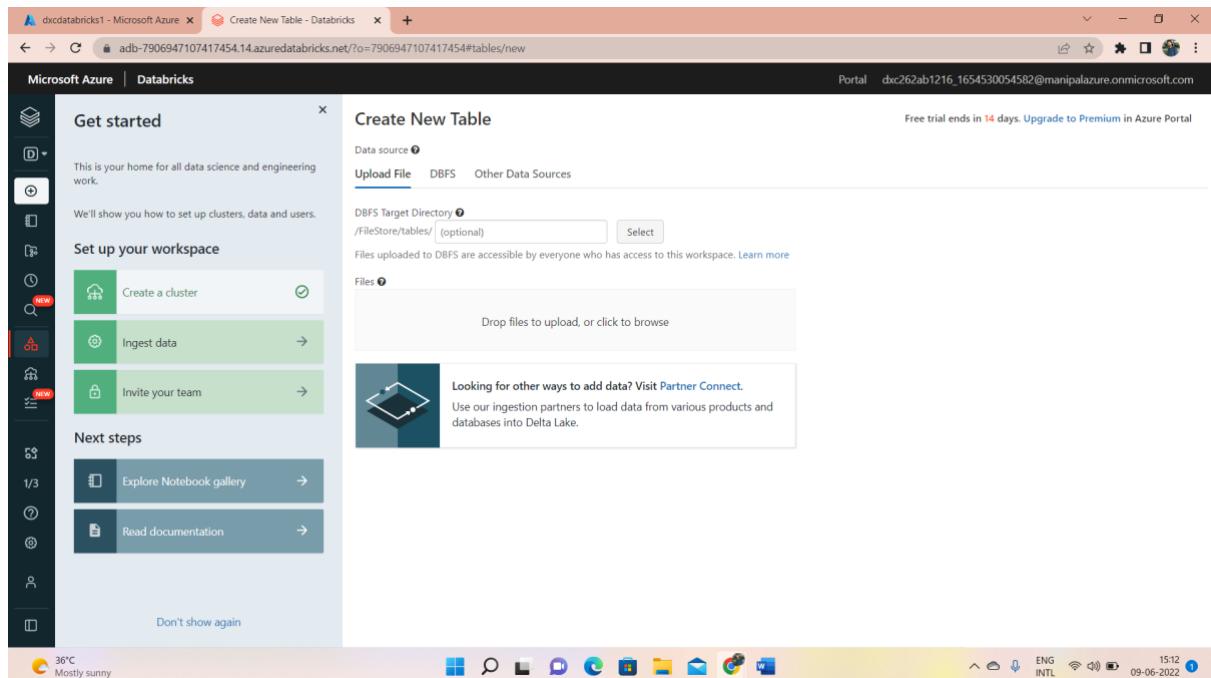
Basics

Subscription	Azure-DXC262AB12Lab
Resource group	apmorgon1
Name	morgon1
Region	East US
Version	V1 (will be deprecated)

[Create](#) < Previous Next Download a template for automation



Practical Lab: Create Azure Databricks Linked Service in ADF



[adx databricks1 - Microsoft Azure](#) [2022-06-09 - DBFS Example - Databricks](#) +

adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#notebook/623680991092700/command/623680991092701

Microsoft Azure | Databricks

Get started

This is your home for all data science and engineering work.

We'll show you how to set up clusters, data and users.

Set up your workspace

- Create a cluster
- Ingest data
- Invite your team

Next steps

- Explore Notebook gallery
- Read documentation

Don't show again

36°C Mostly sunny

2022-06-09 - DBFS Example Python

Detached Cell 1

Overview

This notebook will show you how to create and query a table or DataFrame that you uploaded to DBFS. DBFS is a Databricks File System that allows you to store data for querying inside of Databricks. This notebook assumes that you have a file already inside of DBFS that you would like to read from.

This notebook is written in **Python** so the default cell type is Python. However, you can use different languages by using the `%LANGUAGE` syntax. Python, Scala, SQL, and R are all supported.

```
Cmd 2
```

```
1 # File location and type
2 file_location = "/FileStore/tables/1000_Companies.xlsx"
3 file_type = "xlsx"
4
5 # CSV options
6 infer_schema = "false"
7 first_row_is_header = "false"
8 delimiter = ","
9
10 # The applied options are for CSV files. For other file types, these will be ignored.
11 df = spark.read.format(file_type) \
12     .option("inferSchema", infer_schema) \
13     .option("header", first_row_is_header) \
14     .option("sep", delimiter) \
15     .load(file_location)
16
17 display(df)
```

Python

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Portal dx262ab1216_1654530054582@manipalazure.onmicrosoft.com

Schedule Share

Markdown

ENG INTL 09-06-2022 15:15

[adx databricks1 - Microsoft Azure](#) [Create New Table - Databricks](#) +

adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#tables/new

Microsoft Azure | Databricks

Create New Table

Preview Table

Specify Table Attributes

Specify the Table Name, Database and Schema to add this to the data UI for other users to access

⚠ Invalid characters (.,();=& found in the columns) have been replaced with underscores ⓘ

Table Name companies_csv

Create in Database default

File Type CSV

Column Delimiter ,

First row is header

Infer schema

Multi-line

Create Table

Create Table in Notebook

36°C Mostly sunny

	R_D_Spend	Administration	Marketing_Spend	State	Profit
165349.2	136897.8	471784.1	New York	192261.83	
162597.7	151377.59	443898.53	California	191792.06	
153441.51	101145.55	407934.54	Florida	191050.39	
144372.41	118671.85	383199.62	New York	182901.99	
142107.34	91391.77	366168.42	Florida	166187.94	
131876.9	99814.71	362861.36	New York	156991.12	
134615.46	147198.87	127716.82	California	156122.51	

Free trial ends in 14 days. Upgrade to Premium in Azure Portal

Portal dx262ab1216_1654530054582@manipalazure.onmicrosoft.com

ENG INTL 09-06-2022 15:31

The screenshot shows the Microsoft Azure Databricks interface. At the top, there are two tabs: "dxdatabricks1 - Microsoft Azure" and "Create New Table - Databricks". The URL is "adb-7906947107417454.14.azuredatabricks.net/?o=7906947107417454#table/hive_metastore/default/companies_csv". The top right corner shows "Portal" and an email address "dx262ab1216_1654530054582@manipalazure.onmicrosoft.com". A message "Free trial ends in 14 days. Upgrade to Premium in Azure Portal" is displayed.

The main area displays the schema and sample data for a table named "companies_csv".

Schema:

	col_name	data_type	comment
1	R_D_Spend	float	
2	Administration	string	
3	Marketing_Spend	string	
4	State	string	
5	Profit	string	
6			
7	# Partitioning		

Showing all 8 rows.

Sample Data:

	R_D_Spend	Administration	Marketing_Spend	State	Profit
1	165349.2	136897.8	471784.1	New York	192261.83
2	162597.7	151377.59	443898.53	California	191792.06
3	153441.52	101145.55	407934.54	Florida	191050.39
4	144372.4	118671.85	383199.62	New York	182901.99
5	142107.34	91391.77	366168.42	Florida	166187.94

The bottom status bar shows the weather as "36°C Mostly sunny", system icons, and the date/time "09-06-2022 15:31".

Result: Hence, The creation of Data bricks, Cluster an notebook have created and updated with screenshots of it has been done

Conclusion: At last we found the result of AP Morgan data set

References: <https://www.databricks.com/>