

ASSIGNMENT-----SOLUTION SUBMISSION
ON
AZURE ANALYTICS
BY

NAME : SAI KIRAN ANCHE

BATCH:DXC-262-ANALYTICS-B12-AZURE

TRAINING UNDER : MANIPAL PRO LEARN

DATE OF SUBMISSION : 17-06-2022

EMPLOYEE DOMAIN - AZURE ANALYTICS

ROLL NO: DXC262AB12021

COMPANY – DXC TECHNOLOGY

TRAINER NAME – MR. AJAY KUMAR

NO OF QUESTIONS :08

Assignment	-	17th	June	2022:
<hr/>				
1.	Write a python program to predict car sales of a company by using below data,			
year	:	2010	2011	2012
Sales	in millions:	169	199	262
2013				
2014		301	345	398
2015			501	595
2016			610	700
2017			720	
display outcome using linear regression method				
2.	Write python program to generate possible tuples from any two sample Lists			
3.	Create Azure Databricks & try to connect databricks & powerBI ,			
explain the steps with screenshots.				

4.Create Azure Synapse & connect with Azure Blob, explain the steps with screenshots

5.Create Azure Synapse spark pool & query sample sample JSON file,
explain the steps with screenshots

6.Create Azure Cosmos DB & import sample JSON file, explain the steps with screenshots

7.Connect COSMOS DB & Azure Synapse analytics & explain the steps with screenshots

8.Create azure Data factory & azure Blob, connect Blob & ADF,

import blob files into Data factory & explain the steps with screenshots

Please create a word / pdf document, and send it to : avyuktitraining1@gmail.com

INTRODUCTION

This Assignment is given by manipal pro learn team on the basis of the training done in the forenoon session of this morning. The main objective behind this assignment is to master the theory and enhance knowledge over creating the data bricks and performing the analytics part.

There are 6 questions and they are of same level. All the questions have been focused on what the trainer taught in the earlier session. All the demonstrations have been done successfully and documented .This assignment gave me immense confidence in mastering the domain that has been assigned to me. Special thanks to Unext team for providing the lab access.

1. Write a python program to predict car sales of a company by using below data,

year : 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020

Sales in millions: 169 199 262 301 345 398 501 595 610 700 720

display outcome using linear regression method

The screenshot shows a Google Colab interface with a Jupyter notebook titled "Untitled2.ipynb". The notebook contains Python code for linear regression. The code imports matplotlib.pyplot and numpy, defines years and sales data, prints the sales data, imports linear_model from sklearn, defines variables Year and Sales_in_millions, prints the sales data again, reshapes the sales array, and prints the new sales array. The output cell shows the original sales data and the reshaped array [[69], [199], [262], [301]]. The status bar at the bottom indicates the code completed at 4:18 PM.

```
[2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020][169, 199, 262, 301, 345, 398, 501, 595, 610, 700, 720]

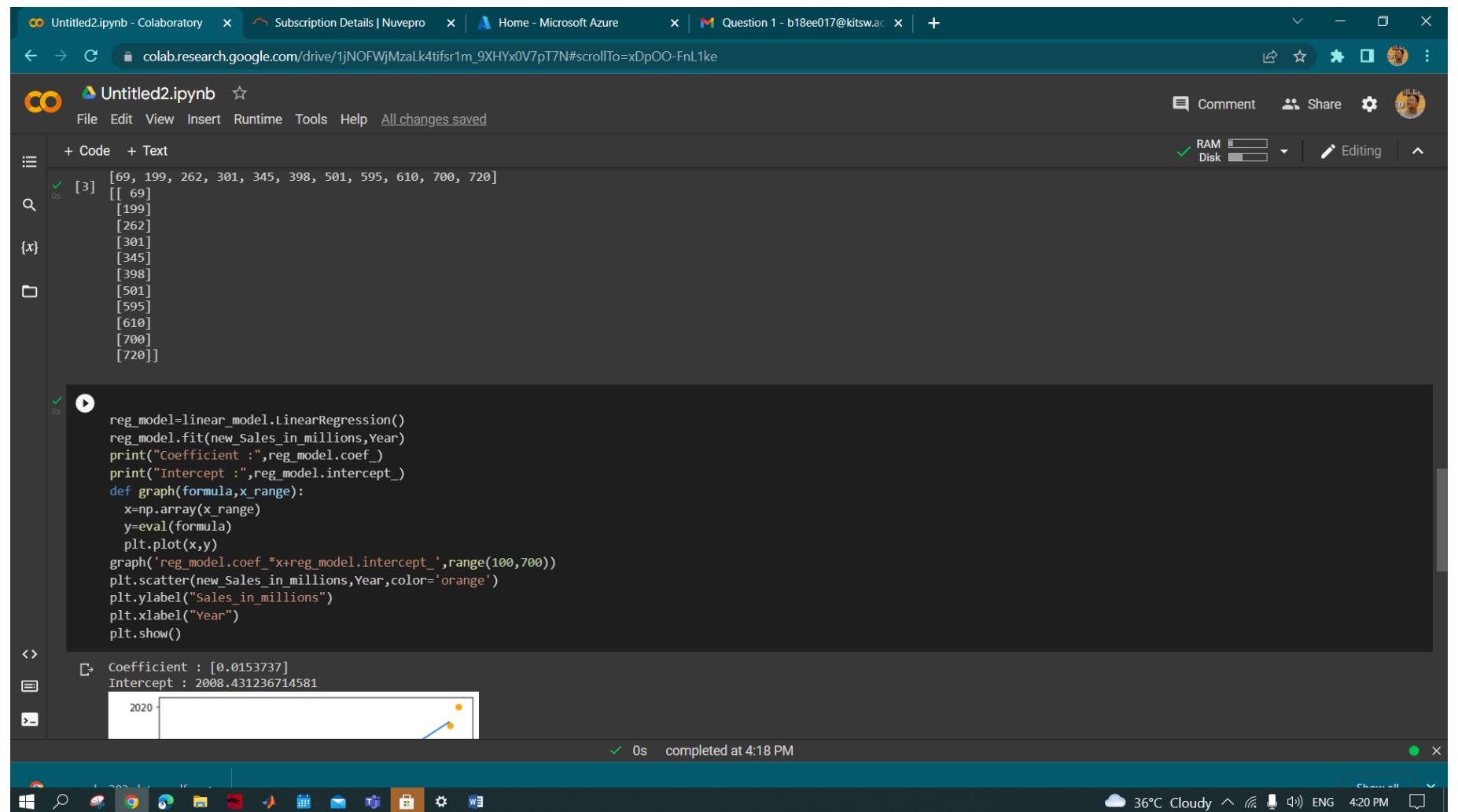
[3] import matplotlib.pyplot as plt
import numpy as np
from sklearn import linear_model

Year=[2010, 2011, 2012, 2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020]
Sales_in_millions=[169 ,199 ,262 ,301 ,345 ,398 ,501 ,595 ,610 ,700 ,720]

print(sales_in_millions)

new_Sales_in_millions=np.array(Sales_in_millions).reshape(-1,1)
print(new_Sales_in_millions)

[[69]
 [199]
 [262]
 [301]]
```



Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | Home - Microsoft Azure | Question 1 - b18ee017@kitsw.ac | +

colab.research.google.com/drive/1jNOFWjMzaLk4tfsr1m_9XHYx0V7pT7N#scrollTo=xDpOO-FnL1ke

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

RAM Disk Editing

```
reg_model=linear_model.LinearRegression()
reg_model.fit(new_Sales_in_millions,Year)
print("Coefficient : ",reg_model.coef_)
print("Intercept : ",reg_model.intercept_)
def graph(formula,x_range):
    x=np.array(x_range)
    y=eval(formula)
    plt.plot(x,y)
graph('reg_model.coef_*x+reg_model.intercept_',range(100,700))
plt.scatter(new_Sales_in_millions,Year,color='orange')
plt.ylabel("Sales in millions")
plt.xlabel("Year")
plt.show()
```

Coefficient : [0.0153737]
Intercept : 2008.431236714581

Sales in millions

Year

0s completed at 4:18 PM

2. Write python program to generate possible tuples from any two sample Lists

The screenshot shows a Google Colab notebook interface. The title bar indicates the notebook is titled "Untitled2.ipynb - Colaboratory". The code cell contains the following Python script:

```
[8] mydata1 = (4, 5)
mydata2 = (7, 8)

print("The original tuple 1: " + str(mydata1))

print("The original tuple 2: " + str(mydata2))

res = [(a, b) for a in mydata1 for b in mydata2]
res = res + [(a, b) for a in mydata2 for b in mydata1]
print("The possible combinations are : " + str(res))
```

The output of the code is displayed below the code cell:

```
The original tuple 1: (4, 5)
The original tuple 2: (7, 8)
The possible combinations are : [(4, 7), (4, 8), (5, 7), (5, 8), (7, 4), (7, 5), (8, 4), (8, 5)]
```

The status bar at the bottom of the screen shows "0s completed at 4:18 PM". The taskbar at the bottom of the desktop shows various application icons, and the system tray indicates a temperature of 36°C, cloudy weather, and the date/time as 4:21 PM.

3. Create Azure Databricks & try to connect databricks & powerBI , explain the steps with screenshots.

The screenshot shows the Microsoft Azure portal's search interface. The search bar at the top contains the query "azure data bricks". Below the search bar, there are several filter tabs: All, Services (99+), Documentation (29), Azure Active Directory (30), Resources (0), Resource Groups (0), and Marketplace (0). The main search results are displayed under the "Services" category. The first result is "Azure Database for MySQL servers", followed by "Azure Databricks", which is highlighted with a yellow box. Other services listed include "SQL databases", "Azure Cosmos DB", "Azure Data Box", "Azure Data Box Gateway", and "Azure Database Migration Projects". To the right of the search results, there is a sidebar with sections for "Reservations", "Documentation", and "Azure Active Directory". The "Documentation" section includes links to "What is Azure Databricks?", "Deploy Azure Databricks in your Azure virtual network (VNet inject...)", "Azure Databricks concepts", "Databricks Connect", and "Diagnostic logging in Azure Databricks". The "Azure Active Directory" section lists users: "clouduser" (User), "Unextdxc_1635523531052" (User), and "Unextdxc_1635523527625" (User). On the far right, there is a large blue arrow pointing right with the text "More services".

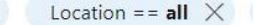
Microsoft Azure Σ 

Home >

Azure Databricks

Manipal Pro Learn (manipalazure.onmicrosoft.com)

 Create  Manage view  Refresh  Export to CSV  Open query  Assign tags

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

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

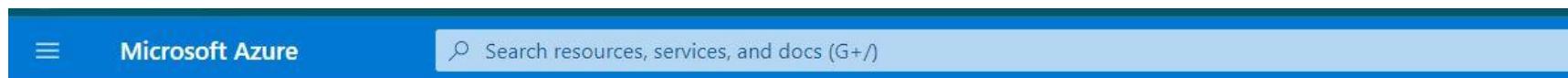


No azure databricks services to display

Unlock insights from all your data and build artificial intelligence (AI) solutions with Azure Databricks, set up your Apache Spark environment in minutes, autoscale, and collaborate on shared projects in an interactive workspace.

[Create azure databricks service](#)

[Learn more !\[\]\(ac7494f141109b59d18bf9c3aeb84d93_img.jpg\)](#)



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 * ⓘ

Azure-DXC262AB12Lab



Resource group * ⓘ

azuresynapse12



Create new

Instance Details

Workspace name *

connectbi12



Region *

East US



Pricing Tier * ⓘ

Trial (Premium - 14-Days Free DBUs)



Review + create

< Previous

Next : Networking >

Microsoft Azure Search resources, services, and docs (G+ /)

Home > azuresynapse12_connectbi12 | Overview

Deployment

Search (Ctrl+ /) Delete Cancel Redeploy Refresh

Overview Inputs Outputs Template

We'd love your feedback! →

Deployment is in progress

Deployment name: azuresynapse12_connectbi12
Subscription: Azure-DXC262AB12Lab
Resource group: azuresynapse12

Start time: 6/17/2022, 4:25:15 PM
Correlation ID: d6598333-9034-4d3e-b5da-48df3f4bc518

Deployment details (Download)

Resource	Type	Status	Operation details
No results.			

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvopro | connectbi12 - Microsoft Azure

portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3a28cdce-3bd7-4219-858e-23ff20f8b998/resourceGroups/azuresynapse12/providers/Microsoft.Databricks/w...

Microsoft Azure Search resources, services, and docs (G+/-) dxc262ab1227 1654530... MANIPAL PRO LEARN (MANIPAL...)

Home > azuresynapse12_connectbi12 >

connectbi12 Azure Databricks Service

Search (Ctrl+)/ Delete

Overview

Activity log

Access control (IAM)

Tags

Settings

Virtual Network Peerings

Encryption

Properties

Locks

Automation

Tasks (preview)

Export template

Support + troubleshooting

New Support Request

Essentials

Status : Active

Resource group : azuresynapse12

Location : East US

Subscription : Azure-DXC26AB12Lab

Subscription ID : 3a28cdce-3bd7-4219-858e-23ff20f8b998

Managed Resource Group : databricks-rg-connectbi12-iw4z2hgexnwu

URL : <https://adb-4912964121709471.11.azure.databricks.net>

Pricing Tier : Trial (Premium - 14-Days Free DBUs)

Tags (edit) : Click here to add tags

JSON View

 Launch Workspace

Upgrade to Premium

Documentation Getting Started Import Data from File Import Data from Azure Storage

SharedZpynb - Collaboratory | Subscription Details | Azure DevOps | Connectivity | Microsoft Azure | Databricks

← → C adb-4912964121709471.11.azuredatabricks.net/?o=4912964121709471#tables/new

Microsoft Azure | Databricks Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com

Get started

This is your home for all data science and engineering tasks. You can manage clusters, data and users.

- Notebook
- Table
- Cluster**
- Job
- Repo
- AutoML Experiment

Invite your team

Next steps

- Explore Notebook gallery
- Read documentation

Data Science & Engineering

Notebook
Create a new notebook for querying, data processing, and machine learning.
[Create a notebook](#)

Data import
Quickly import data, preview its schema, create a table, and query it in...
[Browse files](#)

Guide: Quickstart tutorial
Spin up a cluster, run queries on preloaded data, and display results in 5 minutes.
[Start tutorial](#)

Recents

Name	Last viewed
	 There are no recents yet

Documentation **Release notes** **Blog posts**

Free trial ends in **14** days. [Upgrade to Premium](#) in Azure Portal

Partner Connect
Fivetran, dbt
Tableau, Power BI
[View all partners](#)

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | connectbi12 - Microsoft Azure | Create Cluster - Databricks +

← → C adb-4912964121709471.11.azure.databricks.net/?o=4912964121709471#create/cluster

Microsoft Azure | Databricks Portal dxc262ab1227

Clusters / New Compute Free trial ends

New Cluster

Cancel Create Cluster DBU / hour: 0.75 0 Workers: 0 GB Memory, 0 Cores
1 Driver: 14 GB Memory, 4 Cores

Cluster name: data123

Cluster mode: Single Node

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

Promotional discount applied to Photon during preview

Use Photon Acceleration Preview

Autopilot options: Terminate after 030 minutes of inactivity

Node type: Standard_DS3_v2 14 GB Memory, 4 Cores

DBU / hour: 0.75 Standard_DS3_v2

Advanced options

The screenshot shows the 'Create Cluster' page in the Microsoft Azure Databricks portal. The 'Create Cluster' button is highlighted in green. The cluster name is 'data123', mode is 'Single Node', runtime is 'Runtime: 10.4 LTS (Scala 2.12, Spark 3.2.1)', and node type is 'Standard_DS3_v2'. A promotional discount message is displayed. Autopilot options include terminating after 030 minutes of inactivity. The DBU/hour cost is 0.75.

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | connectbi12 - Microsoft Azure | Create New Table - Databricks +

adb-4912964121709471.11.azuredatabricks.net/?o=4912964121709471#tables/new

Microsoft Azure | Databricks Portal dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com

Create New Table

Data source [?](#)

Upload File DBFS Other Data Sources

DBFS Target Directory [?](#)
/FileStore/tables/ (optional) [Select](#)

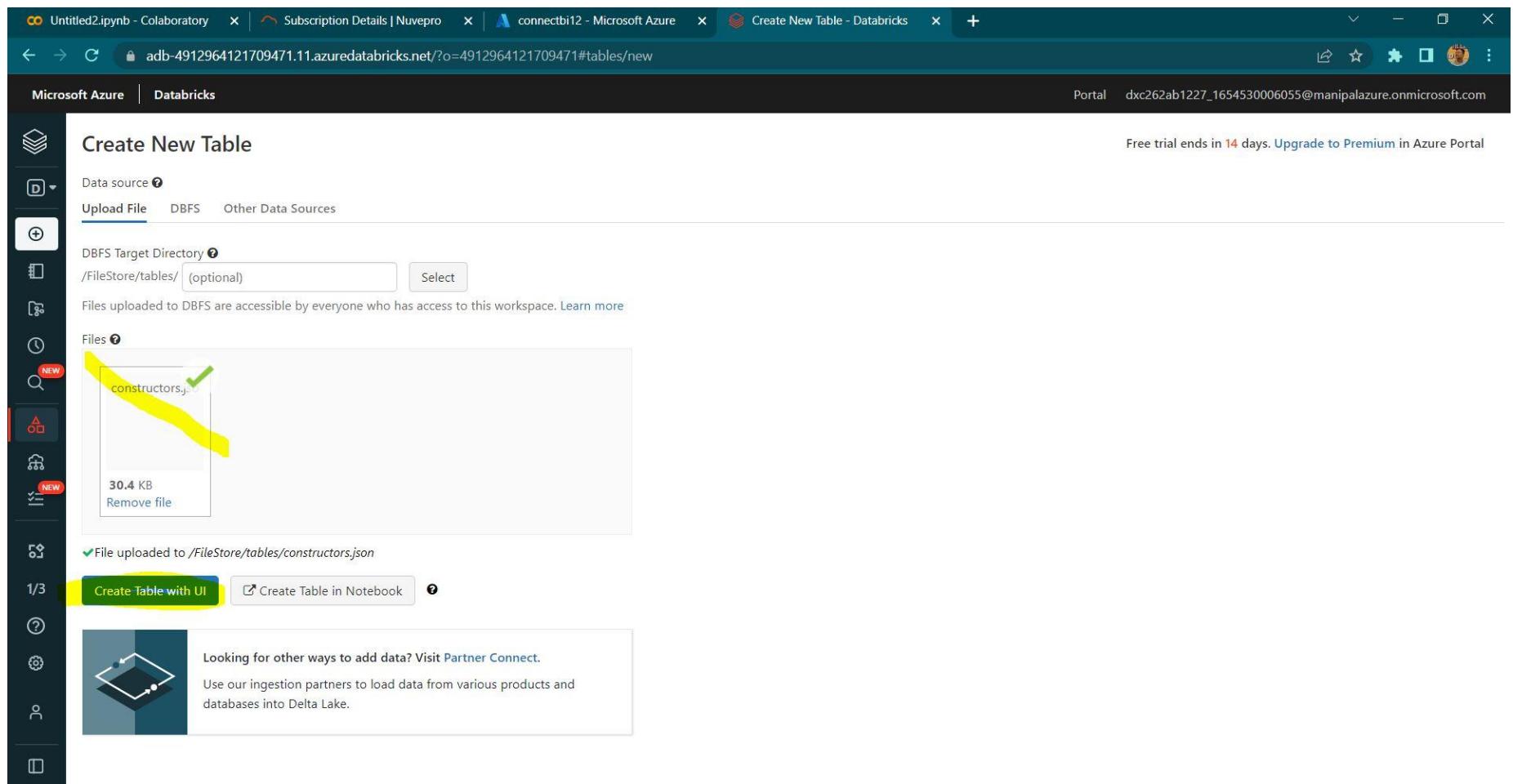
Files [?](#)

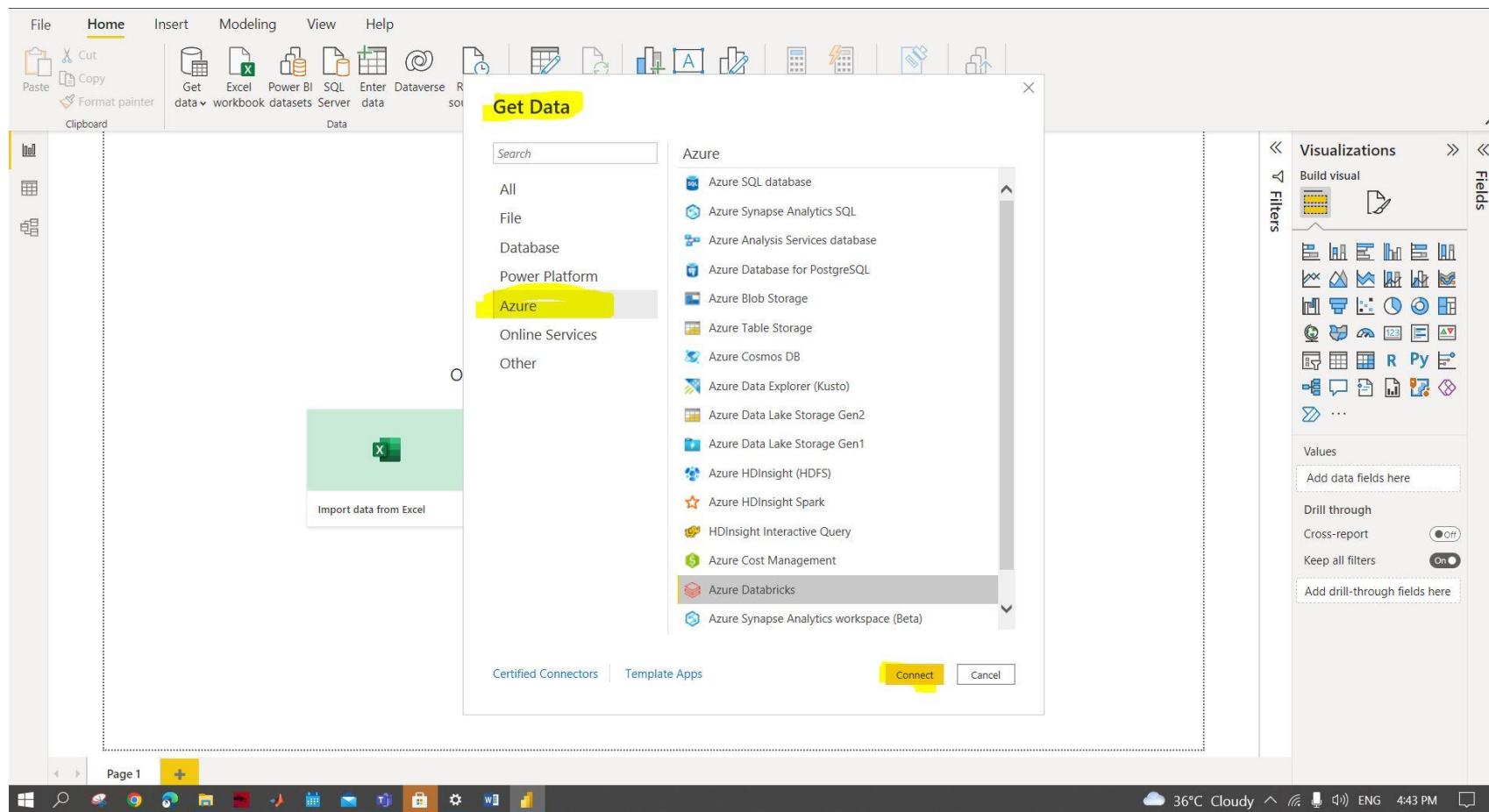
✓ constructors.json
30.4 KB Remove file

✓ File uploaded to /FileStore/tables/constructors.json

[Create Table with UI](#) [Create Table in Notebook](#) [?](#)

Looking for other ways to add data? Visit [Partner Connect](#).
Use our ingestion partners to load data from various products and databases into Delta Lake.





adb-4912964121709471.11.azuredatabricks.net/?o=4912964121709471#partnerconnect

Microsoft Azure | Databricks Portal dxc26ab1227_1654530006055@manipalazure.onmicrosoft.com

Partner Connect

Try partner solutions within minutes, even if you don't have a partner account.

Power BI Desktop All categories

BI and visualization

Microsoft Power BI

Quickly find meaningful insights within your data and easily build rich, visual analytic reports.

Before you go, make sure you have everything you need

To open the connection file, make sure you have Power BI 2.85.681.0 or above.
Download the latest version of Power BI and install it.

Open the connection file and sign in
After Power BI Desktop starts, it will prompt you for your Azure Databricks credentials. You can provide your Databricks user name and password, a Databricks personal access token, or sign in with Azure Active Directory.

Close



Partner Connect



Try partner solutions within minutes, even if you don't have a partner account.



Power BI Desktop

All categories



BI and visualization



Microsoft Power BI

Quickly find meaningful insights within your data and easily build rich, visual analytic reports.



User Settings



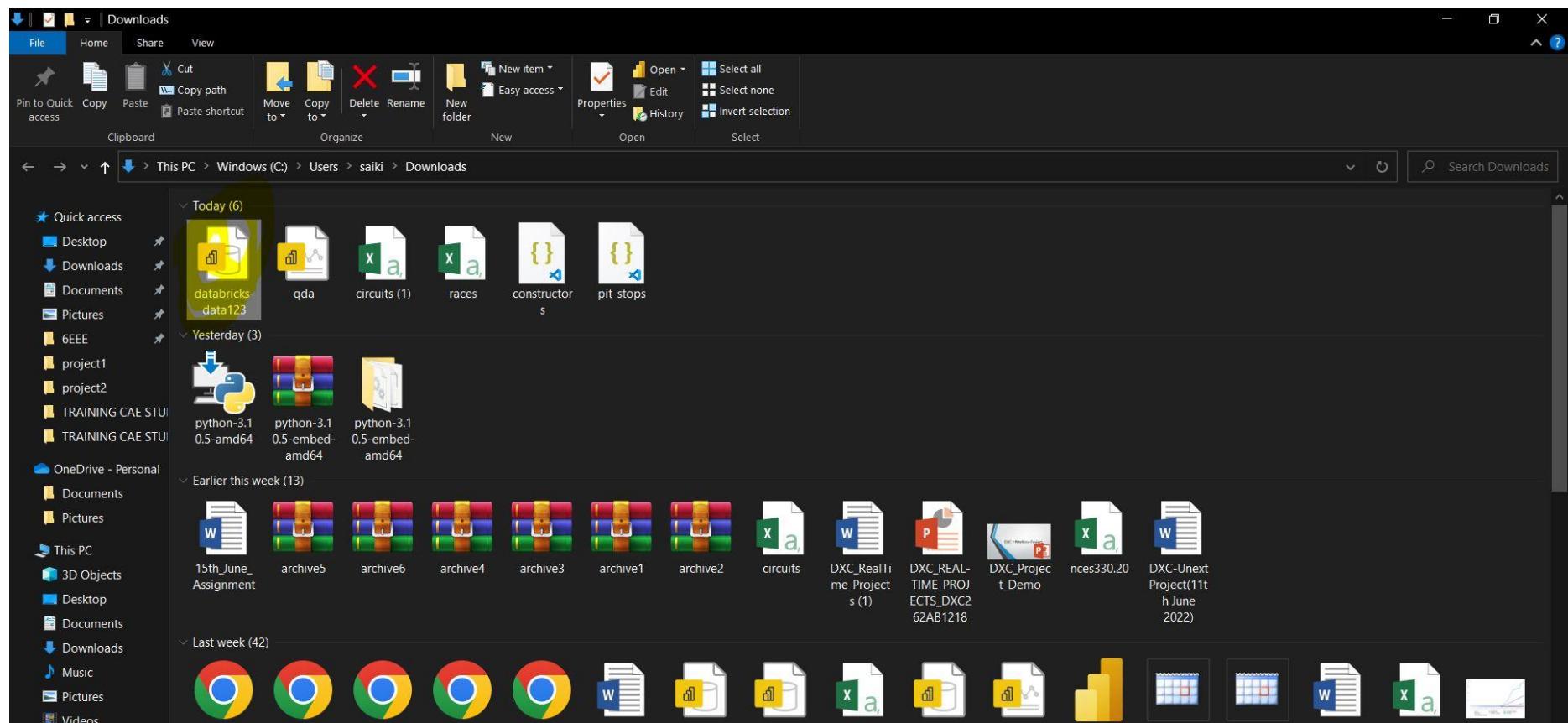
1/3 Admin Console

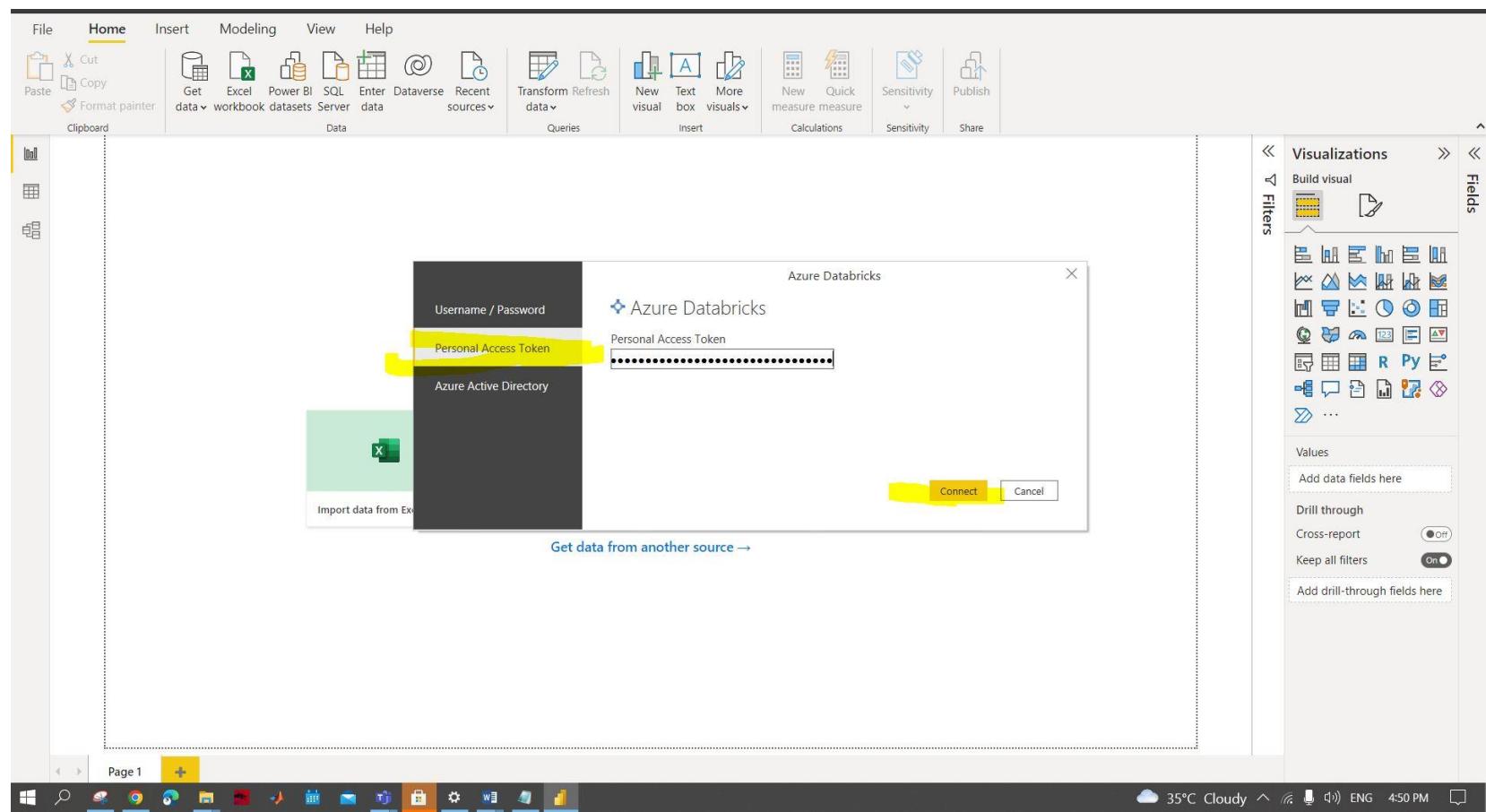


Manage Account



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





Untitled - Power BI Desktop

File Home Insert Modeling View Help

Cut Copy Format painter Paste Get data Excel Power BI workbook datasets Search

Navigator

Display Options ▾

- adb-4912964121709471.11.azuredatabricks.net [1]
- hive_metastore [1]
 - default
- samples [2]
 - nytaxi [1]
 - trips
 - tpch [8]
 - customer
 - lineitem
 - nation
 - orders
 - part
 - partsupp
 - region
 - supplier

trips

DataSource.Error: ODBC: ERROR [42000] [Microsoft][Hardy] (80) Syntax or semantic analysis error thrown in server while executing query. Error message from server: org.apache.hive.service.cli.HiveSQLException: Error running query: org.apache.spark.sql.AnalysisException: Table or view not found: samples.nytaxi.trips; line 7 pos 5
at org.apache.spark.sql.hive.thriftserver.HiveThriftServerErrors\$.runningQueryError (HiveThriftServerErrors.scala:47)
at org.apache.spark.sql.hive.thriftserver.SparkExecuteStatementOperation.org\$apa

Load Transform Data Cancel

Visualizations Build visual

Filters

Fields

Values Add data fields here

Drill through Cross-report Keep all filters

Add drill-through fields here

The screenshot shows the Power BI Desktop interface with the 'Navigator' pane open. The 'trips' dataset is selected. A yellow box highlights an error message: 'DataSource.Error: ODBC: ERROR [42000] [Microsoft][Hardy] (80) Syntax or semantic analysis error thrown in server while executing query. Error message from server: org.apache.hive.service.cli.HiveSQLException: Error running query: org.apache.spark.sql.AnalysisException: Table or view not found: samples.nytaxi.trips; line 7 pos 5'. This indicates that the system cannot find the 'trips' table in the 'samples.nytaxi' database. The 'Visualizations' pane on the right shows various chart and report options.

After refreshing I found our constructors Jason file.

The screenshot shows the Power BI Desktop interface. The top navigation bar includes File, Home, Insert, Modeling, View, and Help. The Home tab is selected. The ribbon below has options like Cut, Copy, Format painter, Get data, Excel, Power BI, and others. The main area features a 'Navigator' pane on the left with a tree view of datasets: 'adb-4912964121709471.11.azuredatabricks.net [1]', 'hive_metastore [1]', 'default [1]' (which is expanded and highlighted in yellow), and 'samples'. To the right is a data preview window titled 'constructors_json' showing a table with columns: constructorId, constructorRef, name, nationality, and url. The table lists 23 rows of constructor data. At the bottom of the preview window are buttons for Load, Transform Data, and Cancel. To the right of the preview is a 'Visualizations' pane with various chart and report icons, and a 'Fields' pane. The status bar at the bottom shows 'Page 1', a taskbar with various icons, and system information including weather (35°C Cloudy), battery level, and time (4:57 PM).

constructorId	constructorRef	name	nationality	url
1	mclaren	McLaren	British	http://en.wikipedia.org/w/index.php?title=McLaren_Mercedes&oldid=5000000
2	bmw_sauber	BMW Sauber	German	http://en.wikipedia.org/w/index.php?title=BMW_Sauber_F1_Racing&oldid=5000000
3	williams	Williams	British	http://en.wikipedia.org/w/index.php?title=Williams_Mercedes_Racing&oldid=5000000
4	renault	Renault	French	http://en.wikipedia.org/w/index.php?title=Renault_F1_Racing&oldid=5000000
5	toro_rosso	Toro Rosso	Italian	http://en.wikipedia.org/w/index.php?title=Toro_Rosso_F1_Racing&oldid=5000000
6	ferrari	Ferrari	Italian	http://en.wikipedia.org/w/index.php?title=Ferrari_F1_Racing&oldid=5000000
7	toyota	Toyota	Japanese	http://en.wikipedia.org/w/index.php?title=Toyota_F1_Racing&oldid=5000000
8	super_aguri	Super Aguri	Japanese	http://en.wikipedia.org/w/index.php?title=Super_Aguri_McLaren_F1_Racing&oldid=5000000
9	red_bull	Red Bull	Austrian	http://en.wikipedia.org/w/index.php?title=Red_Bull_Racing&oldid=5000000
10	force_india	Force India	Indian	http://en.wikipedia.org/w/index.php?title=Force_India_Mercedes_Racing&oldid=5000000
11	honda	Honda	Japanese	http://en.wikipedia.org/w/index.php?title=Honda_F1_Racing&oldid=5000000
12	spyker	Spyker	Dutch	http://en.wikipedia.org/w/index.php?title=Spyker_F1_Racing&oldid=5000000
13	mf1	MF1	Russian	http://en.wikipedia.org/w/index.php?title=MF1_Racing&oldid=5000000
14	spyker_mf1	Spyker MF1	Dutch	http://en.wikipedia.org/w/index.php?title=Spyker_MF1_Racing&oldid=5000000
15	sauber	Sauber	Swiss	http://en.wikipedia.org/w/index.php?title=Sauber_F1_Racing&oldid=5000000
16	bar	BAR	British	http://en.wikipedia.org/w/index.php?title=BAR_Mercedes_Racing&oldid=5000000
17	jordan	Jordan	Irish	http://en.wikipedia.org/w/index.php?title=Jordan_Mercedes_Racing&oldid=5000000
18	minardi	Minardi	Italian	http://en.wikipedia.org/w/index.php?title=Minardi_F1_Racing&oldid=5000000
19	jaguar	Jaguar	British	http://en.wikipedia.org/w/index.php?title=Jaguar_F1_Racing&oldid=5000000
20	prost	Prost	French	http://en.wikipedia.org/w/index.php?title=Prost_GP_F1_Racing&oldid=5000000
21	arrows	Arrows	British	http://en.wikipedia.org/w/index.php?title=Arrows_F1_Racing&oldid=5000000
22	benetton	Benetton	Italian	http://en.wikipedia.org/w/index.php?title=Benetton_F1_Racing&oldid=5000000
23	brawn	Brawn	British	http://en.wikipedia.org/w/index.php?title=Brawn_Mercedes_Racing&oldid=5000000

Untitled - Power BI Desktop

File Home Help

Clipboard Data Queries Relationships Security Q&A Publish

Get data workbook datasets Server data Transform Refresh data Manage relationships Manage roles as Q&A setup Language schema Sensitivity Share

Clipboard Data Queries Relationships Security Q&A Publish

Properties Fields

Cards

Show the database in the header when applicable

constructor_json

constructorId constructorRef name nationality url

Collapsible card settings:

- Show related fields when card is collapsed: Yes
- Pin related fields to top of card: No

All tables +

Windows taskbar: All tables, +, Cloud 35°C Cloudy, ENG, 5:00 PM

4.Create Azure Synapse & connect with Azure Blob, explain the steps with screenshots

← → C portal.azure.com/#create/Microsoft.Synapse

Microsoft Azure

Home > Azure Synapse Analytics >

Create Synapse workspace

Subscription *

Resource group *
[Create new](#)

Managed resource group

Workspace details

Name your workspace, select a location, and choose a primary Data Lake Storage Gen2 file system to serve as the default location for logs and job output.

Workspace name *

Region *

Select Data Lake Storage Gen2 * From subscription Manually via URL

Account name *
[Create new](#)

File system name *
[Create new](#)

! Additional configuration is required. After you create your workspace, perform these tasks:

- Assign other users to the **Contributor** role on workspace
- Assign other users the appropriate [Synapse RBAC roles](#) using Synapse Studio

Contact an **Owner** of the storage account, and ask them to perform the

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

Home > Azure Synapse Analytics >

Create Synapse workspace ...

Authentication

Choose the authentication method for access to workspace resources such as SQL pools. The authentication method can be changed later on. [Learn more ↗](#)

Authentication method ①

- Use both local and Azure Active Directory (Azure AD) authentication
 Use only Azure Active Directory (Azure AD) authentication

SQL Server admin login * ①

SQL Password ①

Confirm password

Password and confirm password must match.

System assigned managed identity permission

Select to grant the workspace network access to the Data Lake Storage Gen2 account using the workspace system identity.

[Learn more ↗](#)

Allow network access to Data Lake Storage Gen2 account. ①

i The selected Data Lake Storage Gen2 account does not restrict network access using any network access rules, or you selected a storage account manually via URL under Basics tab. [Learn more ↗](#)

Workspace encryption

⚠ Double encryption configuration cannot be changed after opting into using a customer-managed key at the time of workspace creation.

Review + create

< Previous

Next: Networking >

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | Microsoft.Azure.SynapseAnalytics

portal.azure.com/#view/HubsExtension/DeploymentDetailsBlade/~/overview/id/%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroups%2Fazuresynapse12%2BMANIPUL

Microsoft Azure Search resources, services, and docs (G+) ✉️ 📈 🏷️ 🛡️ 🌐 🌐 🌐 🌐

Home > Microsoft.Azure.SynapseAnalytics-20220617170302 | Overview

Deployment

Search (Ctrl+ /) « Delete Cancel Redeploy Refresh

Overview Inputs Outputs Template

We'd love your feedback! →

✓ Your deployment is complete

Deployment name: Microsoft.Azure.SynapseAnalytics-20220617170... Start time: 6/17/2022, 5:06:34 PM
Subscription: Azure-DXC262AB12Lab Correlation ID: 6e39ee15-125e-4fe9-ab81-b8764512a4ff Download

Deployment details (Download) Next steps

Go to resource group

 Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill
Set up cost alerts >

 Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud

Free Microsoft tutorials
Start learning today >

Work with an expert
Azure experts are service providers who can help manage your assets and be your first line of support.
Find an Azure expert >

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | dxblob - Microsoft Azure

portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3a28cdce-3bd7-4219-858e-23ff20f8b998/resourceGroups/azuresynapse12/providers/Microsoft.Storage/stor...

Microsoft Azure Search resources, services, and docs (G+) dx262ab1227 1654530... MANIPAL PRO LEARN

Home > Microsoft.Azure.SynapseAnalytics-20220617170302 > azuresynapse12 > dxblob

dxblob | Containers

Storage account

Search (Ctrl+I) Container Change access level Restore containers Refresh Delete

Overview Activity log Tags Diagnose and solve problems Access Control (IAM) Data migration Events Storage browser (preview)

Containers File shares Queues Tables

\$logs thesource

Name Last modified Public access level Lease state

Name	Last modified	Public access level	Lease state
\$logs	6/17/2022, 3:17:43 PM	Private	Available
thesource	6/17/2022, 3:18:09 PM	Blob	Available

Show deleted containers

35°C Cloudy ENG 5:12 PM

Containers

Name	Last modified	Public access level	Lease state
\$logs	6/17/2022, 3:17:43 PM	Private	Available
thesource	6/17/2022, 3:18:09 PM	Blob	Available

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | thesource - Microsoft Azure | +

portal.azure.com/#view/Microsoft_Azure_Storage/ContainerMenuBlade/~/overview/storageAccountId/%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroup...

Microsoft Azure Search resources, services, and docs (G+) dxc262ab1227 1654530... MANIPAL PRO LEARN

Home > Microsoft.Azure.SynapseAnalytics-20220617170302 > auresynapse12 > dxcblob >

thesource Container

Search (Ctrl+J) Upload Change access level Refresh Delete Change tier Acquire lease Break lease View snapshots Create snapshot

Overview Diagnose and solve problems Access Control (IAM)

Authentication method: Access key (Switch to Azure AD User Account)
Location: thesource

Search blobs by prefix (case-sensitive) Show deleted blobs

Add filter

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
circuits (1).csv	6/17/2022, 3:18:32 PM	Hot (Inferred)		Block blob	9.81 KiB	Available
constructors.json	6/17/2022, 3:18:32 PM	Hot (Inferred)		Block blob	29.7 KiB	Available
pit_stops.json	6/17/2022, 3:18:37 PM	Hot (Inferred)		Block blob	1.31 MiB	Available
races.csv	6/17/2022, 3:18:33 PM	Hot (Inferred)		Block blob	114.11 KiB	Available

Coordinator replied to a conversation you're in
DXC-262-Analytics-B12-Azure / General

Microsoft Azure | sqladminuser1234

Synapse Analytics workspace
sqladminuser1234

New ▾

Ingest Explore and analyze Visualize

Discover more

Knowledge center Browse partners

Recent resources

35°C Cloudy 5:14 PM

The screenshot displays the Microsoft Azure Synapse Analytics workspace interface. At the top, there's a blue header bar with the Microsoft Azure logo, the workspace name 'sqladminuser1234', and a user profile icon. Below the header, the title 'Synapse Analytics workspace' and the workspace name 'sqladminuser1234' are prominently displayed. A 'New' button with a dropdown arrow is located just below the workspace title. The main content area features a large, stylized globe graphic with a bar chart overlaid, representing data analysis. Three main action buttons are shown: 'Ingest' (Perform a one-time or scheduled data load), 'Explore and analyze' (Learn how to get insights from your data), and 'Visualize' (Build interactive reports with Power BI capabilities). Below these buttons, there are links for 'Discover more' (Knowledge center and Browse partners) and 'Recent resources'. The bottom of the screen shows a standard Windows taskbar with various pinned icons and system status indicators like weather and battery level.

The screenshot shows the Microsoft Azure Synapse workspace interface. The left sidebar has a 'Data' tab selected, which is highlighted with a yellow box. Under 'Data', there are several options: 'Workspace' (selected), 'Linked' (highlighted with a yellow box), 'Connect to external data' (highlighted with a yellow box), 'Integration dataset', and 'Browse gallery'. A search bar at the top says 'Search' and a user profile 'MANIPAL PRO LEARN' is in the top right. The main area is a placeholder with the text 'Select an item' and a small icon of a cylinder and a screen.

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Az | sqladminuser1234 - Azure Synap | +

Microsoft Azure | sqladminuser1234 | Search | : | dxc26ab1227_1654530006055@manipalazure.onmicrosoft.com | MANIPAL PRO LEARN

Synapse live | Validate all | Publish all

Data | Workspace | Linked | Filter resources by name | Azure Data Lake Storage Gen2 | 2

Connect to external data

Once a connection is created, the underlying data of that connection will be available for analysis in the Data hub or for pipeline activities in the Integrate hub.

Azure Blob Storage	Azure Cosmos DB (MongoDB API)	Azure Cosmos DB (SQL API)
Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen2	

Use the resource exp

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Azure | sqladminuser1234 - Azure Synapse | Settings

Microsoft Azure | sqladminuser1234 | Search | MANIPAL PRO LEARN

web.azure-synapse.net/en/authoring/explore/linked?workspace=%2Fsubscriptions%2Fa28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroups%2Fazuresynapse12%2Fproviders%2FMicr...

Synapse live | Validate all | Publish all

Data | Workspace | Linked | Filter resources by name | Azure Data Lake Storage Gen2 | 2

New linked service

Azure Blob Storage | Learn more

Connect via integration runtime * | AutoResolveIntegrationRuntime

Authentication type | Account key

Connection string | Azure Key Vault

Account selection method | From Azure subscription | Enter manually

Azure subscription | Azure-DXC26AB12Lab (3a28cdce-3bd7-4219-858e-23ff20f8b998)

Storage account name * | dxcblob

Additional connection properties | New

Test connection | To linked service | To file path

Annotations | New

Parameters | New

Connection successful

35°C Cloudy | ENG | 5:18 PM

The screenshot shows the Microsoft Azure portal interface for configuring a linked service. The left sidebar shows a 'Data' section with 'Workspace' and 'Linked' tabs, and a list of 'Azure Data Lake Storage Gen2' resources. The main right pane is titled 'New linked service' and is configured for 'Azure Blob Storage'. Under 'Authentication type', 'Account key' is selected. The 'Connection string' tab is active. In the 'Account selection method' section, 'From Azure subscription' is chosen, and the dropdown shows 'Azure-DXC26AB12Lab (3a28cdce-3bd7-4219-858e-23ff20f8b998)'. The 'Storage account name' field contains 'dxcblob'. At the bottom, a green success message 'Connection successful' is visible.

The screenshot shows the Microsoft Azure Synapse Studio interface. The top navigation bar includes the URL web.azure-synapse.net/en/authoring/explore/linked?workspace=%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroups%2Fazuresynapse12%2Fproviders%2FMicr..., a search bar, and a user profile icon for dxc262ab1227_1654530006055@manipalazure.onmicrosoft.com MANIPAL PRO LEARN.

The left sidebar features a navigation menu with icons for Home, Data, Machine Learning, Pipelines, Databricks, and More. The "Data" section is selected, showing a "Workspace" tab and a "Linked" tab. A search bar allows filtering resources by name. Below the tabs, a list shows "Azure Data Lake Storage Gen2" with a count of 2.

The main content area displays a large, semi-transparent placeholder icon of two cylinders and a screen with code brackets (</>), indicating "Select an item". Below this, a sub-instruction reads: "Use the resource explorer to select or create a new item".

The bottom of the screen shows the Windows taskbar with various pinned application icons, and a system tray at the bottom right showing weather (35°C Cloudy), battery status, and connectivity information.

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Az | sqladminuser1234 - Azure Synap | Settings | +

Microsoft Azure | sqladminuser1234 | dxc26ab1227_1654530006055@manipalazure.onmicrosoft.com | MANIPAL PRO LEARN

Synapse live | Validate all | Publish all

Data

Workspace Linked

Filter resources by name

Azure Blob Storage 1

- AzureBlobStorage1 (dxcblob)
 - thesource

Azure Data Lake Storage Gen2 2

thesource

New SQL script | New notebook | New data flow | New integration dataset | Upload | Download | New folder | More

theresource

Name	Access Tier	Access Tier Last Modified	Last Modified	Blob Type	Content Type	Size	Status	Remaining Days	Deleted Time	Lease State
circuits (1).csv	Hot		6/17/2022, 3:18:32 PM	Block Blob	text/csv	9.8 KB	Active			
constructors.json	Hot		6/17/2022, 3:18:32 PM	Block Blob	application/json	29.7 KB	Active			
pit_stops.json	Hot		6/17/2022, 3:18:37 PM	Block Blob	application/json	1.3 MB	Active			
races.csv	Hot		6/17/2022, 3:18:33 PM	Block Blob	text/csv	114.1 KB	Active			

5.Create Azure Synapse spark pool & query sample JSON file,
explain the steps with screenshots

The screenshot shows the Microsoft Azure Synapse Analytics studio interface. The left sidebar displays the 'Data' workspace under the 'Linked' tab. A folder named 'thesource' is selected. The main area shows a table of files from 'AzureBlobStorage1 (dxcblob)'. The 'constructors.json' file is highlighted. The toolbar at the top has several options: 'New SQL script', 'New notebook' (which is highlighted), 'New data flow', 'New integration dataset', 'Upload', 'Download', 'New folder', and 'More'. A search bar and a 'Search by prefix (case-sensitive)' input field are also present.

Name	Access	New Spark table	Modified	Last Modified	Blob Type	Content Type	Size	Status	Remaining Days	Deleted Time	Lease State
circuits (1).csv	Hot			6/17/2022, 3:18:32 PM	Block Blob	text/csv	9.8 KB	Active			
constructors.json	Hot			6/17/2022, 3:18:32 PM	Block Blob	application/json	29.7 KB	Active			
pit_stops.json	Hot			6/17/2022, 3:18:37 PM	Block Blob	application/json	1.3 MB	Active			
races.csv	Hot			6/17/2022, 3:18:33 PM	Block Blob	text/csv	114.1 KB	Active			

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Az | sqladminuser1234 - Azure Synap | Settings | +

Microsoft Azure | sqladminuser1234 | Search | ? | Help | dxc26ab1227_1654530006055@manipalazure.onmicrosoft.com | MANIPAL PRO LEARN

Synapse live | Validate all | Publish all 1

Integrate | thesource | Notebook 1 | Run all | Undo | Publish | Outline | Attach to | Select Apache Spark pool | Language | PySpark (Python) | Variables | Manage pools | ...

Please select a Spark pool to attach before running cell!

```
1 %pyspark
2 blob_account_name = "dxcblob"
3 blob_container_name = "thesource"
4 from pyspark.sql import SparkSession
5
6 sc = SparkSession.builder.getOrCreate()
7 token_library = sc._jvm.com.microsoft.azure.synapse.tokenlibrary.TokenLibrary
8 blob_sas_token = token_library.getConnectionString("AzureBlobStorage1")
9
10 spark.conf.set(
11     'fs.azure.sas.%s.%s.blob.core.windows.net' % (blob_container_name, blob_account_name),
12     blob_sas_token)
13 df = spark.read.load('wasbs://thesource@dxcblob.blob.core.windows.net/constructors.json', format='json')
14 display(df.limit(10))
```

No items to show
Try creating a new item using the + button above. Learn more

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Az | sqladminuser1234 - Azure Synap | Settings | + | - | X

Microsoft Azure | sqladminuser1234 | web.azureynapse.net/en/management/apachesparkpools?workspace=%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroups%2Fazuresynapse12%2Fprovider... | dxc26ab1227_1654530006055@manipalazure.onmicrosoft.com | MANIPAL PRO LEARN

Synapse live | Validate all | Publish all (1)

Analytics pools | SQL pools | **Apache Spark pools** | Data Explorer pools (preview) | External connections | Linked services | Microsoft Purview | Integration | Triggers | Integration runtimes | Security | Access control | Credentials | Managed private endpoints | Configurations + libraries | Workspace packages | Data flow libraries (preview) | Apache Spark configurations | Source control

Apache Spark pool

Apache Spark pools can be tuned to run different kinds of Apache Spark workloads using specific configuration libraries, permissions, etc. Learn more

+ New | Refresh | Filter by name

Showing 0-0 of 0 item

Name	Node size family	Size
No items to show		

Try changing your filter or create new Apache Spark pool

New Apache Spark pool

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Az | sqladminuser1234 - Azure Synap | Settings | + | - | X

Microsoft Azure | sqladminuser1234 | web.azure-synapse.net/en/management/apachesparkpools?workspace=%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23ff20f8b998%2FresourceGroups%2Fazuresynapse12%2Fprovider... | dxc26ab1227_1654530006055@manipalazure.onmicrosoft.com | MANIPAL PRO LEARN

Synapse live | Validate all | Publish all

Analytics pools | SQL pools | Apache Spark pools | Data Explorer pools (preview) | External connections | Linked services | Microsoft Purview | Integration | Triggers | Integration runtimes | Security | Access control | Credentials | Managed private endpoints | Configurations + libraries | Workspace packages | Data flow libraries (preview) | Apache Spark configurations | Source control | Git configuration

New Apache Spark pool

Basics * Additional settings * Tags Review + create

Create an Synapse Analytics Apache Spark pool with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults, or visit each tab to customize.

Apache Spark pool details

Name your Apache Spark pool and choose its initial settings.

Apache Spark pool name * assignjason

Isolated compute * Enabled Disabled

Node size family * Memory Optimized

Node size * Small (4 vCores / 32 GB)

Autoscale * Enabled Disabled

Number of nodes * 3

Estimated price Est. cost per hour Failed to fetch billing info

Dynamically allocate executors * Enabled Disabled

35°C Cloudy | ENG | 5:34 PM

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | sqladminuser1234 - Microsoft Azure | sqladminuser1234 - Azure Synapse | Settings | +

← → ⌂ 🔍 web.azure-synapse.net/en/authoring/explore/linked/notebooks/Notebook%201?workspace=%2Fsubscriptions%2F3a28cdce-3bd7-4219-858e-23f20f8b998%2FresourceGroups%2Fazur...

Microsoft Azure | sqladminuser1234

Search

Synapse live | Validate all | Publish all (1)

Data

Workspace | Linked

Filter resources by name

Azure Blob Storage (1)

AzureBlobStorage1 (dxcblob)

thesource

Azure Data Lake Storage Gen2 (2)

theresource

Notebook 1

Run all | Undo | Publish | Outline | Attach to | Select Apache

Not started

Please select a Spark pool to attach before running cell!

```
1 %%pyspark
2 blob_account_name = "dxcblob"
3 blob_container_name = "thesource"
4 from pyspark.sql import SparkSession
5
6 sc = SparkSession.builder.getOrCreate()
7 token_library = sc._jvm.com.microsoft.azure.synapse.token
8 blob_sas_token = token_library.getConnectionString("Azur
9
10 spark.conf.set(
11     'fs.azure.sas.%s.%s.blob.core.windows.net' % (blob_
12         blob_sas_token)
13 df = spark.read.load('wasbs://thesource@dxcblob.blob.cor
14 display(df.limit(10))
```

Press shift + enter to run

+ Code

Notifications

Dismiss all

Deploying (Yellow Alert)

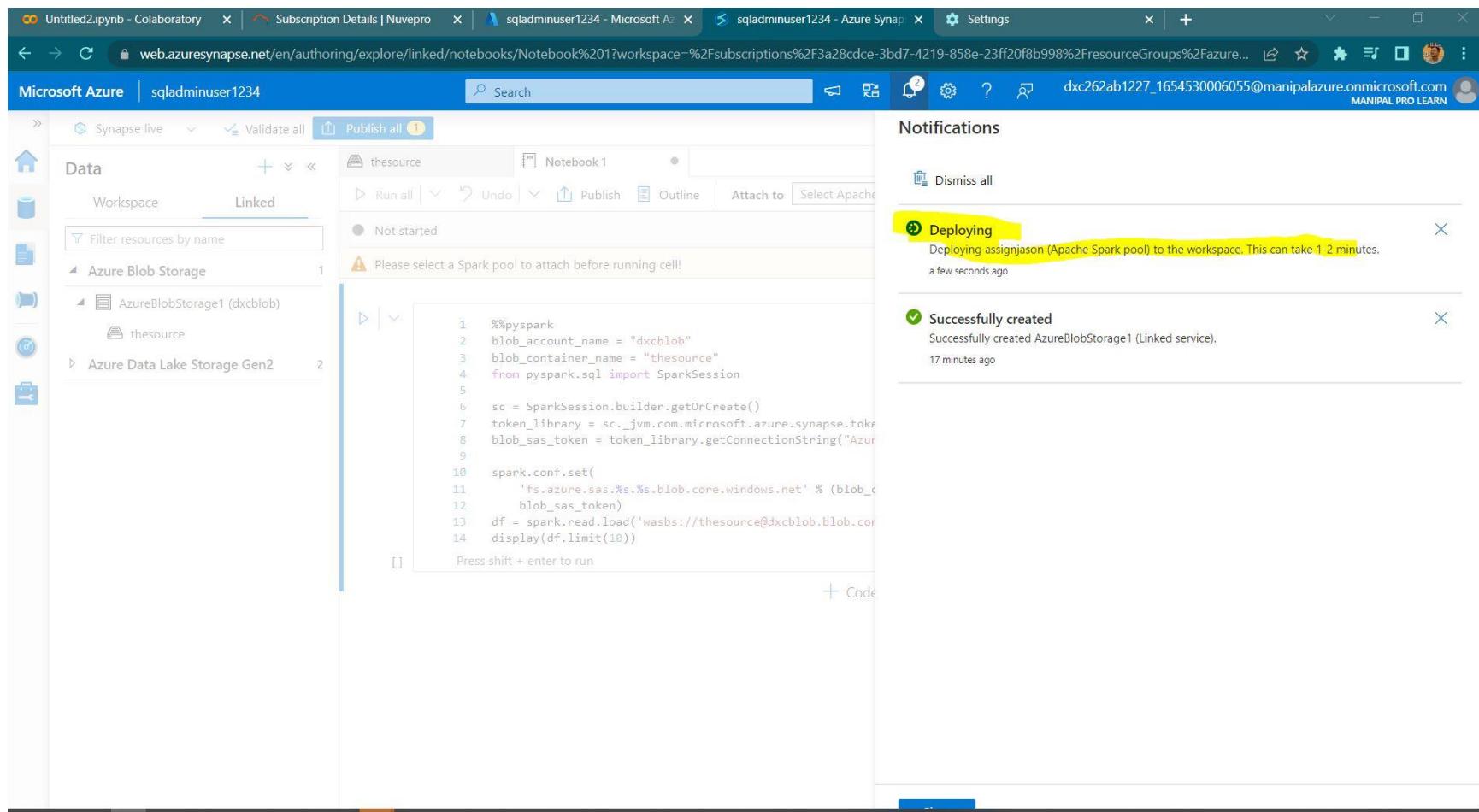
Deploying assignjason (Apache Spark pool) to the workspace. This can take 1-2 minutes.

a few seconds ago

Successfully created (Green Success)

Successfully created AzureBlobStorage1 (Linked service).

17 minutes ago



The screenshot shows the Microsoft Azure Synapse Notebooks interface. The left sidebar displays a 'Data' section with 'Workspace' and 'Linked' tabs, and a list of resources under 'Azure Blob Storage' and 'Azure Data Lake Storage Gen2'. The main area shows a notebook titled 'Notebook 1' with the following code:

```
1 %%pyspark
2 blob_account_name = "dxcblob"
3 blob_container_name = "thesource"
4 from pyspark.sql import SparkSession
5
6 sc = SparkSession.builder.getOrCreate()
7 token_library = sc._jvm.com.microsoft.azure.synapse.tokenlibrary.TokenLibrary
8 blob_sas_token = token_library.getConnectionString("AzureBlobStorage1")
9
10 spark.conf.set(
11     'fs.azure.sas.%s.%s.blob.core.windows.net' % (blob_container_name, blob_account_name),
12     blob_sas_token)
13 df = spark.read.load('wasbs://thesource@dxcblob.blob.core.windows.net/constructors.json', format='json')
14 display(df.limit(10))
```

The code uses PySpark to read a JSON file from a blob storage container named 'thesource' in the 'dxcblob' account. The notebook is currently set to run in 'PySpark (Python)' mode.

The screenshot shows the Microsoft Azure Synapse Notebook interface. The left sidebar displays 'Data' resources, including 'Workspace' and 'Linked' storage accounts like 'Azure Blob Storage' and 'Azure Data Lake Storage Gen2'. The main workspace contains a notebook titled 'Notebook 1' with the following PySpark code:

```
%pyspark
blob_account_name = "dxcblob"
blob_container_name = "thesource"
from pyspark.sql import SparkSession

sc = SparkSession.builder.getOrCreate()
token_library = sc._jvm.com.microsoft.azure.synapse.tokenlibrary.TokenLibrary
blob_sas_token = token_library.getConnectionString("AzureBlobStorage1")

spark.conf.set(
    'fs.azure.sas.%s.%s.blob.core.windows.net' % (blob_container_name, blob_account_name),
    blob_sas_token)
df = spark.read.load('wasbs://thesource@dxcblob.blob.core.windows.net/constructors.json', format='json')
display(df.limit(10))
```

The code output indicates a successful execution: [1] Job execution Succeeded Spark 2 executors 8 cores. Below the code, a table view shows the first five rows of the 'constructors' dataset:

constructorId	constructorRef	name	nationality	url
1	mclaren	McLaren	British	http://en.wikipedia.org/w
2	bmw_sauber	BMW Sauber	German	http://en.wikipedia.org/w
3	williams	Williams	British	http://en.wikipedia.org/w
4	renault	Renault	French	http://en.wikipedia.org/w
5	toro_rosso	Toro Rosso	Italian	http://en.wikipedia.org/w

6.Create Azure Cosmos DB & import sample JSON file, explain the steps with screenshots

The screenshot shows the Microsoft Azure portal interface. At the top, there is a search bar with the text "cosom". Below the search bar, the results are displayed under the heading "All". The first result, "Azure Cosmos DB", is highlighted with a yellow box. Other results include "Azure Managed Instance for Apache Cassandra" and "Azure Cosmos DB API for MongoDB". The left sidebar contains sections for Services, Marketplace, Documentation, and Resource groups. The bottom navigation bar includes links for Subscriptions, Resource groups, All resources, Dashboard, and Tools, which include Microsoft Learn, Azure Monitor, Microsoft Defender for Cloud, and Cost Management.

��්‍රිංකලාපිත තොරතුරු | Subscription Details | Azure Active Directory | Azure Cosmos DB - Microsoft | Azure Synapse | Settings

← → C portal.azure.com/#view/HubsExtension/BrowseResource/resourceType/Microsoft.DocumentDb%2FdatabaseAccounts

Microsoft Azure Search resources, services, and docs (G+)

MANIPAL PRO LEARN dxc262ab1227_1654530...

Home > Azure Cosmos DB Manipal Pro Learn

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

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

No grouping List view

Name ↑↓	Status ↑↓	Subscription ↑↓	Write Region ↑↓	Read Region ↑↓
---------	-----------	-----------------	-----------------	----------------

No Azure Cosmos DB accounts to display

Create a globally distributed, multi-model, fully managed database using API of your choice. Or try it for free, up to 20k RU/s, for 30 days with unlimited renewal.

Create Azure Cosmos DB account Try now ↗



Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | Select API option - Microsoft Azure | sqladminuser1234 - Azure Synap | Settings

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

Microsoft Azure Search resources, services, and docs (G+)

Home > Azure Cosmos DB >

Select API option

Which API best suits your workload?

Azure Cosmos DB is a fully managed NoSQL database service for building scalable, high performance applications. [Learn more](#)

To start, select the API to create a new account. The API selection cannot be changed after account creation.

Core (SQL) - Recommended

Azure Cosmos DB's core, or native API for working with documents. Supports fast, flexible development with familiar SQL query language and client libraries for .NET, JavaScript, Python, and Java.

[Create](#) [Learn more](#)

Azure Cosmos DB API for MongoDB

Fully managed database service for apps written for MongoDB. Recommended if you have existing MongoDB workloads that you plan to migrate to Azure Cosmos DB.

[Create](#) [Learn more](#)

Cassandra

Fully managed Cassandra database service for apps written for Apache Cassandra. Recommended if you have existing Cassandra workloads that you plan to migrate to Azure Cosmos DB.

[Create](#) [Learn more](#)

Azure Table

Fully managed database service for apps written for Azure Table storage. Recommended if you have existing Azure Table storage workloads that you plan to migrate to Azure Cosmos DB, but do not want to re-write your application to use the SQL API.

[Create](#) [Learn more](#)

Gremlin (Graph)

Fully managed graph database service using the Gremlin query language, based on Apache TinkerPop project. Recommended for new workloads that need to store relationships between data.

[Create](#) [Learn more](#)

Microsoft Azure Search resources, services, and docs (G+/-)

Home > Azure Cosmos DB > Select API option >

Create Azure Cosmos DB Account - Core (SQL) ...

Basics Global Distribution Networking Backup Policy Encryption Tags Review + create

Azure Cosmos DB is a fully managed NoSQL database service for building scalable, high performance applications. [Try it for free](#), for 30 days with unlimited renewals. Go to production [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 * azuresynapse12 [Create new](#)

Instance Details

Account Name * cosmosdatajason

Location * (US) West US

Capacity mode Provisioned throughput Serverless [Learn more about capacity mode](#)

Review + create Previous Next: Global Distribution

Microsoft Azure Search resources, services, and docs (G+) MANIPAL PRO LEARN

Home > Microsoft.Azure.CosmosDB-20220617175323 | Overview X

 Deployment

Delete Cancel Redeploy Refresh

Overview Inputs Outputs Template

 We'd love your feedback! →

Deployment is in progress

 Deployment name: Microsoft.Azure.CosmosDB-20220617175323
Subscription: Azure-DXC262AB12Lab
Resource group: auresynapse12

Start time: 6/17/2022, 5:53:28 PM
Correlation ID: c9bf0338-dcd2-4999-9148-863bf5053f6b 

 Deployment details [\(Download\)](#)

Resource	Type	Status	Operation details
 cosmosdatajson	Microsoft.DocumentDb/databaseAccounts	OK	Operation details

Microsoft Azure Search resources, services, and docs (G+) 3 MANIPAL PRO LEARN

Home > cosmosdatajason

cosmosdatajason | Data Explorer

Azure Cosmos DB account

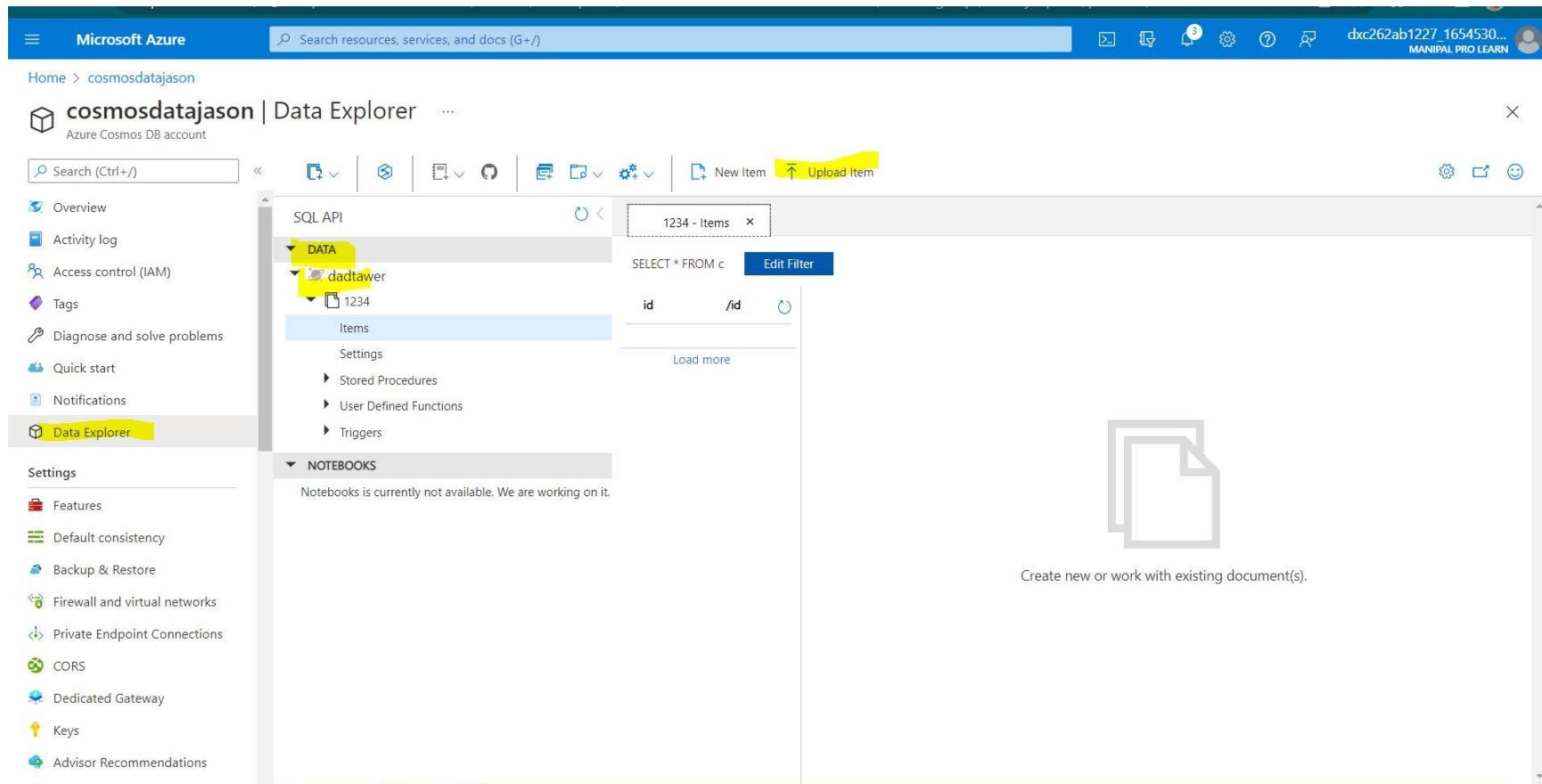
Search (Ctrl+ /) Upload Item

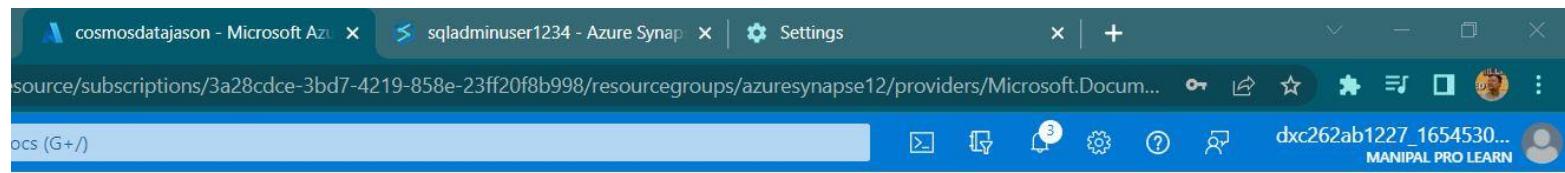
Overview Activity log Access control (IAM) Tags Diagnose and solve problems Quick start Notifications Data Explorer

SQL API DATA dadtawer 1234 Items id /id Load more Settings Stored Procedures User Defined Functions Triggers

NOTEBOOKS Notebooks is currently not available. We are working on it.

Create new or work with existing document(s).





A screenshot of the Azure Cosmos DB portal. On the left, a table view shows a single row with columns "id" and "/id". The "id" column contains "1234 - Items". On the right, a modal window titled "Upload Items" is open, showing a file input field with "pit_stops.json" selected. A yellow highlight is drawn around this file name. In the bottom right corner of the screen, a dark blue notification bubble appears, containing the text "Rohit replied to a conversation you're in" and "DXC-262-Analytics-B12-Azure / General".

Untitled2.ipynb - Colaboratory | Subscription Details | Nuvepro | cosmosdatajason - Microsoft Azure | sqladminuser1234 - Azure Synapse | Settings

portal.azure.com/#@manipalazure.onmicrosoft.com/resource/subscriptions/3a28cdce-3bd7-4219-858e-23ff20f8b998/resourcegroups/azuresynapse12/providers/Microsoft.DocumentDB/cosmosdbAccounts/cosmosdatajason

Microsoft Azure Search resources, services, and docs (G+)

Home > cosmosdatajason

cosmosdatajason | Data Explorer

Azure Cosmos DB account

Search (Ctrl+J) New Container Enable Azure Synapse Link New Notebook Connect to GitHub

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Quick start Notifications Data Explorer

SQL API DATA NOTEBOOKS Notebooks is currently not available. We are working on it.

Welcome to C

Globally distributed, multi-model data

Launch quick start New Container

Recents Top 3 things you need to know

Advanced Modeling Patterns Partitioning Best Practices

OK

New Container

Container Id: dadtawer

Partition key: /id

Unique keys

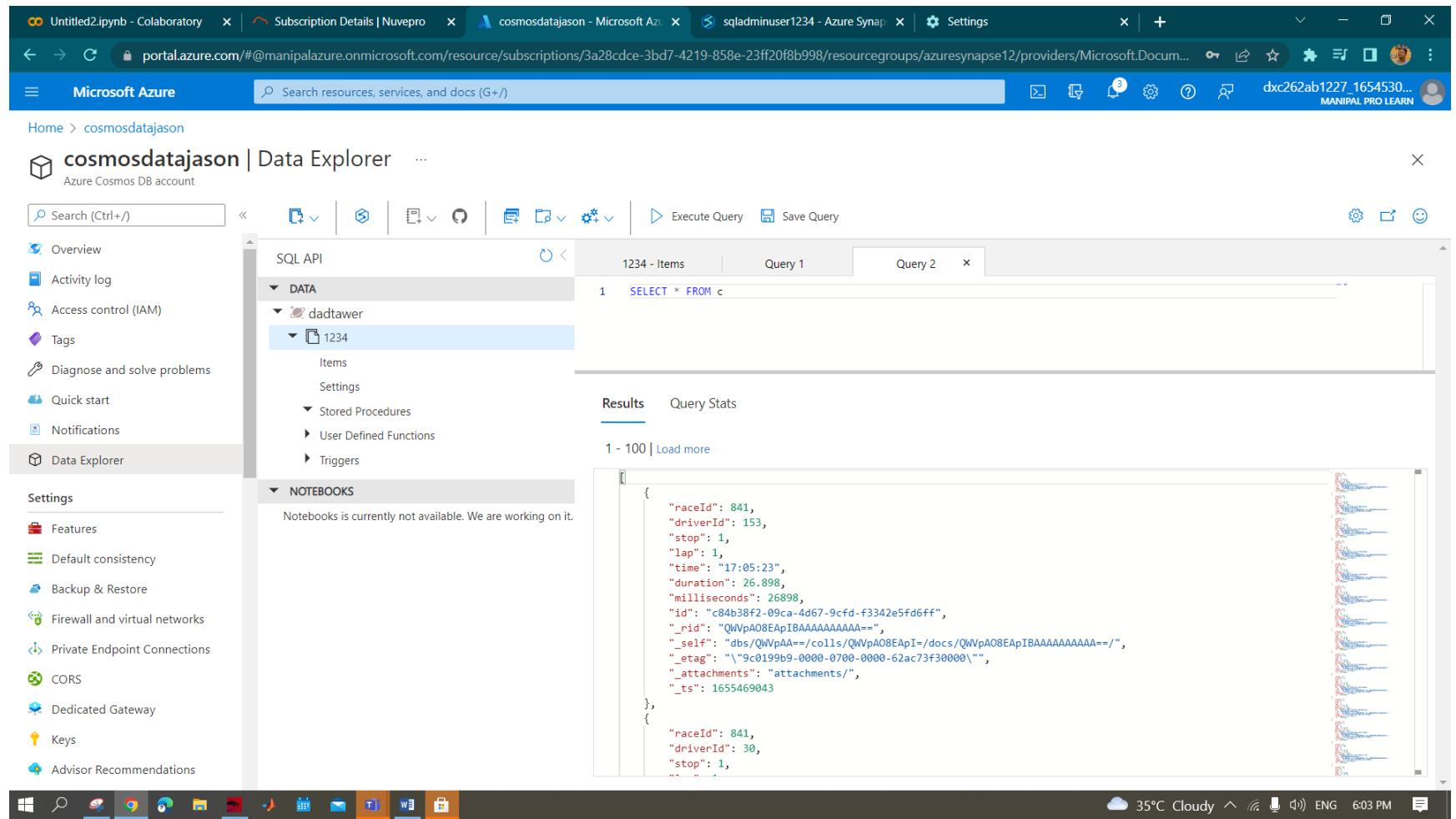
Add unique key

Analytical store: Off

Enable

35°C Cloudy 5:58 PM

The screenshot shows the Azure Data Explorer interface for the 'cosmosdatajason' account. The left sidebar includes links for Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Quick start, Notifications, and Data Explorer (which is selected). The main area features a 'Welcome to C' message and sections for Launch quick start and New Container. A modal dialog for creating a new container is open, asking for Container Id (1234), Partition key (/id), and Unique keys. The status bar at the bottom provides weather information (35°C Cloudy) and system details (ENG, 5:58 PM).



7. Connect COSMOS DB & Azure Synapse analytics & explain the steps with screenshots

The screenshot shows the Azure Synapse Analytics workspace interface. On the left, the Data workspace sidebar is open, showing options like Workspace, Azure Blob Storage, and Azure Data Lake Storage. A notebook titled "Notebook 1" is currently selected. The notebook area displays the following PySpark code:

```
%pyspark
blob_account_name = "dxcstorageaccount12345"
blob_container_name = "sourceblob1"
from pyspark.sql import SparkSession

sc = SparkSession.builder.getOrCreate()
token_library = sc._jvm.com.microsoft.azure.synapse.tokenlibrary.TokenLibrary
blob_sas_token = token_library.getConnectionString("AzureBlobStorage1")

spark.conf.set(
    'fs.azure.sas.%s.%s.blob.core.windows.net' % (blob_container_name, blob_account_name),
    blob_sas_token)
df = spark.read.load('wasbs://sourceblob1@dxcstorageaccount12345.blob.core.windows.net/races (1).csv', format='csv')
## If header_exists uncomment line below
##, header=True
)
display(df.limit(10))
```

The code is executing successfully, as indicated by the green checkmark icon and the message "3 sec - Command executed in 2 sec 827 ms by dxc262ab1215_1654530060194 on 3:31:50 PM, 6/17/22". Below the code, the resulting DataFrame is displayed in a table format:

_c0	_c1	_c2	_c3	_c4
raceld	year	round	circuitId	name
1	2009	1	1	Australian Grand Prix

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live Validate all Publish all 1

Data

Workspace Linked

Filter resources by name

- Azure Blob Storage 3
- Azure Cosmos DB 1
 - CosmosDb1 (ToDoList) 1
 - Items
- Azure Data Lake Storage Gen2 2

Notebook 1

Run all Undo Publish Outline Attach to Select Apache

Not started

Please select a Spark pool to attach before running cell!

+ Code + Markdown

```
1 # Load a streaming Spark DataFrame from a Cosmos DB container
2 # To select a preferred list of regions in a multi-region cluster
3
4 # For Spark 2.4+
5 dfStream = spark.readStream\
6   .format("cosmos.oltp")\
7   .option("spark.synapse.linkedService", "CosmosDb1")\
8   .option("spark.cosmos.container", "Items")\
9   .option("spark.cosmos.changeFeed.readEnabled", "true")\
10  .option("spark.cosmos.changeFeed.startFromTheBeginning", "true")\
11  .option("spark.cosmos.changeFeed.checkpointLocation", "file:///tmp/cosmos_changefeed_checkpoint")\
12  .option("spark.cosmos.changeFeed.queryName", "stream")\
13  .load()
14
15 # For Spark 3.1+
16 # dfStream = spark.readStream\
17 #   .format("cosmos.oltp.changeFeed")\
18 #   .option("spark.synapse.linkedService", "CosmosDb1")\
19 #   .option("spark.cosmos.container", "Items")\
20 #   .option("spark.cosmos.changeFeed.startFrom", "Beginning")\
21 #   .option("spark.cosmos.changeFeed.mode", "Incremental")\
22 #   .load()
23
```

Press shift + enter to run

Continue Cancel

Connect to external data

Once a connection is created, the underlying data of that connection will be available for analysis in the Data hub or for pipeline activities in the Integrate hub.

		
Azure Blob Storage	Azure Cosmos DB (MongoDB API)	Azure Cosmos DB (SQL API)
		
Azure Data Explorer (Kusto)	Azure Data Lake Storage Gen2	

We use optional cookies to provide a better experience. [Learn more](#)

Synapse live Validate all Publish all 1

Data Workspace Linked

Filter resources by name

Azure Blob Storage 3

- AzureBlobStorage1 (dxcstorageacc...)
- AzureBlobStorage2 (dxcstorageacc...)
- Sample Datasets

Azure Data Lake Storage Gen2 2

Notebook 1

sourceblob1

Run all Undo Publish Outline Attach to dxsparkdem...

Session timed out. Run the notebook to start a new session.

```
1 %%pyspark
2 blob_account_name = "dxcstorageaccount12345"
3 blob_container_name = "sourceblob1"
4 from pyspark.sql import SparkSession
5
6 sc = SparkSession.builder.getOrCreate()
7 token_library = sc._jvm.com.microsoft.azure.synapse.tok...
8 blob_sas_token = token_library.getConnectionString("Azur...
9
10 spark.conf.set(
11     'fs.azure.sas.%s.%s.blob.core.windows.net' % (blob_c...
12     blob_sas_token)
13 df = spark.read.load('wasbs://sourceblob1@dxcstorageacc...
14 ## If header_exists uncomment line below
15 ##, header=True
16 )
17 display(df.limit(10))
```

3 sec - Command executed in 2 sec 827 ms by dxc26ab1215_165453000

View Table Chart Export results

c0	c1	c2
raceld	year	rou...
1	2009	1

New linked service

Azure Cosmos DB (SQL API) Learn more

AutoResolveIntegrationRuntime

Authentication type

Account key

Connection string Azure Key Vault

Account selection method

From Azure subscription Enter manually

Azure subscription

Azure-DXC262AB12Lab (414da265-1fd8-4b63-919d-307787722583)

Azure Cosmos DB account name * dxccosmosdb123

Database name * ToDoList

Additional connection properties

+ New

Annotations

+ New

> Parameters

Connection successful

Test connection Cancel

Create Back

Synapse live Validate all Publish all (1)

Data Workspace Linked

Filter resources by name

Not started

A Please select a Spark pool to attach before running cell!

+ Code + Markdown

```
1 # Load a streaming Spark DataFrame from a Cosmos DB container
2 # To select a preferred list of regions in a multi-region Cosmos DB account, add .option("spark
3
4 # For Spark 2.4
5 dfStream = spark.readStream\
6     .format("cosmos.oltp")\
7     .option("spark.synapse.linkedService", "CosmosDb1")\
8     .option("spark.cosmos.container", "Items")\
9     .option("spark.cosmos.changeFeed.readEnabled", "true")\
10    .option("spark.cosmos.changeFeed.startFromTheBeginning", "true")\
11    .option("spark.cosmos.changeFeed.checkpointLocation", "/localReadCheckpointFolder")\
12    .option("spark.cosmos.changeFeed.queryName", "streamQuery")\
13    .load()
14
15 # For Spark 3.1 +
16 # dfStream = spark.readStream\
17 #     .format("cosmos.oltp.changeFeed")\
18 #     .option("spark.synapse.linkedService", "CosmosDb1")\
19 #     .option("spark.cosmos.container", "Items")\
20 #     .option("spark.cosmos.changeFeed.startFrom", "Beginning")\
21 #     .option("spark.cosmos.changeFeed.mode", "Incremental")\
22 #     .load()
23
```

Press shift + enter to run

Properties

General Related (0)

Name * Notebook 1

Description

Type .ipynb notebook

Size 1,999 bytes

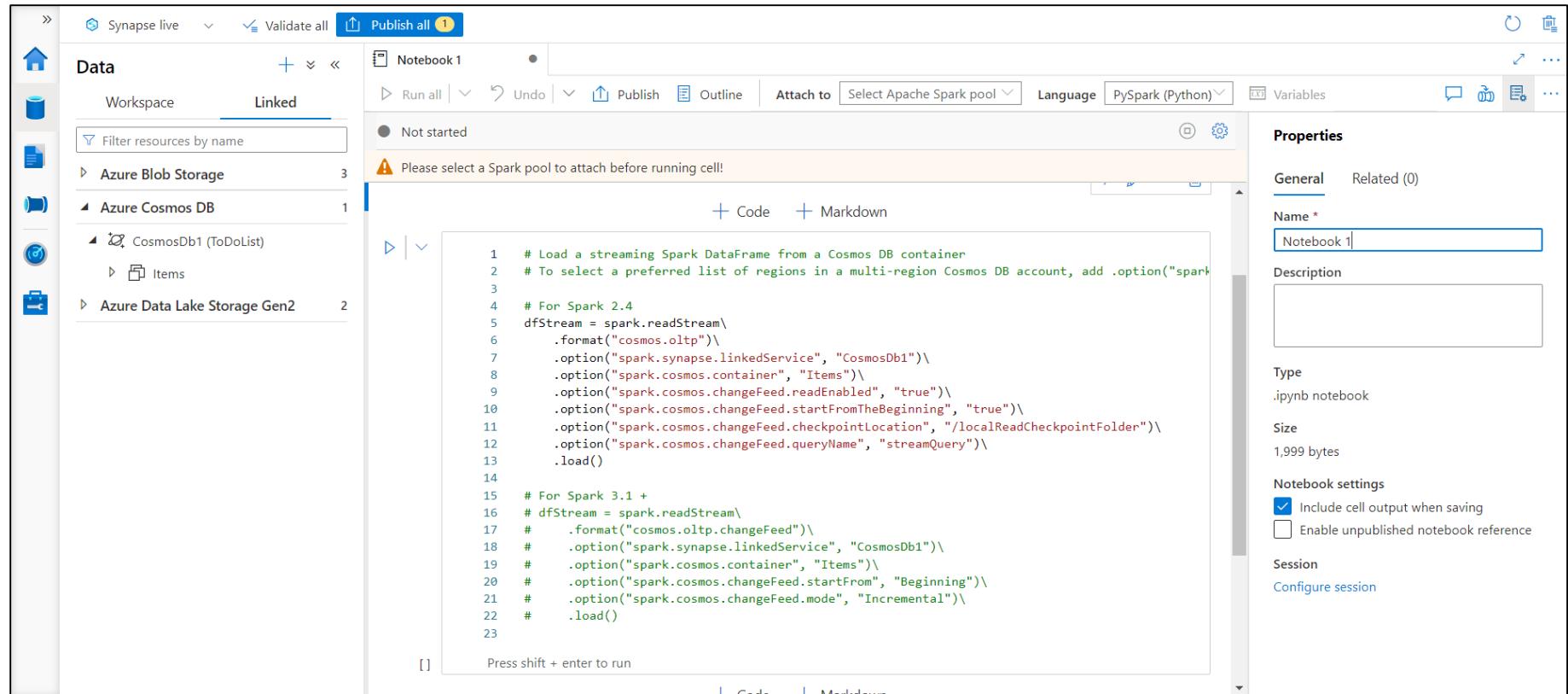
Notebook settings

Include cell output when saving

Enable unpublished notebook reference

Session

Configure session



8.Create azure Data factory & azure Blob, connect Blob & ADF,

import blob files into Data factory & explain the steps with screenshots

Home >

Data factories

Manipal 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 ↑

No data factories to display

Try changing or clearing your filters.

Create data factory Learn more ↗

Give feedback



Create Data Factory

X

[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 * ⓘ

Azure-DXC262AB12Lab

Resource group * ⓘ

dxcrg123

[Create new](#)

Instance details

Name * ⓘ

DXCdatafactory12345

Region * ⓘ

East US

Version * ⓘ

V2 (Recommended)

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

Create Data Factory

Basics **Git configuration** Networking Advanced Tags Review + create

Azure Data Factory allows you to configure a Git repository with either Azure DevOps or GitHub. Git is a version control system that allows for easier change tracking and collaboration.

[Learn more about Git integration in Azure Data Factory](#)

Configure Git later ⓘ



Microsoft Azure Search resources, services, and docs (G+)

Home > Data factories >

Create Data Factory

Validation Passed

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	dxcrg123
Name	DXCdatafactory12345
Region	East US
Version	V2 (Recommended)

Networking

Connect via	Public endpoint
-------------	-----------------

Actions

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

Home > Microsoft.DataFactory-20220617164623 | Overview X

 Microsoft.DataFactory-20220617164623 | Deployment

« Delete Cancel Redeploy Refresh

 We'd love your feedback! →

 Your deployment is complete

Deployment name: Microsoft.DataFactory-20220617164623
Subscription: Azure-DXC262AB12Lab
Resource group: dxcrg123

Start time: 6/17/2022, 4:47:30 PM
Correlation ID: 3e007683-8188-4aa5-ac61-d119d23dd309 Download

Deployment details Download

Next steps

Go to resource

 Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
[Set up cost alerts >](#)

 Microsoft Defender for Cloud
Secure your apps and infrastructure
[Go to Microsoft Defender for Cloud >](#)

 Free Microsoft tutorials
[Start learning today >](#)

 Work with an expert
Azure experts are service provider partners who can help manage your assets on Azure and be your first line of support.
[Find an Azure expert >](#)

Home / Microsoft Data Factory / DXCdatafactory12345 /

DXCdatafactory12345

Data factory (V2)

Search (Ctrl+ /) Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Essentials

Resource group (move) : [dxcrg123](#)

Status : Succeeded

Type : Data factory (V2)

Location : East US

Getting started : [Quick start](#)

Subscription (move) : [Azure-DXC262AB12Lab](#)

Subscription ID : 414da265-1fd8-4b63-919d-307787722583

JSON View

Settings

Networking

Managed identities

Properties

Locks

Getting started

Quick start

Monitoring

Alerts

Metrics

Diagnostic settings

Logs

Getting started

Open Azure Data Factory Studio

Start authoring and monitoring your data pipelines and data flows.

Open 

Read documentation

Learn how to be productive quickly. Explore concepts, tutorials, and samples.

Learn more 

Monitoring

PipelineRuns

100

80

ActivityRuns

100

80

60

<https://adf.azure.com/en/home?factory=%2Fsubscriptions%2F414da265-1fd8-4b63-919d-30...>

[Set up code repository](#)

Data factory
DXCdatafactory12345

[New](#)

 **Ingest**
Copy data at scale once or on a schedule.

 **Orchestrate**
Code-free data pipelines.

 **Transform data**
Transform your data using data flows.

 **Configure SSIS**
Manage & run your SSIS packages in the cloud.

Discover more

 [Browse partners \(preview\)](#)

 [Pipeline templates](#)

 [SAP pipeline templates](#)

« Data Factory Validate all Publish all

Home

Author

Monitor

Manage

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries (preview)

Security

Credentials

Customer managed key

Managed private endpoints

Linked services

Linked service defines the connection information to a data store or compute. [Learn more](#)

+ New

Filter by name Annotations : Any

Showing 0 - 0 of 0 items

Name ↑ Type ↑ Related ↑ Annotations ↑

No linked service to show

If you expected to see results, try changing your filters or create a new linked services.

Create linked service



Microsoft Azure | DXCdatafactory12345

Validate all Publish all

Home Author Monitor Manage

Connections Linked services Integration runtimes Microsoft Purview

Source control Git configuration ARM template

Author Triggers Global parameters Data flow libraries (preview)

Security Credentials Customer managed key Managed private endpoints

Linked services

Linked service defines the connection information to a data store or compute

+ New

Filter by name Annotations : Any

Showing 0 - 0 of 0 items

Name ↑ Type ↑

New linked service

Data store Compute

Search

All Azure Database File Generic protocol NoSQL Services and apps

		
Amazon Redshift	Amazon S3	Amazon S3 Compatible
		
Apache Impala	Asana (Preview)	Azure Blob Storage
		
Continue	Cancel	

Home

Author

Monitor

Manage

Data Factory

Validate all

Publish all

Connections

Linked services

Integration runtimes

Microsoft Purview

Source control

Git configuration

ARM template

Author

Triggers

Global parameters

Data flow libraries (preview)

Security

Credentials

Customer managed key

Managed private endpoints

Linked services

Linked service defines the connection information to a data store or compute service.

New

Filter by name

Annotations : Any

Showing 0 - 0 of 0 items

Name ↑ Type ↑

If you expected to see result

New linked service

Azure Blob Storage [Learn more](#)

Name * AzureBlobStorage1

Description

Connect via integration runtime * [①](#) AutoResolveIntegrationRuntime

Authentication type Account key

Connection string Azure Key Vault

Account selection method [①](#)

From Azure subscription Enter manually

Azure subscription [①](#) Azure-DXC262AB12Lab (414da265-1fd8-4b63-919d-307787722583)

Storage account name * dxcstorageaccount12345

Additional connection properties

New

Connection successful [①](#)

Test connection Cancel

Create Back

RESULT

Almost all the test cases have been solved and presented successfully in the present document.

CONCLUSIONS

All the case studies have been solved successfully with all the concepts that have been covered in the training session. It's really a great experience of learning while solving the cases. This case study gave me immense confidence regarding my ability to upskill in new technologies.