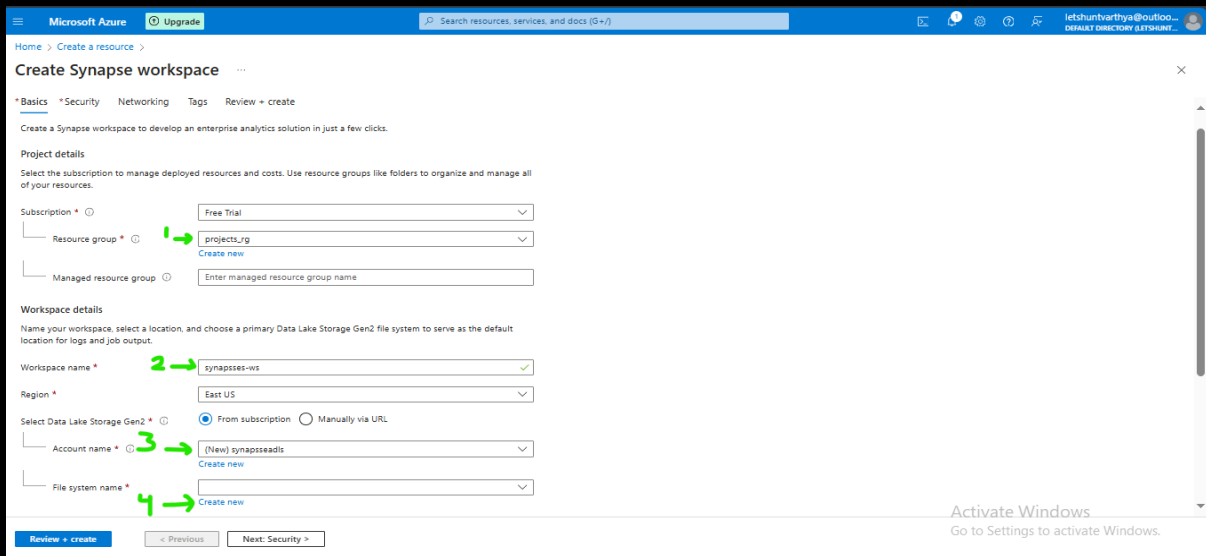
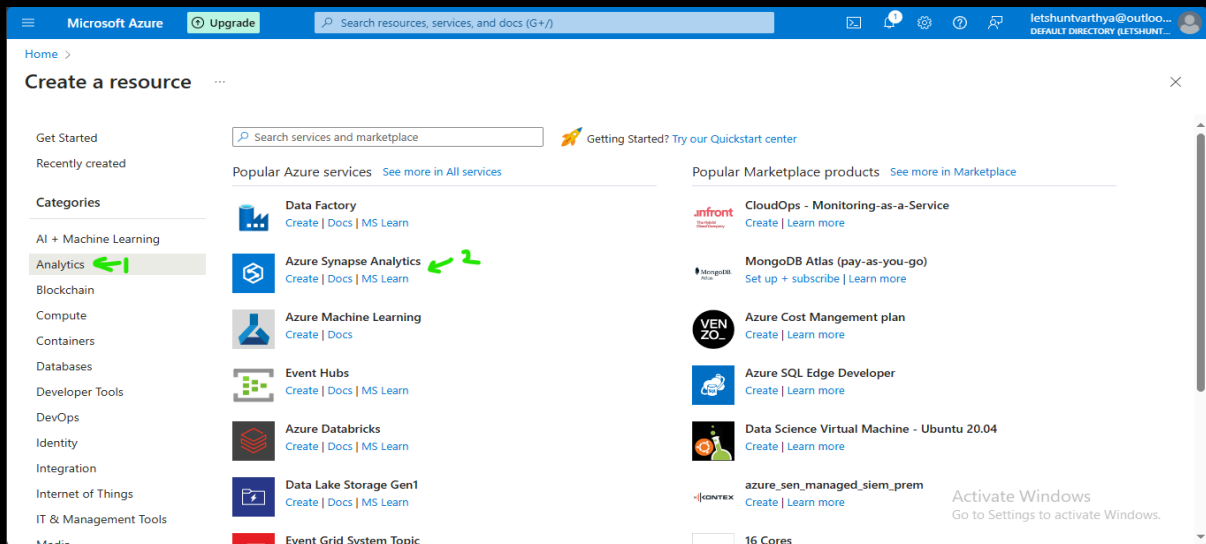
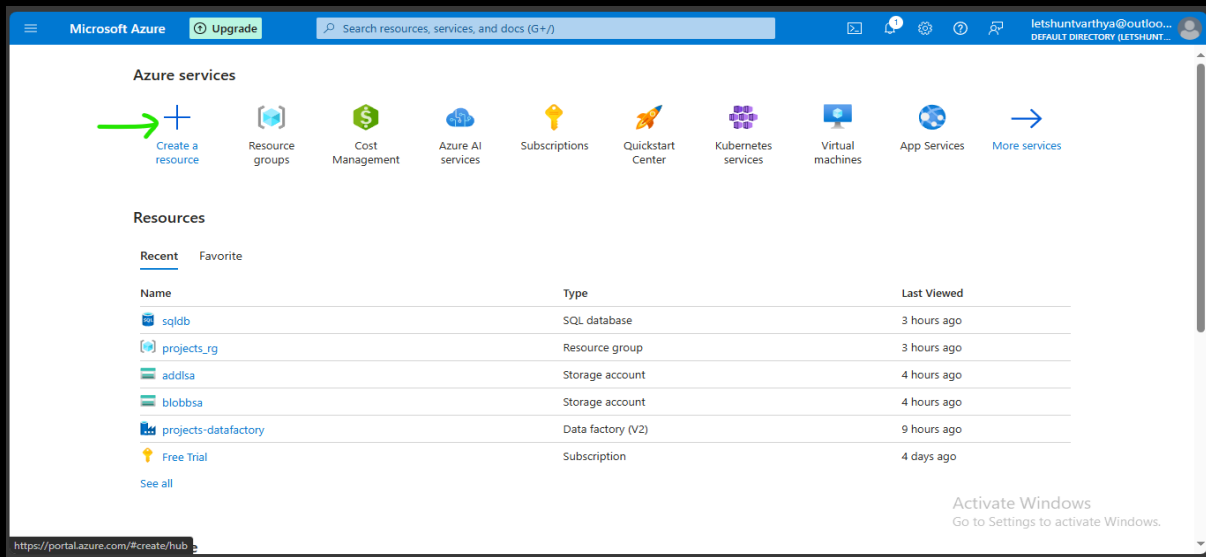


- Pipeline → adls2sqlserver using azure synapse analytics(SparkPool)

## Azure Synapse Creation:



Microsoft Azure

Upgrade

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

letshuntvarthya@outloo...  
DEFAULT DIRECTORY (LETSHUNT...)

Home > Create a resource >

## Create Synapse workspace

\* Basics \* Security Networking Tags Review + create

Create a Synapse workspace to develop an enterprise analytics solution in just a few clicks.

**Project details**

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

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](#)

Review + create

< Previous

Next: Security >

Activate Windows  
Go to Settings to activate Windows.

Microsoft Azure

Upgrade

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

letshuntvarthya@outloo...  
DEFAULT DIRECTORY (LETSHUNT...)

Home > Create a resource >

## Create Synapse workspace

\* Basics \* Security Networking Tags Review + create

Configure security options for your 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 ☒ Use both local and Microsoft Entra ID authentication ☐ Use only Microsoft Entra ID authentication

SQL Server admin login \*

SQL Password  ✓

Confirm password  ✓

**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.

**Workspace encryption**

Review + create

< Previous

Next: Networking >

Activate Windows  
Go to Settings to activate Windows.

Microsoft Azure

Upgrade

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

letshuntvarthya@outloo...  
DEFAULT DIRECTORY (LETSHUNT...)

Home > Create a resource >

## Create Synapse workspace

\* Basics \* Security Networking Tags Review + create

Configure networking options for your workspace.

**Managed virtual network**

Choose whether to set up a dedicated Azure Synapse-managed virtual network for your workspace. [Learn more](#)

Managed virtual network: ☐ Enable ☒ Disable

**Firewall rules**

Azure Synapse Studio and other client tools will only connect to the workspace endpoints if this setting is selected. Connections from specific IP addresses or all Azure services can be allowed or disallowed after the workspace is provisioned.

Allow connections from all IP addresses to your workspace's endpoints. You can restrict these permissions to just Azure datacenter IP addresses and/or specific IP address ranges after creating the workspace.

☒ Allow connections from all IP addresses

**Encrypted connections**

Review + create

< Previous

Next: Tags >

Activate Windows  
Go to Settings to activate Windows.

Microsoft Azure

Upgrade

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

letshuntvarthya@outloo...  
DEFAULT DIRECTORY (LETSHUNT...)

Home > Create a resource >

Create Synapse workspace

\* Basics \* Security Networking Tags Review + create

Tags are name/value pairs that enable you to categorize resources and view consolidated billing by applying the same tag to multiple resources and resource groups. [learn more](#)

Note that if you create tags and then change resource settings on other tabs, your tags will be automatically updated.

Name ⓘ

Value ⓘ

Resource

:

Synapse workspace

Review + create

< Previous

Next: Review + create >

Activate Windows  
Go to Settings to activate Windows.

Microsoft Azure

Upgrade

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

letshuntvarthya@outloo...  
DEFAULT DIRECTORY (LETSHUNT...)

Home > Create a resource >

Create Synapse workspace

✓ Validation succeeded

\* Basics \* Security Networking Tags Review + create

Product Details

Azure Synapse Analytics workspace by Microsoft  
[Terms of use](#) | [Privacy policy](#)

Serverless SQL est. cost/TB ⓘ  
415.97 INR

Terms

By clicking Create, I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above, (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) 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. For additional details see [Azure Marketplace Terms](#).

Basics

SubscriptionFree Trial

Resource groupprojects\_rg

RegionEast US

Create

< Previous

Next >

Download a template for automation

Activate Windows  
Go to Settings to activate Windows.

Microsoft Azure

Upgrade

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

letshuntvarthya@outloo...  
DEFAULT DIRECTORY (LETSHUNT...)

Home >

Microsoft.Azure.SynapseAnalytics-20240528214258 | Overview

Deployment

Search

Delete

Cancel

Redeploy

Download

Refresh

Overview

Inputs

Outputs

Template

✓ Your deployment is complete

Deployment name : Microsoft.Azure.SynapseAnalytics-20240528214258Start time : 5/28/2024, 10:03:44 PM

Subscription : Free TrialCorrelation ID : fdfea5b3-02e8-4159-953d-0de47167c348

Resource group : projects\_rg

Deployment details

Next steps

Go to resource group

Give feedback

Tell us about your experience with deployment

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](#)

Activate Windows  
Go to Settings to activate Windows.

Home > Microsoft.Azure.SynapseAnalytics-20240528214258 | Overview > projects\_rg > synapseseadls

### synapseseadls | Containers

Storage account

Search

+ Container + Change access level + Restore containers + Refresh + Delete + Give feedback

Search containers by prefix

Show deleted containers

Name	Last modified	Anonymous access level	Lease state
container	5/28/2024, 10:09:56 PM	Private	Available
datasets	5/28/2024, 10:04:14 PM	Private	Available

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

Home > Microsoft.Azure.SynapseAnalytics-20240528214258 | Overview > projects\_rg > synapseseadls | Containers >

### datasets

Container

Search

+ Upload + Add Directory + Refresh + Rename + Delete + Change tier + Acquire

Authentication method: Access key (Switch to Microsoft Entra user account)

Location: datasets

Search blobs by prefix (case-sensitive)

Upload blob

1 file(s) selected: 2666141-Bike\_Data (1).csv

Drag and drop files here or Browse for files

Overwrite if files already exist

Advanced

Upload

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

Home > Microsoft.Azure.SynapseAnalytics-20240528214258 | Overview > projects\_rg >

### synapsesws

Synapse workspace

Search

+ New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) + Refresh + Reset SQL admin password + Delete

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Microsoft Entra ID

Properties

Locks

Analytics pools

SQL pools

Apache Spark pools

Data Explorer pools (preview)

Security

Encryption

Networking

Identity

Essentials

Resource group (move) : projects\_rg

Status : Succeeded

Location : East US

Subscription (move) : Free Trial

Subscription ID : c92e5287-784e-4a20-9e6c-8549947f2ee6

Managed virtual network : No

Managed Identity object ... : ef682be3-454e-4f06-a2fb-42081af16316

Workspace web URL : https://web.azure.synapse.net/workspace=%2fsubscriptions%2f...

Tags (edit) : Add tags

Networking : Show firewall settings

Primary ADLS Gen2 acco... : https://synapseseadls.dfs.core.windows.net

Primary ADLS Gen2 file s... : datasets

SQL admin username : sqladminuser

SQL Microsoft Entra admin : live.com#letshuntvarthya@outlook.com

Dedicated SQL endpoint : synapsesws-sql.azure.synapse.net

Serverless SQL endpoint : synapsesws-ondemand.sql.azure.synapse.net

Development endpoint : https://synapsesws-dev.azure.synapse.net

Getting started

Open Synapse Studio

Start building your fully-integrated analytics solution and unlock new insights.

Open

Read documentation

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

Learn more

Analytics pools

Search to filter items...

Activate Windows

Go to Settings to activate Windows.

Microsoft Azure Synapse Analytics synapsesws

We use optional cookies to provide a better experience. Learn more

Accept Reject More options

Synapse live Validate all Publish all

Analytics pools

SQL pools

Apache Spark pools

Data Explorer pools (prev...

External connections

Linked services

Microsoft Purview

Integration

Triggers

SQL pools

The serverless SQL pool, Built-in, is immediately available for your workspace. Dedicated SQL pools can be configured to adapt to team or organizational requirements and constraints. Learn more

+ New + Refresh

Filter by name

Showing 1-1 of 1 items (1 Serverless, 0 Dedicated)

Name	Type	Status	Size
Built-in	Serverless	Online	Auto

Microsoft Azure | Synapse Analytics | synapses-ws

We use optional cookies to provide a better experience

### New dedicated SQL pool

Basics \* Additional settings \* Tags Review + create

Create a dedicated SQL pool with your preferred configurations. Complete the Basics tab then go to Review + Create to provision with smart defaults. [Learn more](#)

**Dedicated SQL pool details**

Name your dedicated SQL pool and choose its initial settings.

Dedicated SQL pool name \*

Performance level  **DW300c**

Estimated price  **Est. cost per hour 376.87 INR**  
[View pricing details](#)

[Review + create](#) [Next: Additional settings >](#)

Activate Windows  
Go to Settings to activate Windows. [Cancel](#)

Microsoft Azure | Synapse Analytics | synapses-ws

We use optional cookies to provide a better experience

### New dedicated SQL pool

Basics \* Additional settings \* Tags **Review + create**

**Validation succeeded.**

**Product details**

Azure Synapse Analytics dedicated SQL pool by Microsoft  
[Terms of use](#) | [Privacy policy](#) **Est. cost per hour 376.87 INR**  
[View pricing details](#)

**Terms**

By clicking "Create", I (a) agree to the legal terms and privacy statement(s) associated with the Marketplace offering(s) listed above; (b) authorize Microsoft to bill my current payment method for the fees associated with the offering(s), with the same billing frequency as my Azure subscription; and (c) 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. For additional details see [Azure Marketplace Terms](#)

**Data source**

Dedicated SQL pool name dedicatedsql  
Performance level DW300c

**Additional settings**

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

Activate Windows  
Go to Settings to activate Windows. [Cancel](#)

Microsoft Azure | Synapse Analytics | synapses-ws


We use optional cookies to provide a better experience

### 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](#)

Showing 0-0 of 0 item

Name	Node size family	Size
<div> <b>No items to show</b> Try changing your filter or create new Apache Spark pool <a href="#">New Apache Spark pool</a></div>		

Activate Windows  
Go to Settings to activate Windows.

**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 \* 1 → sparkws

Isolated compute \* ☐ Enabled ☒ Disabled

Node size family \* Memory Optimized

Node size \* 2 → Small (4 vCores / 32 GB)

Autoscale \* ☒ Enabled ☐ Disabled

Number of nodes \* 4 → 3 8 3

Estimated price ☐ Est. cost per hour 137.77 to 367.39 INR [View pricing details](#)

Dynamically allocate executors \* ☐ Enabled ☒ Disabled

5 → Review + create Next: Additional settings >

Activate Windows  
Go to Settings to activate Windows.

Click create

**Data**

Workspace Linked

Filter resources by name

1 → Azure Data Lake Storage Gen2 2

3 → synapses-ws (Primary - synapse...

4 → datasets (Primary)

5 → container

6 → (Attached Containers)

Database1 • datasets

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

New SQL script New notebook New data flow New integration dataset Upload More

← → datasets

Name	Last Modified	Content Type	Size
synapse	28/5/2024, 10:21:06 pm	Folder	
2666141-Bike_Data (1).csv	28/5/2024, 10:12:59 pm		6.6 MB

Preview

7 → Select TOP 100 rows

Create external table

Bulk load

Showing 1 to 2 of 2 cached item

Activate Windows  
Go to Settings to activate Windows.

Database1 • datasets

SQL script 1

Other users in your workspace may have access to modify this item. Do not use this item unless you trust all users who may have access to the workspace.

Run Undo Publish Query plan Connect to Built-in Use database master

```

1 -- This is auto-generated code
2 SELECT
3     TOP 100 *
4 FROM
5     OPENROWSET(
6         BULK 'https://synapseadls.dfs.core.windows.net/datasets/2666141-Bike_Data (1).csv',
7         FORMAT = 'CSV',
8         PARSE_VERSION = '2.0'
9     ) AS [result]

```

Results Messages

View Table Chart Export results

C1	C2	C3	C4	C5	C6	C7	C8
Region	Country	Customer	Business Segm...	Category	Model	Color	SalesDate
North America	United States	Advanced Bike ...	Components	Road Frames	LL Road Frame	Red	4/1/2020
North America	United States	Central Discou...	Bikes	Mountain Bikes	Mountain-100	Silver	4/1/2020
North America	United States	Leisure Sales	Clothing	Leisure	Leisure-Sleeve L...	Multi	4/1/2020

00:00:15 Query executed successfully.

**Properties**

General Related (0)

Name \* SQL script 1

Description

Type .sql script

Size 241 bytes

Results settings per query ☒ First 5000 rows (default)

☐ All rows

Activate Windows  
Go to Settings to activate Windows.

- We can see the data in the spark pool as well.

The screenshot shows the Microsoft Azure Synapse Analytics interface. On the left, the 'Data' pane shows a workspace with a linked 'dedicatedsql (SQL)' database. A context menu is open over the 'dbo.bikedata\_tb' table, showing options like 'New SQL script', 'New notebook', 'New data flow', 'New integration dataset', and 'Refresh'. The main pane displays a SQL query: `SELECT TOP (100) [Region], [Country], [Customer], [BusinessSegment], [Category], [Model], [Color], [SalesDate], [ListPrice], [UnitPrice], [OrderQty]`. Below the query, the results are shown in a table format.

Region	Country	Customer	BusinessSegment	Category	Model	Color	SalesDate	ListPrice	UnitPrice	OrderQty
North America	United States	Valuable Bike P...	(NULL)	Pumps	Minipump	(NULL)	2017-10-13	19.99		1
North America	Canada	Valley Toy Store	(NULL)	Mountain Bikes	Mountain-200	Silver	2019-10-13	2319.99		1
North America	United States	Small Bike Shop	(NULL)	Bib-Shorts	Men's Bib-Shorts	Multi	2017-12-07	89.99		5

At the bottom, a status bar indicates: '00:00:02 Query executed successfully.'

- Let's do the transformation using spark pool

The screenshot shows a Jupyter Notebook in the Azure Synapse Analytics environment. The notebook is titled 'Notebook 1' and is attached to a 'sparkws' workspace. The code in the notebook is as follows:

```
1 %%spark
2 df = spark.read.load('abfss://rawdata@synapseadl1sa.dfs.core.windows.net/2666141-Bike_Data (1).csv', format='csv'
3 ## If header exists uncomment line below
4 ,header=True,inferSchema=True
5 )
6 display(df.limit(10))
```

The execution output shows 'Job execution Succeeded' and 'Spark 2 executors 8 cores'. Below this, the results are displayed in a table format.

Region	Country	Customer	Business Segment	Category	Model	Color
North America	United States	Advanced Bike Components	Components	Road Frames	LL Road Frame	Red
North America	United States	Central Discount Store	Bikes	Mountain Bikes	Mountain-100	Silver
North America	United States	Leading Sales & Repair	Clothing	Jerseys	Long-Sleeve Logo Jersey	Multi
North America	United States	Paint Supply	Components	Mountain Frames	HL Mountain Frame	Black
North America	United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-450	Red
North America	United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-650	Black
North America	United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-650	Red
North America	United States	Modern Bike Store	Bikes	Road Bikes	Road-450	Red
North America	United States	Modern Bike Store	Bikes	Road Bikes	Road-650	Red
North America	Canada	Corner Bicycle Supply	Accessories	Helmets	Sport-100	Black

The screenshot shows a Jupyter Notebook in the Azure Synapse Analytics environment, continuing from the previous one. The code in the notebook is as follows:

```
1 from pyspark.sql.functions import *
2 from pyspark.sql.types import *
3 from pyspark.sql import *

1 data = df.select('Country', 'Customer', 'Business Segment', 'Category', 'Model', 'SalesDate', 'ListPrice', 'UnitPrice', 'OrderQty')

1 data = data.withColumn("Sales", col("ListPrice") * col("OrderQty"))

1 data = data.withColumn("Cost", col("UnitPrice") * col("OrderQty"))
```

The execution output shows 'Job execution Succeeded' and 'Spark 2 executors 8 cores'. Below this, the results are displayed in a table format.

Country	Customer	Business Segment	Category	Model	SalesDate	ListPrice	UnitPrice	OrderQty	Sales	Cost
United States	Advanced Bike Components	Components	Road Frames	LL Road Frame	2017-10-13	19.99		1	19.99	
United States	Central Discount Store	Bikes	Mountain Bikes	Mountain-100	2019-10-13	2319.99		1	2319.99	
United States	Leading Sales & Repair	Clothing	Jerseys	Long-Sleeve Logo Jersey	2017-12-07	89.99		5	449.95	
United States	Paint Supply	Components	Mountain Frames	HL Mountain Frame	2017-12-07	89.99		5	449.95	
United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-450	2017-12-07	89.99		5	449.95	
United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-650	2017-12-07	89.99		5	449.95	
United States	Scooters and Bikes Store	Bikes	Road Bikes	Road-650	2017-12-07	89.99		5	449.95	
United States	Modern Bike Store	Bikes	Road Bikes	Road-450	2017-12-07	89.99		5	449.95	
United States	Modern Bike Store	Bikes	Road Bikes	Road-650	2017-12-07	89.99		5	449.95	
Canada	Corner Bicycle Supply	Accessories	Helmets	Sport-100	2017-12-07	89.99		5	449.95	

```
1 data = data.withColumn("Profit",col("Sales")-col("OrderQty"))
[57] ✓ <1 sec - Command executed in 157 ms by letshuntvarthya on 3:53:50 AM, 5/29/24

1 # database configuration
2 server="synapse-ws.sql.azuresynapse.net"
3 port = "1433"
4 database = "dedicatedsql"
5 db_properties = {"user":"sqladminuser","password":"Azure@123"}
[58] ✓ <1 sec - Command executed in 152 ms by letshuntvarthya on 3:58:37 AM, 5/29/24

1 url = 'jdbc:sqladminuser://synapse-ws.sql.azuresynapse.net:1433;dedicatedsql'
2 # url = 'jdbc:sqladminuser://{0};{1};database={2}'.format(server,port,database)'
[59] ✓ <1 sec - Command executed in 149 ms by letshuntvarthya on 4:09:07 AM, 5/29/24

...

1 output = DataFeameworkriter(data)
[60] ✓ <1 sec - Command executed in 157 ms by letshuntvarthya on 4:11:44 AM, 5/29/24
```

```
+ Code + Markdown

1 output.jdbc(url = url,table="bikedata_tb",mode="Overwrite",properties=db_properties)
```

Microsoft Azure | Synapse Analytics | synapses-ws

Synapse live | Validate all | Publish all

Data | Workspace | Linked

Filter resources by name

- Lake database 2
- SQL database 2
  - dedicatedsql (SQL)
    - Tables
      - dbo.bikedata\_tb

Run | Undo | Publish | Query plan | Connect to | Use database

```
1 SELECT TOP (100) [Region]
2 ,[Country]
3 ,[Customer]
4 ,[BusinessSegment]
5 ,[Category]
6 ,[Model]
7 ,[Color]
8 ,[SalesDate]
9 ,[ListPrice]
10 ,[UnitPrice]
11 ,[OrderQty]
```

00:00:02 Query executed successfully.

Region	Country	Customer	BusinessSegm...	Category	Model	Color	SalesDate	ListPrice	U
North America	United States	Valuable Bike P...	(NULL)	Pumps	Minipump	(NULL)	2017-10-13	19.99	1
North America	Canada	Valley Toy Store	(NULL)	Mountain Bikes	Mountain-200	Silver	2019-10-13	2319.99	1.
North America	United States	Small Bike Shop	(NULL)	Bib-Shorts	Men's Bib-Shorts	Multi	2017-12-07	89.99	5