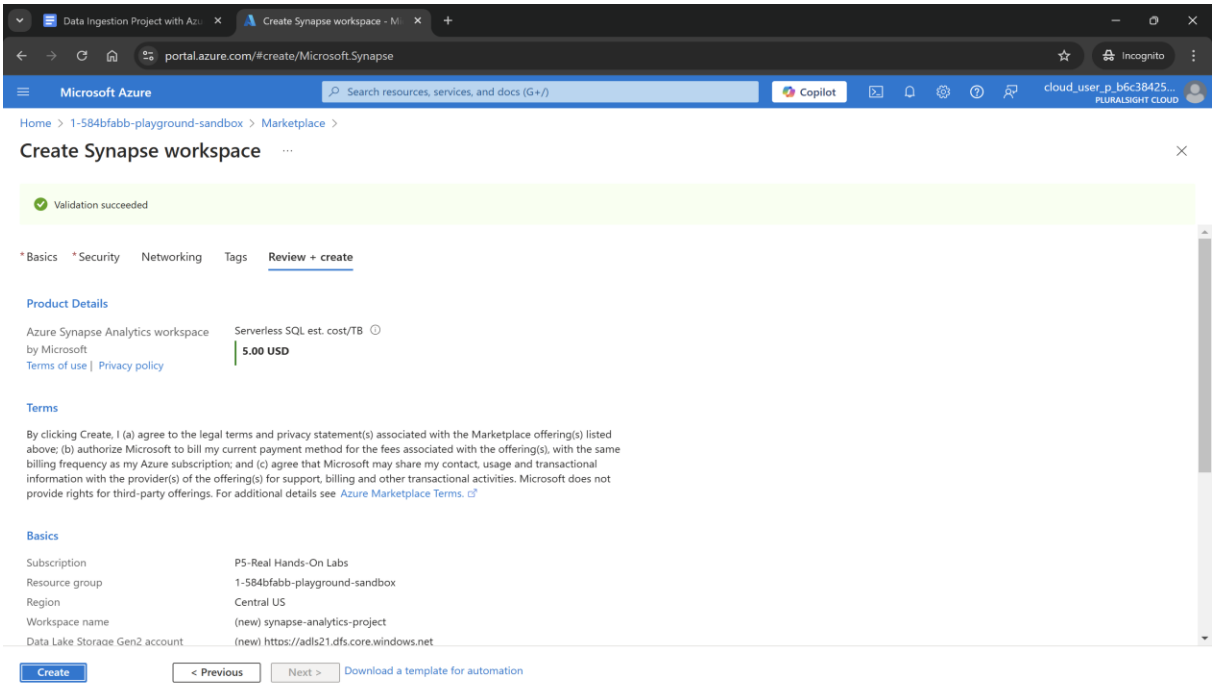
Data Ingestion using Azure Synapse Analytics

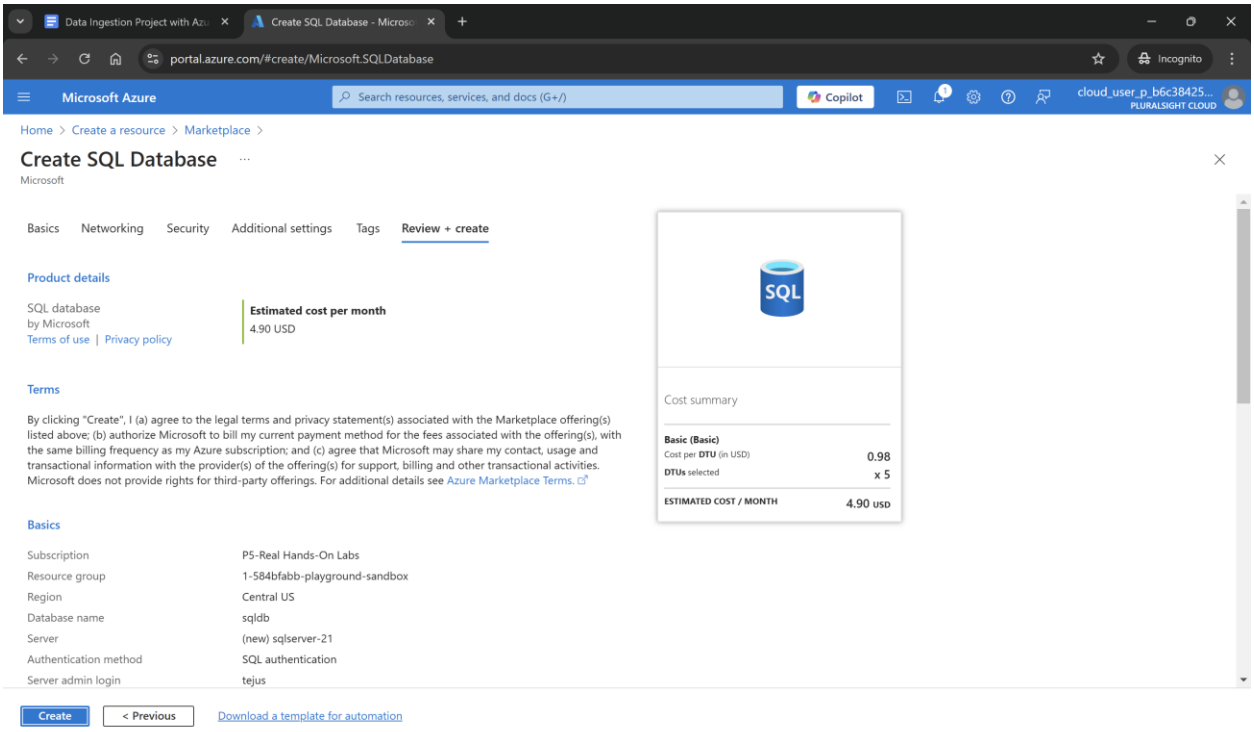
Steps Involved:

1. **Create a Synapse Resource**
2. **Create a SQL Server, Database**
3. **Create an Azure Key Vault, and store secret**
4. **Link HTTP Server, SQL Database, Key Vault, ADLS2 resources**
5. **Create appropriate Source, Target integration datasets with dynamic parameters**
6. **Create a pipeline with appropriate Source, Sink with dynamic parameters**
7. **Create SQL script on serverless/dedicated SQL pool to create External source format, External data source, External table for ingested dataset**
8. **Analyze the table with queries**

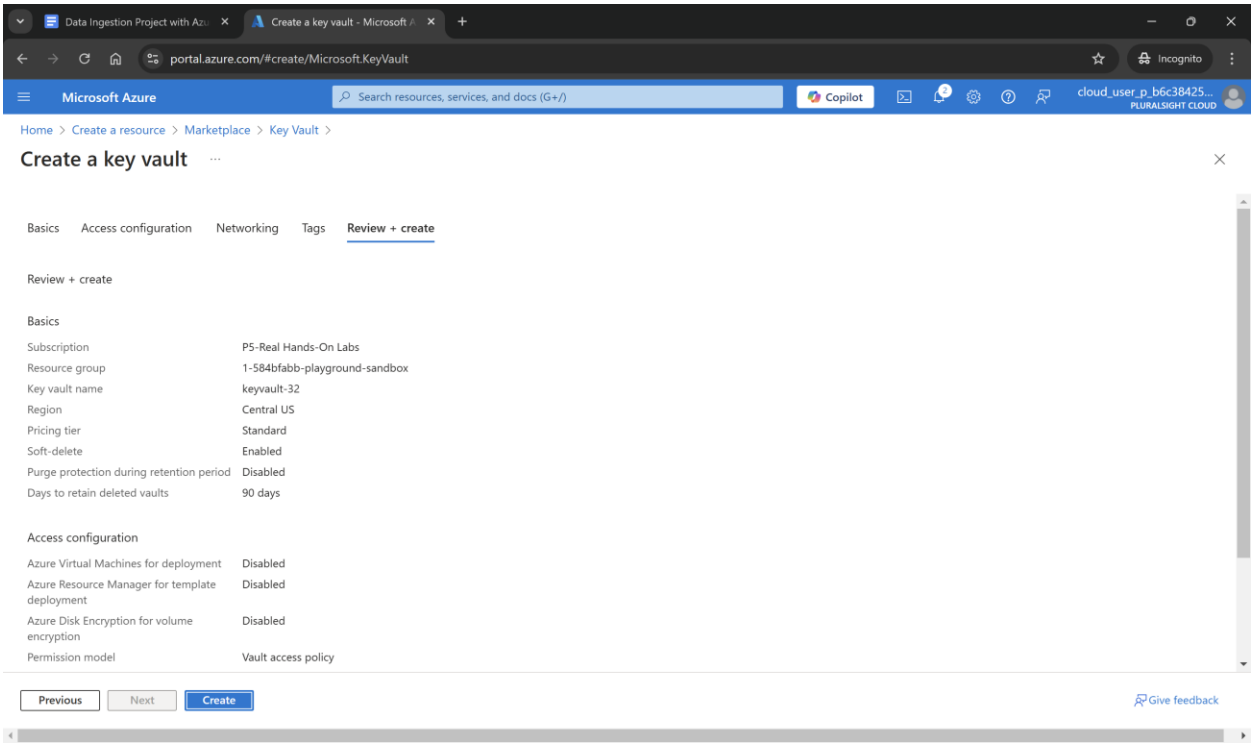
Step 1.

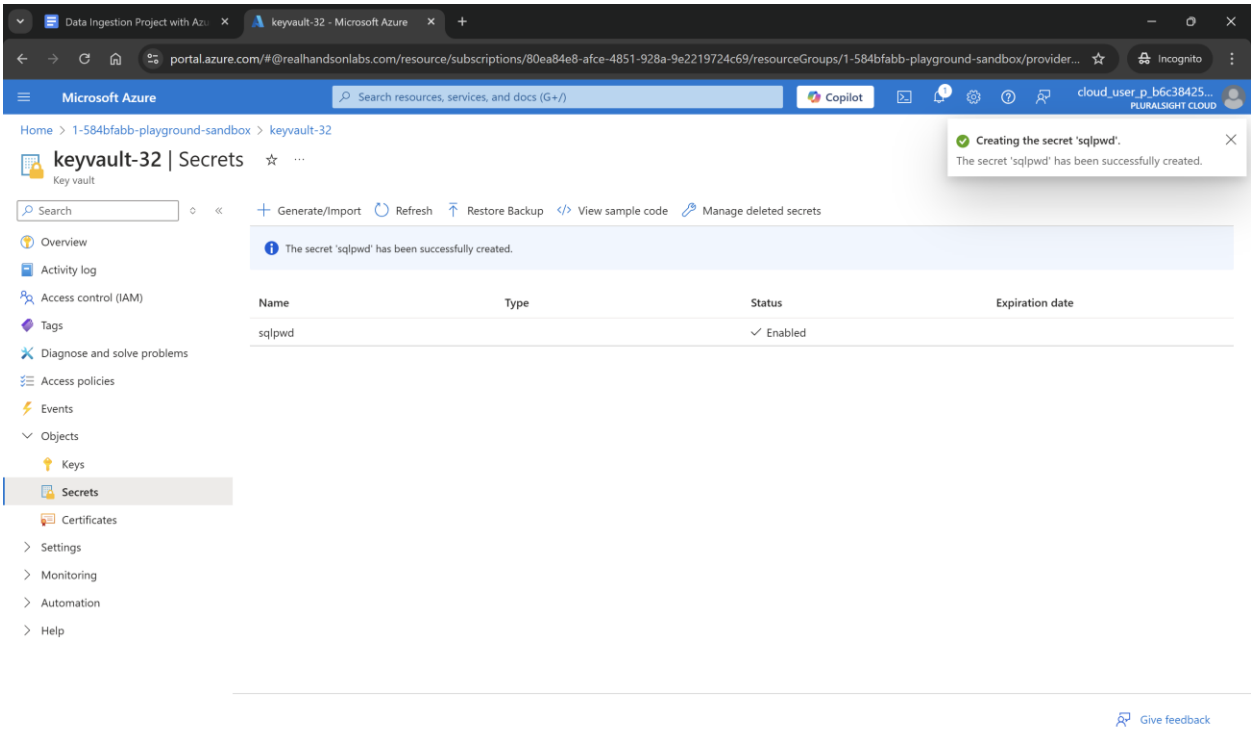


Step 2.

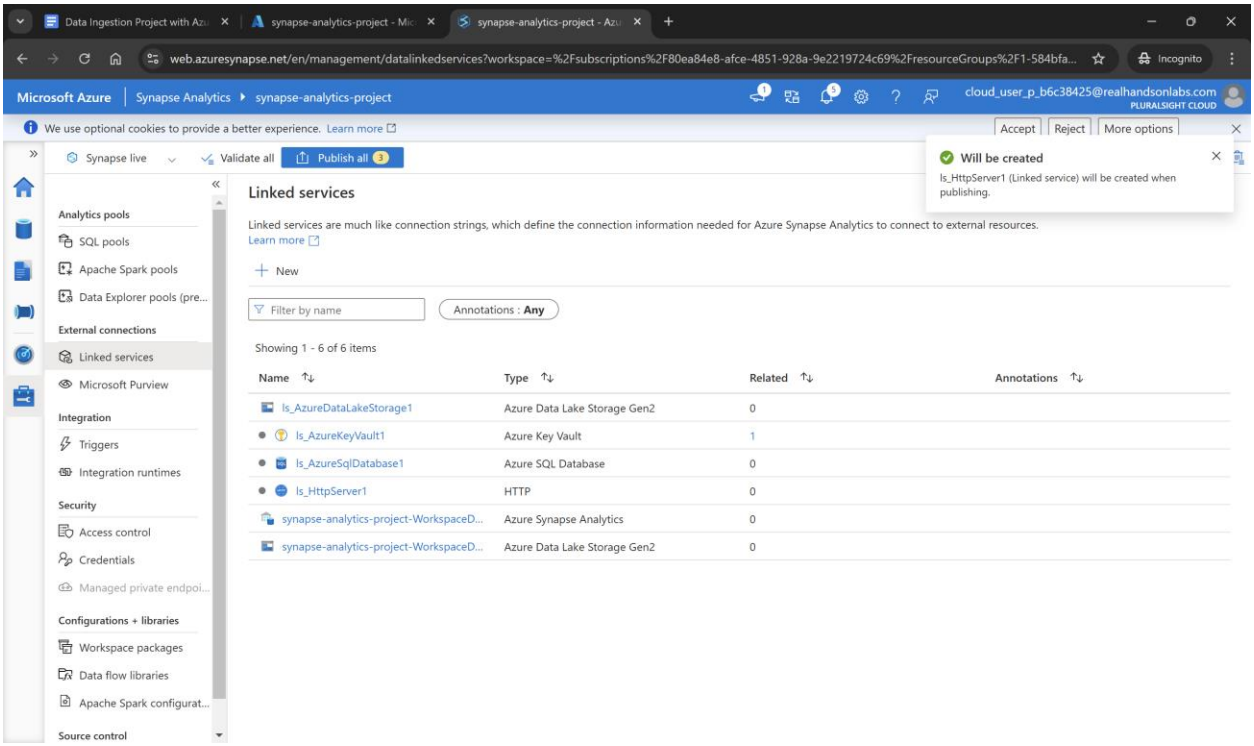

Data Ingestion Project with Azu
Create SQL Database - Microso
portal.azure.com/#create/Microsoft.SQLDatabase
Incognito
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
cloud_user_p_b6c38425...
PLURALSIGHT CLOUD
Home > Create a resource > Marketplace >
Create SQL Database
Microsoft
Basics Networking Security Additional settings Tags Review + create
Product details
SQL database
by Microsoft
Terms of use | Privacy policy
Estimated cost per month
4.90 USD
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
Subscription P5-Real Hands-On Labs
Resource group 1-584bfabb-playground-sandbox
Region Central US
Database name sqldb
Server (new) sqlserver-21
Authentication method SQL authentication
Server admin login tejus
SQL
Cost summary
Basic (Basic)
Cost per DTU (in USD) 0.98
DTUs selected x 5
ESTIMATED COST / MONTH 4.90 USD
Create
< Previous
Download a template for automation

Step 3.

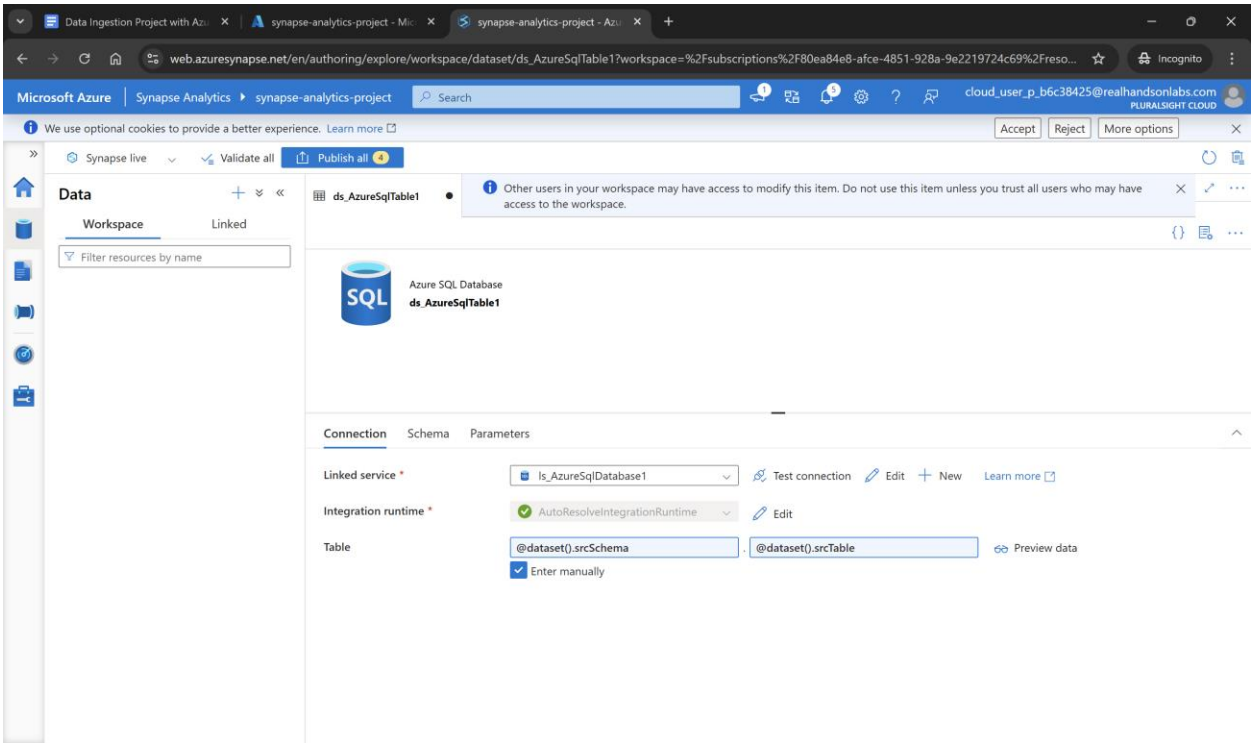

Data Ingestion Project with Azu
Create a key vault - Microsoft A
portal.azure.com/#create/Microsoft.KeyVault
Incognito
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
cloud_user_p_b6c38425...
PLURALSIGHT CLOUD
Home > Create a resource > Marketplace > Key Vault >
Create a key vault
Basics Access configuration Networking Tags Review + create
Review + create
Basics
Subscription P5-Real Hands-On Labs
Resource group 1-584bfabb-playground-sandbox
Key vault name keyvault-32
Region Central US
Pricing tier Standard
Soft-delete Enabled
Purge protection during retention period Disabled
Days to retain deleted vaults 90 days
Access configuration
Azure Virtual Machines for deployment Disabled
Azure Resource Manager for template deployment Disabled
Azure Disk Encryption for volume encryption Disabled
Permission model Vault access policy
Previous
Next
Create
Give feedback

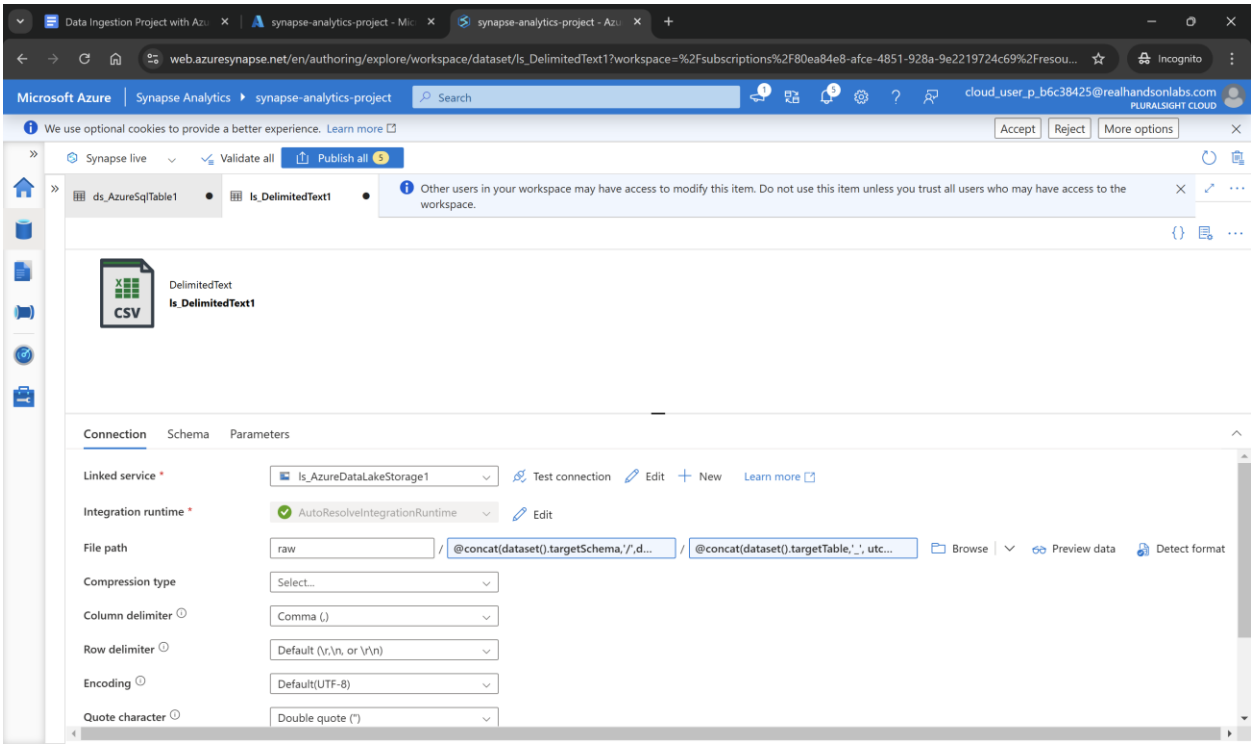

Data Ingestion Project with Azu
keyvault-32 - Microsoft Azure
portal.azure.com/#@realhandsonlabs.com/resource/subscriptions/80ea84e8-afce-4851-928a-9e2219724c69/resourceGroups/1-584bfabb-playground-sandbox/provider...
Incognito
Microsoft Azure
Search resources, services, and docs (G+/)
Copilot
cloud_user_p_b6c38425...
PLURALSIGHT CLOUD
Home > 1-584bfabb-playground-sandbox > keyvault-32
keyvault-32 | Secrets
Key vault
Creating the secret 'sqlpwd'.
The secret 'sqlpwd' has been successfully created.
Search
Generate/Import
Refresh
Restore Backup
View sample code
Manage deleted secrets
The secret 'sqlpwd' has been successfully created.
Overview
Activity log
Access control (IAM)
Tags
Diagnose and solve problems
Access policies
Events
Objects
Keys
Secrets
Certificates
Settings
Monitoring
Automation
Help
Name
Type
Status
Expiration date
sqlpwd
Enabled
Give feedback

Step 4.

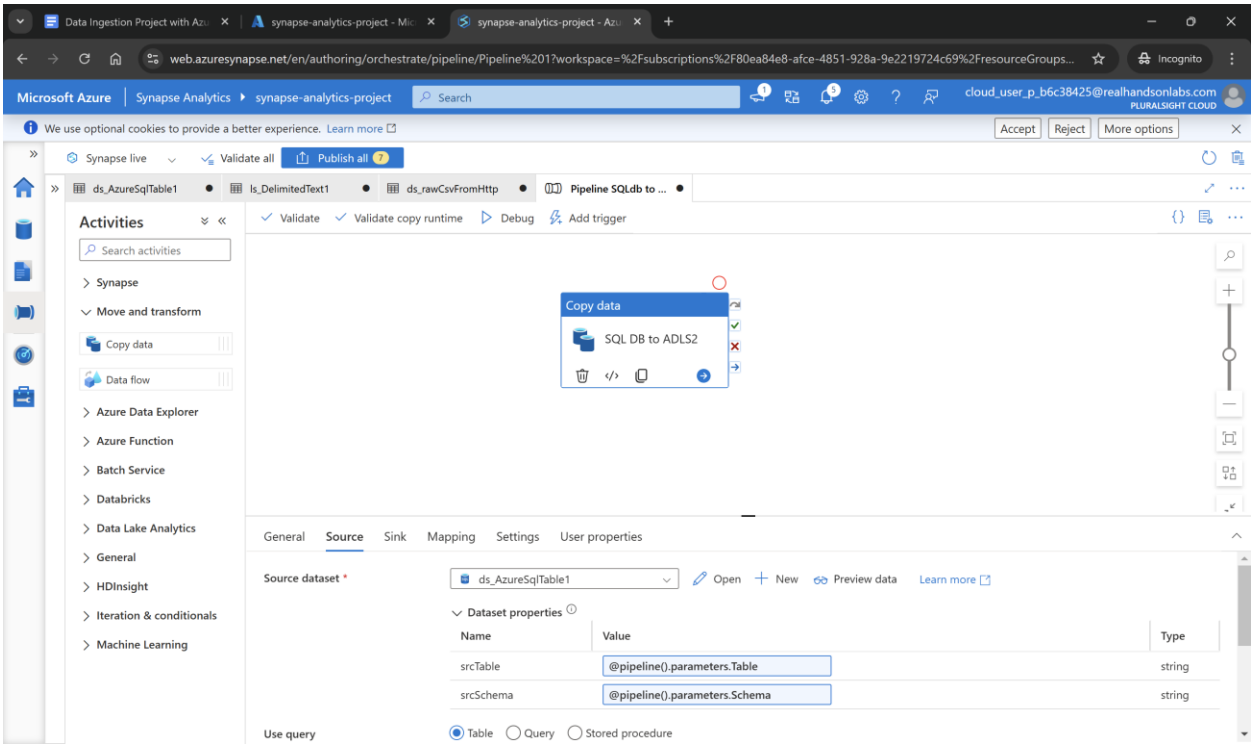

Data Ingestion Project with Azu
synapse-analytics-project - Mic
synapse-analytics-project - Azu
web.azuresynapse.net/en/management/datalinkedservices?workspace=%2Fsubscriptions%2F80ea84e8-afce-4851-928a-9e2219724c69%2FresourceGroups%2F1-584bfa...
Incognito
Microsoft Azure
Synapse Analytics
synapse-analytics-project
cloud_user_p_b6c38425@realhandsonlabs.com
PLURALSIGHT CLOUD
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 3
Will be created
ls_HttpServer1 (Linked service) will be created when publishing.
Analytics pools
SQL pools
Apache Spark pools
Data Explorer pools (pre...
External connections
Linked services
Microsoft Purview
Integration
Triggers
Integration runtimes
Security
Access control
Credentials
Managed private endpoi...
Configurations + libraries
Workspace packages
Data flow libraries
Apache Spark configurat...
Source control
Linked services
Linked services are much like connection strings, which define the connection information needed for Azure Synapse Analytics to connect to external resources.
Learn more
New
Filter by name
Annotations : Any
Showing 1 - 6 of 6 items
Name
Type
Related
Annotations
ls_AzureDataLakeStorage1
Azure Data Lake Storage Gen2
0
ls_AzureKeyVault1
Azure Key Vault
1
ls_AzureSqlDatabase1
Azure SQL Database
0
ls_HttpServer1
HTTP
0
synapse-analytics-project-WorkspaceD...
Azure Synapse Analytics
0
synapse-analytics-project-WorkspaceD...
Azure Data Lake Storage Gen2
0

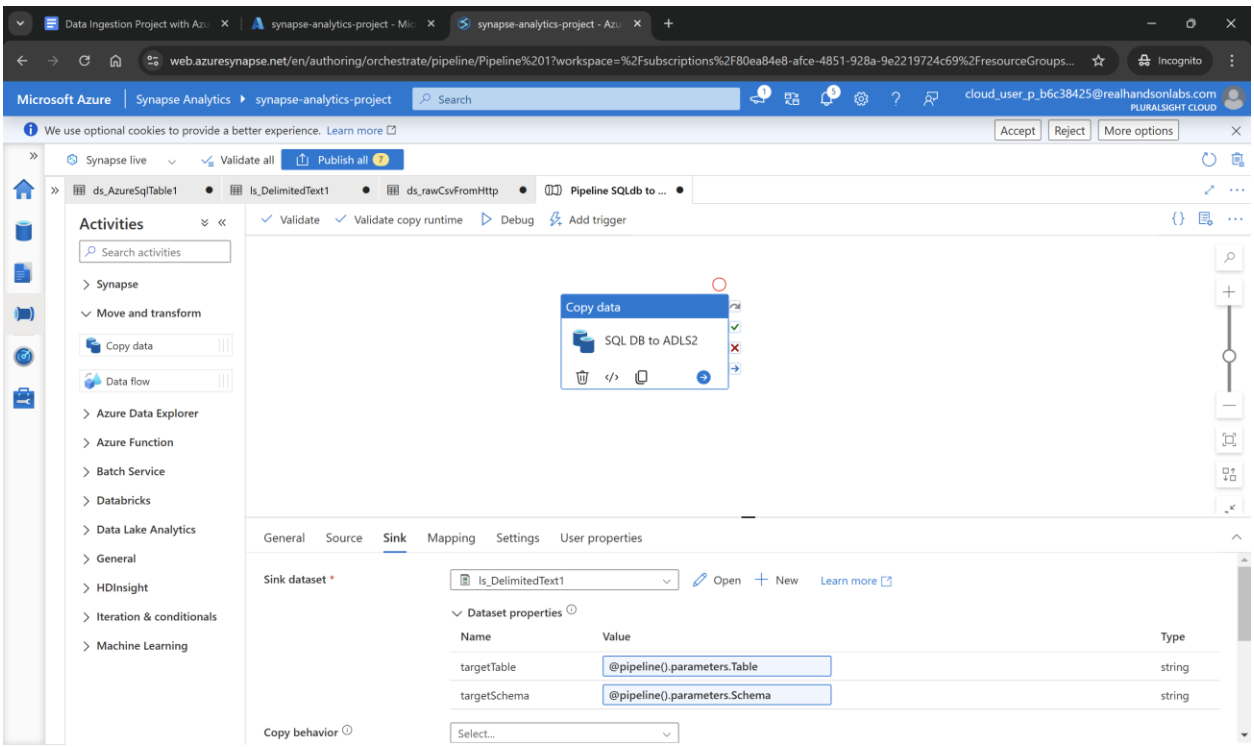
Step 5.

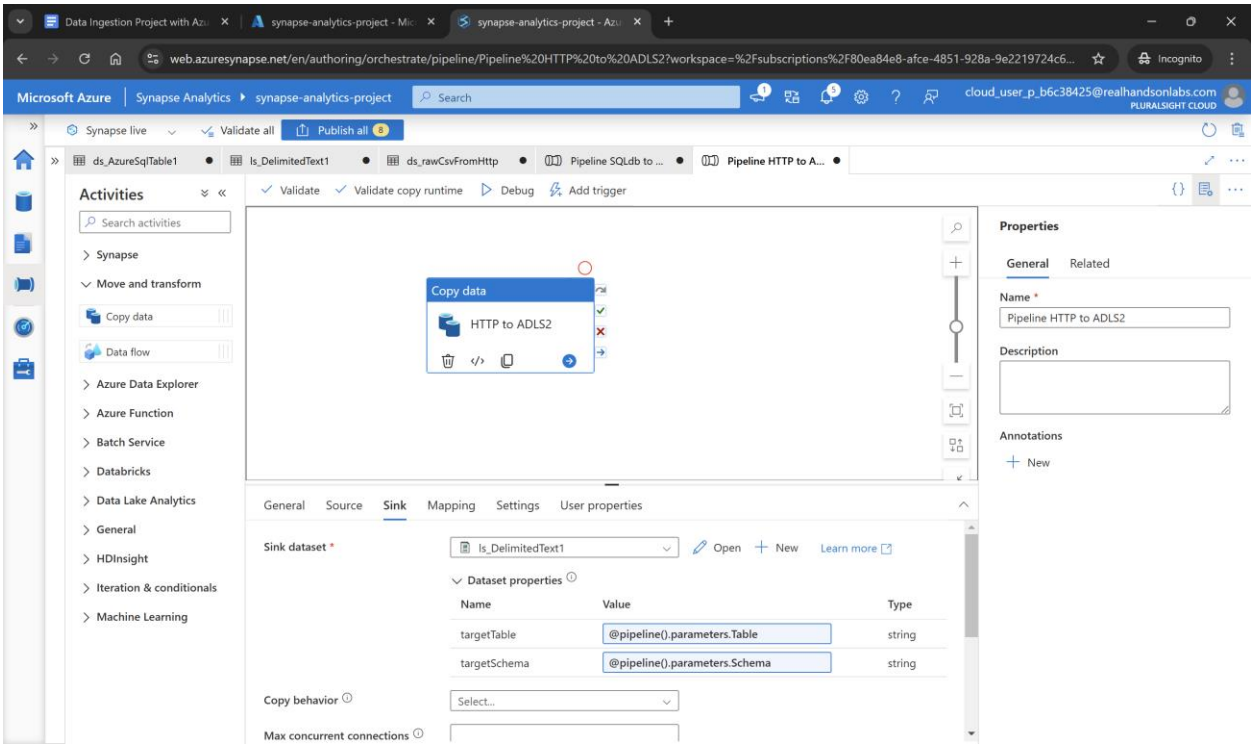

Data Ingestion Project with Azu
synapse-analytics-project - Mic
synapse-analytics-project - Azu
web.azuresynapse.net/en/authoring/explore/workspace/dataset/ds_AzureSqlTable1?workspace=%2Fsubscriptions%2F80ea84e8-afce-4851-928a-9e2219724c69%2Freso...
Incognito
Microsoft Azure
Synapse Analytics
synapse-analytics-project
Search
cloud_user_p_b6c38425@realhandsonlabs.com
PLURALSIGHT CLOUD
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 4
Data
Workspace
Linked
Filter resources by name
ds_AzureSqlTable1
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.
SQL
Azure SQL Database
ds_AzureSqlTable1
Connection
Schema
Parameters
Linked service *
ls_AzureSqlDatabase1
Test connection
Edit
New
Learn more
Integration runtime *
AutoResolveIntegrationRuntime
Edit
Table
@dataset().srcSchema
@dataset().srcTable
Preview data
Enter manually


Data Ingestion Project with Azu
synapse-analytics-project - Mic
synapse-analytics-project - Azu
web.azuresynapse.net/en/authoring/explore/workspace/dataset/ls_DelimitedText1?workspace=%2Fsubscriptions%2F80ea84e8-afce-4851-928a-9e2219724c69%2Fresou...
Incognito
Microsoft Azure
Synapse Analytics
synapse-analytics-project
Search
cloud_user_p_b6c38425@realhandsonlabs.com
PLURALSIGHT CLOUD
We use optional cookies to provide a better experience. Learn more
Accept
Reject
More options
Synapse live
Validate all
Publish all 5
ds_AzureSqlTable1
ls_DelimitedText1
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.
CSV
DelimitedText
ls_DelimitedText1
Connection
Schema
Parameters
Linked service *
ls_AzureDataLakeStorage1
Test connection
Edit
New
Learn more
Integration runtime *
AutoResolveIntegrationRuntime
Edit
File path
raw
@concat(dataset().targetSchema,'/',d...
@concat(dataset().targetTable,'_', utc...
Browse
Preview data
Detect format
Compression type
Select...
Column delimiter
Comma (,)
Row delimiter
Default (\r,\n, or \r\n)
Encoding
Default(UTF-8)
Quote character
Double quote (")

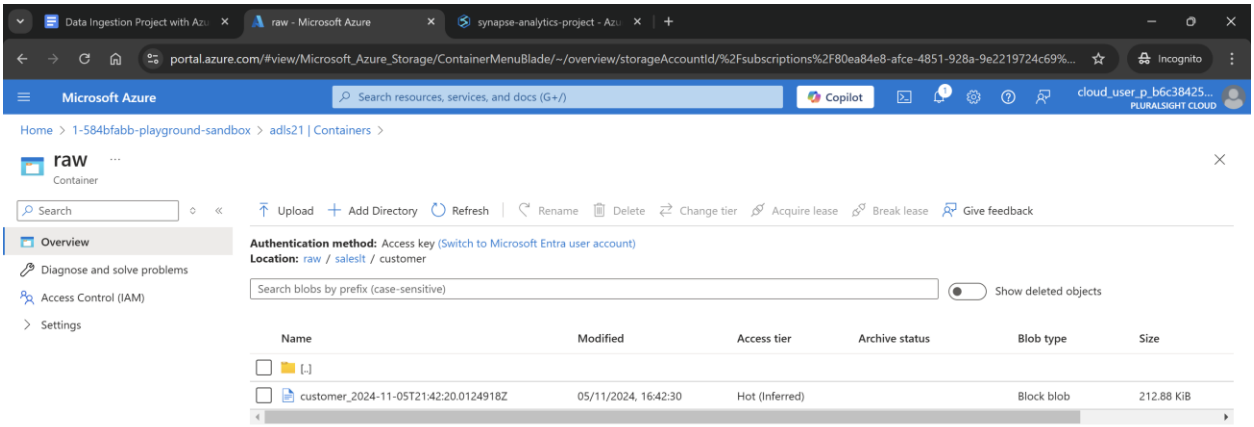
Step 6.

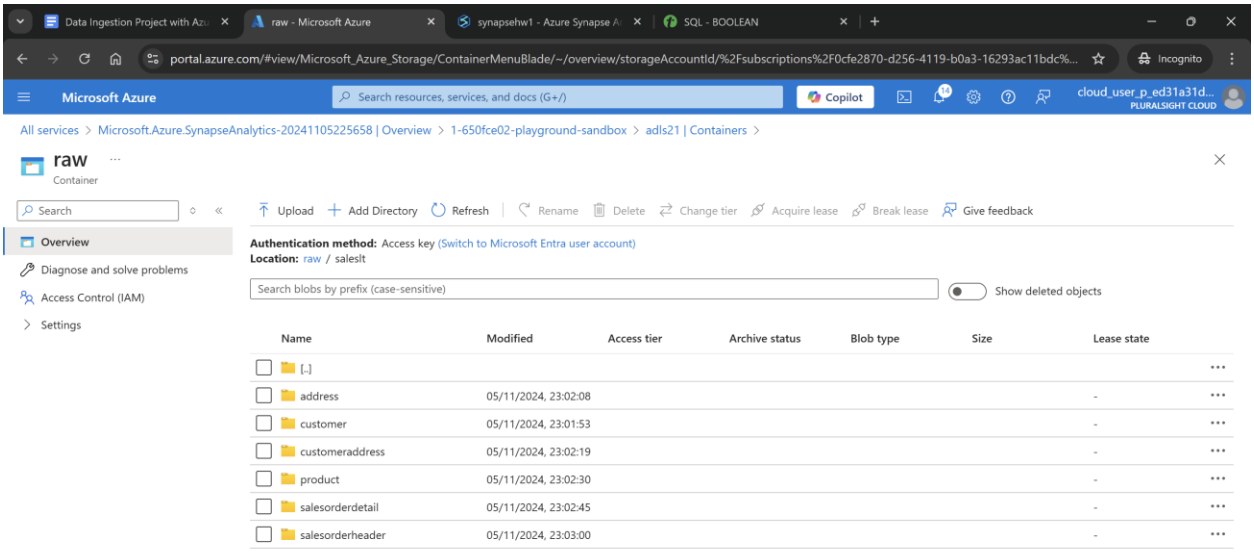

Microsoft Azure
Synapse Analytics
synapse-analytics-project
Search
Activities
Search activities
Synapse
Move and transform
Copy data
Data flow
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Validate
Validate copy runtime
Debug
Add trigger
Copy data
SQL DB to ADLS2
General
Source
Sink
Mapping
Settings
User properties
Source dataset *
ds_AzureSqlTable1
Open
New
Preview data
Learn more
Dataset properties
Name
Value
Type
srcTable
@pipeline().parameters.Table
string
srcSchema
@pipeline().parameters.Schema
string
Use query
Table
Query
Stored procedure


Microsoft Azure
Synapse Analytics
synapse-analytics-project
Search
Activities
Search activities
Synapse
Move and transform
Copy data
Data flow
Azure Data Explorer
Azure Function
Batch Service
Databricks
Data Lake Analytics
General
HDInsight
Iteration & conditionals
Machine Learning
Validate
Validate copy runtime
Debug
Add trigger
Copy data
SQL DB to ADLS2
General
Source
Sink
Mapping
Settings
User properties
Sink dataset *
ls_DelimitedText1
Open
New
Learn more
Dataset properties
Name
Value
Type
targetTable
@pipeline().parameters.Table
string
targetSchema
@pipeline().parameters.Schema
string
Copy behavior
Select...



After executing pipeline

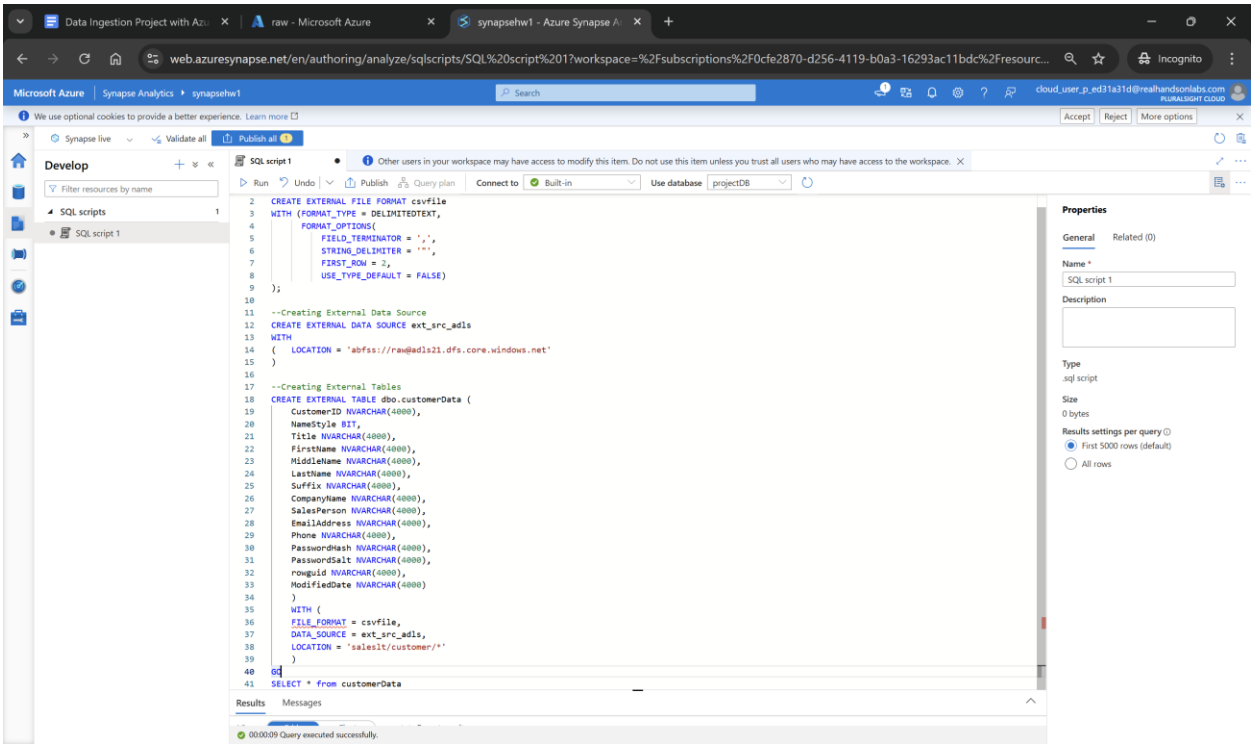


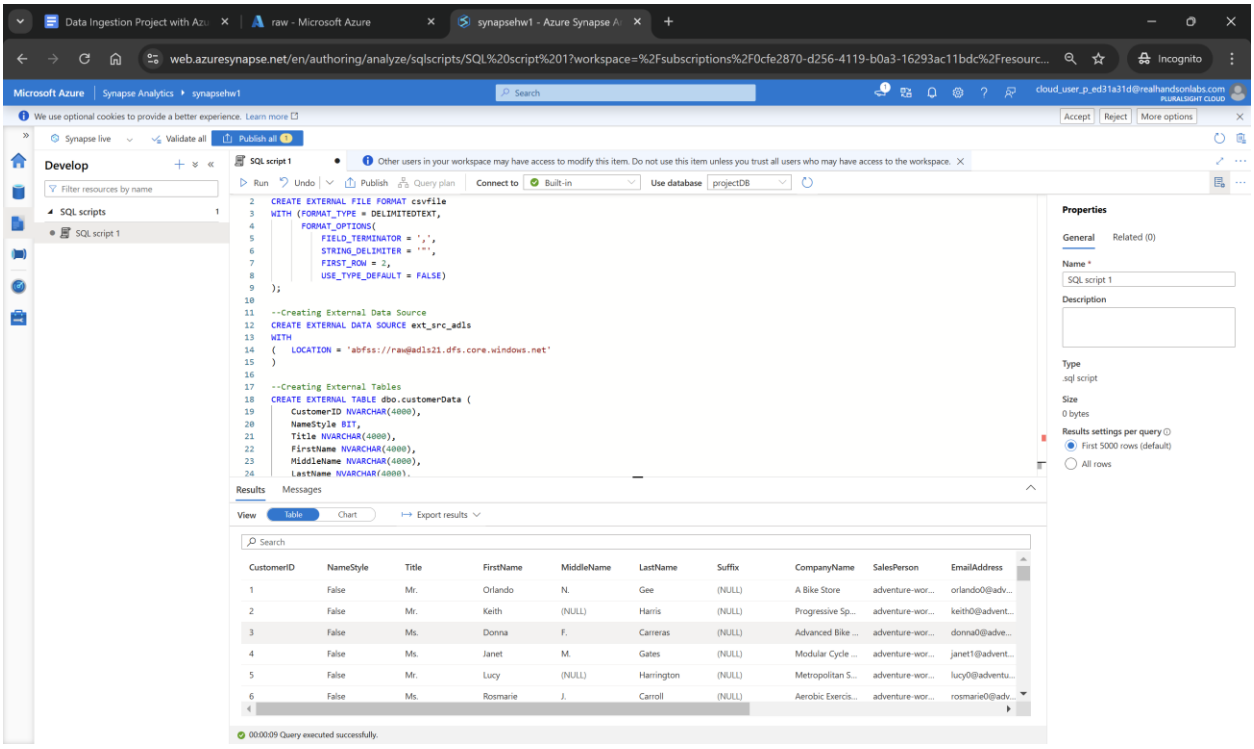


Step 7.

Creating External Table for Customer Data

Similarly create Tables for a ddress, customeraddress, product, salesorderdetail, salesorderheader



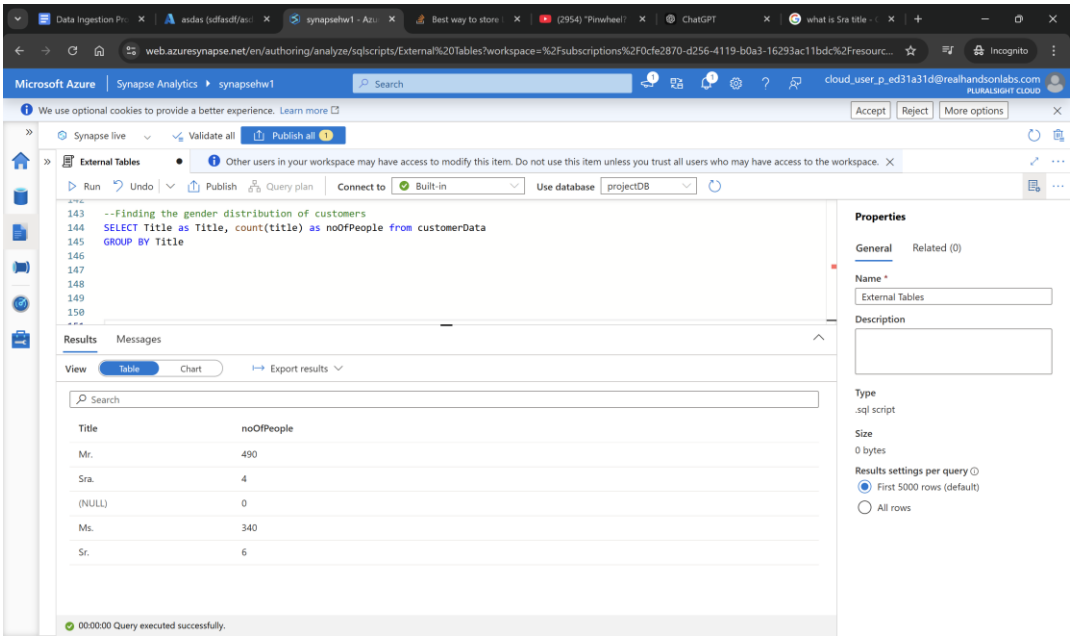


Step 8.

Querying data to perform data analysis

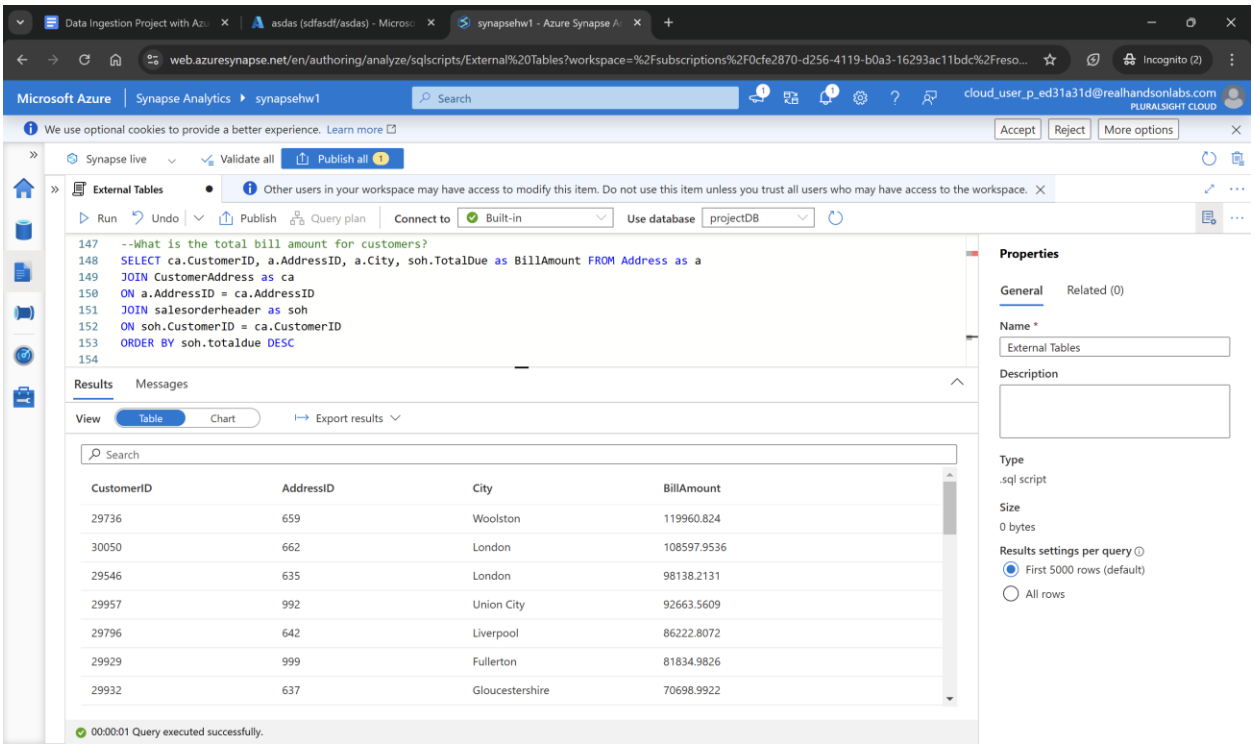
1. `--Finding the gender distribution of customers`

```
SELECT Title as Title, count(title) as noOfPeople from customerData
GROUP BY Title
```



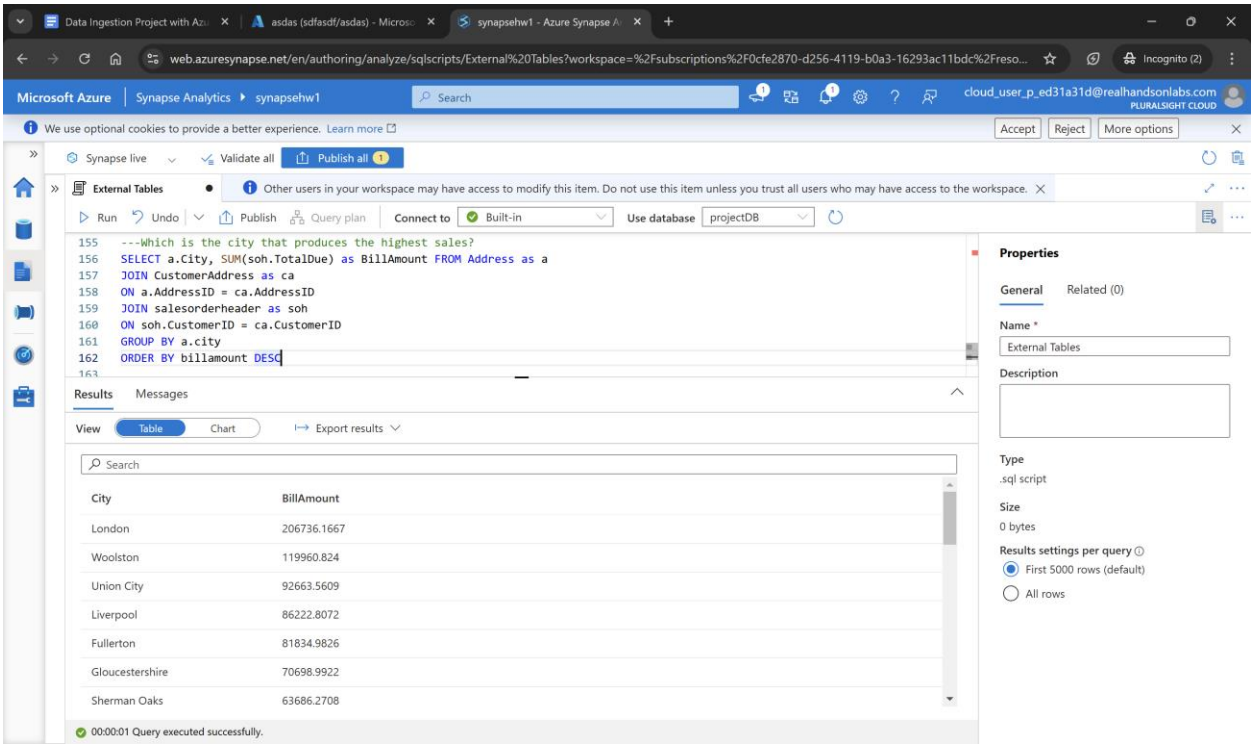
2. --What is the total bill amount for customers?

```
SELECT ca.CustomerID, a.AddressID, a.City, soh.TotalDue as BillAmount FROM
Address as a
JOIN CustomerAddress as ca
ON a.AddressID = ca.AddressID
JOIN salesorderheader as soh
ON soh.CustomerID = ca.CustomerID
ORDER BY soh.totaldue DESC
```



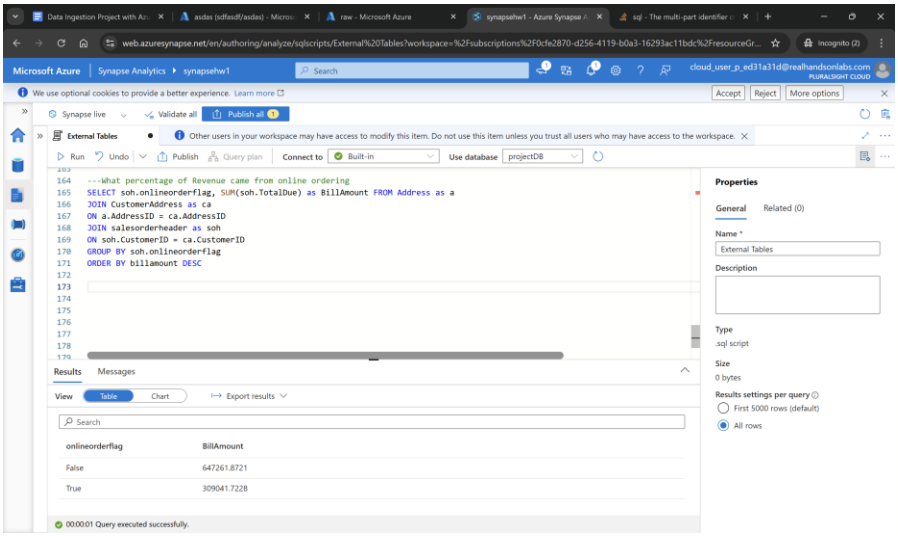
3. ---Which is the city that produces the highest sales?

```
SELECT a.City, SUM(soh.TotalDue) as BillAmount FROM Address as a
JOIN CustomerAddress as ca
ON a.AddressID = ca.AddressID
JOIN salesorderheader as soh
ON soh.CustomerID = ca.CustomerID
GROUP BY a.city
ORDER BY billamount DESC
```



4. ---What percentage of Revenue came from online ordering

```
SELECT soh.onlineorderflag, SUM(soh.TotalDue) as BillAmount FROM Address as a
JOIN CustomerAddress as ca
ON a.AddressID = ca.AddressID
JOIN salesorderheader as soh
ON soh.CustomerID = ca.CustomerID
GROUP BY soh.onlineorderflag
ORDER BY billamount DESC
```



```sql
5. --Find out the most popular product based on sales

SELECT p.name, sum(sod.lineTotal) as TotalSales  from product as p
JOIN salesorderdetail as sod
ON p.productid = sod.productid
GROUP by p.name
Order by totalsales desc
```

```sql
6. --What is the lifetime of the products sold?

SELECT (p.productid), (DATEDIFF(YEAR, p.sellstartdate, p.sellenddate)) as
productLifetime from product as p
where (DATEDIFF(YEAR, p.sellstartdate, p.sellenddate)) is not NULL
ORDER by p.productid asc
```