

Module 7 Lab

Querying an Azure Synapse Analytics Serverless SQL Pool

Lab Exercises

Data Note: The two data files used in this lab were produced by the following third parties. While you should be able to download and use them from the original sources, they are included with the lab files you've previously downloaded:

1. ***ZipCodeData.csv***. Contains a list of all physical zip codes in the United States, along with the city and state, latitude and longitude, population, and land area.
2. ***NationalCarSurvey.csv***: Contains the results of a national survey with over five million responses. Each row represents one respondent, and includes the respondent's zip code, average price of a car in respondent's household, the number of cars owned by members of the household, total commute time of the respondent, along with age group, gender, marital status and annual household income.

Summary

Key West Cars would like you to begin assessing the impact of a variety of factors on sales. This data is often in the form of large and/or semi-structured text files.

In this series of exercises, you'll aggregate the survey data in order to match it with Zip codes in the CarDealer database. In the next module we'll use Power BI to combine the data from both our semi-structured text files and the CarDealer database to visualize the relationship between population, income, and car sales by city and state.

Specifically, you're going to:

1. Create a Synapse Analytics Workspace;
2. Upload text files containing the survey and Zip code data to your Synapse Analytics workspace;
3. Create views using a serverless SQL Pool within your Synapse Analytics Workspace to return the necessary data from these text files as relational tables;

Exercise 1: Create an Azure Synapse Analytics Workspace and Upload Data Files

Goal: Create an Azure Synapse Analytics Workspace, and upload two data files to your Azure Data Lake Storage account:

Solution Steps:

1. In your Azure portal, on the **Home** page, click the **Create a Resource** button, and search for **Azure Synapse Analytics**.
2. In the **Azure Synapse Analytics** blade, click **Create**.
3. Select the **CarDealer** resource group.
4. Enter a unique workspace name. As was the case with your Azure SQL Database logical server, this must be a globally-unique name and will form part of the network address you'll use to connect to your Azure Synapse workspace. For example: **myuniquewsnamexyz**.
5. Select the region nearest you, such as **East US** or **West Europe**.
6. Under Select Data Lake Storage Gen2, ensure the **From subscription** radio button is selected.
7. Click **Create new** and enter a unique name for your Data Lake Storage Gen2 account name. For example: **mydlsaccountname**.
8. Click **Create new** and enter a name for your file system. For example: **myuniquefsname**.
9. Click **Next: Security**.

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

Home > Azure Synapse Analytics >

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](#)

Assign myself the Storage Blob Data Contributor role on the Data Lake Storage Gen2 account to interactively query it in the workspace.

i We will automatically grant the workspace identity data access to the specified Data Lake Storage Gen2 account, using the [Storage Blob Data Contributor](#) role. To enable other users to use this storage account 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
- Assign yourself and other users to the **Storage Blob Data Contributor** role on the storage account

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

10. Enter and confirm a password for sqladminuser. For this lab, use something you'll remember, such as **Pa55w.rd**.
11. Click **Review + Create**.

Home > Azure Synapse Analytics >

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 Azure Active Directory (Azure AD) authentication
 Use only Azure Active Directory (Azure AD) 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.

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.

Choose to encrypt all data at rest in the workspace with a key managed by you (customer-managed key). This will provide double encryption with encryption at the infrastructure layer that uses platform-managed keys. [Learn more](#)

Double encryption using a customer-managed key Enable Disable

Review + create < Previous Next: Networking >

12. Click **Create** after the validation. Wait for deployment to finish. Might take a few minutes.
13. After deployment completes, click the hamburger icon (the three stacked –'s in the upper left corner), and select **All resources**, then click the name of your **Synapse workspace**.
14. In the Synapse workspace blade, click **Open** under **Open Synapse Studio**.

Home > All resources >

myuniquewsname myuniquewsname
Synapse workspace

Search (Ctrl+/) << + New dedicated SQL pool + New Apache Spark pool + New Data Explorer pool (preview) Refresh Reset SQL admin password >>

Overview

- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems

Settings

- Azure Active Directory
- Properties
- Locks

Analytics pools

- SQL pools
- Apache Spark pools
- Data Explorer pools (preview)

Security

- Encryption
- Networking
- Identity
- Private endpoint connections
- Approved Azure AD tenants
- Azure SQL Auditing
- Microsoft Defender for Cloud

Essentials JSON View

Resource group ([move](#))
[CarDealer](#)

Status
Succeeded

Location
East US

Subscription ([move](#))
[Microsoft Partner Network](#)

Subscription ID
1e805e53-f99f-47f3-87a5-45f4274beeb5

Managed virtual network
No

Managed Identity object ID
26d37805-4785-4802-81f2-a5d89623943c

Workspace web URL
<https://web.azuresynapse.net?workspace=%2fsubscriptions%2f1e805e...>

Tags ([edit](#))
[Click here to add tags](#)

Networking
[Show firewall settings](#)

Primary ADLS Gen2 account URL
<https://myuniqueidsgen2.dfs.core.windows.net>

Primary ADLS Gen2 file system
myuniquefsname

SQL admin username
sqladminuser


SQL Active Directory admin
john@cubansandwichpress.com


Dedicated SQL endpoint
myuniquewsname.sql.azuresynapse.net

Serverless SQL endpoint
myuniquewsname-ondemand.sql.azuresynapse.net

Development endpoint
<https://myuniquewsname.dev.azuresynapse.net>

Getting started

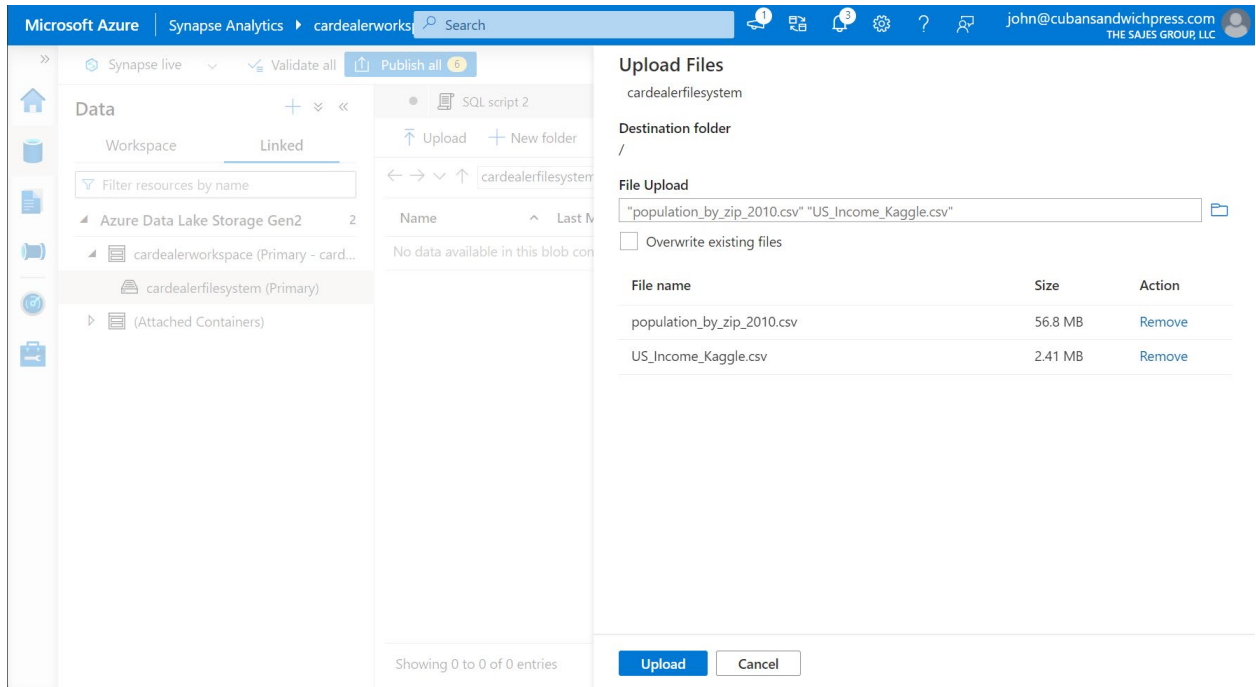
 **Open Synapse Studio**
Start building your fully-integrated analytics solution and unlock new insights.
[Open](#)

 **Read documentation**
Learn how to be productive quickly. Explore concepts, tutorials, and samples.
[Learn more](#)

Analytics pools

Search to filter items...

- Expand the leftmost panel to fully expose the **Home, Data, Develop, Integrate, Monitor,** and **Manage** sections.
- Click the **Data** section, then click **Linked** (to the right of **Workspace**).
- Expand Azure Data Lake Storage Gen2, and expand your primary storage account, and click the name of your file system.



18. Click **Upload**, and upload the following file:
 - a. C:\SQLForDA\LabFiles**ZipCodeData.csv**
19. Repeat the above step and upload two more files:
 - a. C:\SQLForDA \LabFiles**NationalCarSurvey.csv**

Exercise 2: Create a View for Zip Code Data

Goal:

Create a view that returns the city, state, latitude, longitude, population and land area for each Zip code.

Before we create the view, we'll need to create a database on the serverless SQL pool.

Solution Steps:

1. In Synapse Studio, navigate to your Azure Data Lake Storage Gen2 filesystem on the Data page.
2. Right-click **ZipCodeData.csv**, and select **New SQL script > Select TOP 100 rows**.
3. Add **HEADER_ROW=TRUE** to the **OPENROWSET** function so that columns headers display correctly, then click **Run**.
4. Modify the query to return the desired columns as follows:

```
SELECT
    Zipcode
    ,City
    ,State
    ,Lat
    ,Long
```

```

    ,Population
    ,LandArea
FROM
    OPENROWSET(
        BULK
        'https://mydlsaccountname.dfs.core.windows.net/mydlsfilesystem/ZipCodeData.csv',
        FORMAT = 'CSV',
        PARSER_VERSION = '2.0',
        header_row=TRUE
    ) AS [result];

```

- Next, we'll create a database in the serverless SQL pool. Click the **Develop** button in left side navigation panel, then click the **plus (+)** sign, and choose **SQL script**. Create a database using UTF8 encoding. For example:

```

Create database SupplementalCarData
    COLLATE Latin1_General_100_CI_AS_SC_UTF8

```

- Select your original script, then change the default database in the **Use database** list to SupplementalCarData. Create a view named *vZipCodeData* for the query. For example:

Create view ZipCodeData

```

as
SELECT
    Zipcode
    ,City
    ,State
    ,Lat
    ,Long
    ,Population
    ,LandArea
FROM
    OPENROWSET(
        BULK
        'https://mydlsaccountname.dfs.core.windows.net/mydlsfilesystem/ZipCodeData.csv',
        FORMAT = 'CSV',
        PARSER_VERSION = '2.0',
        header_row=TRUE
    ) AS [result];

```

Exercise 3: Create a View for Survey Data

Goal:

Create a view that returns the mean car price, annual household income, number of cars owned per household, and commute time for each Zip code.

Solution Steps:

1. In Synapse Studio, navigate to your Azure Data Lake Storage Gen2 filesystem on the Data page.
2. Right-click **NationalCarSurvey.csv**, and select **New SQL script > Select TOP 100 rows**.
3. Be sure the **SupplementalCarData** database you created above is selected in the **Use database** drop-down list;
4. Add **HEADER_ROW=TRUE** to the **OPENROWSET** function so that columns headers display correctly, then click **Run**.
5. Modify the query to return the desired columns as follows:

```
SELECT
    ZipCode
    ,avg(MeanCarPrice) as [Mean Car Price]
    ,avg(AnnualHouseholdIncome) as [Mean Household Income]
    ,avg(NumberOfCarsOwned) as [Mean Number of Cars Owned]
    ,avg(CommuteTime) as [Mean Commute Time]
FROM
    OPENROWSET(
        BULK
        'https://mydlsaccountname.dfs.core.windows.net/mydlsfilesystem/NationalCarSurvey.csv',
        FORMAT = 'CSV',
        PARSER_VERSION = '2.0',
        HEADER_ROW=TRUE
    ) AS [result]
group by ZipCode;
```

6. Finally, let's create a view in the **SupplementalData** database for this query.

```
create view SurveyResults
as
SELECT
    ZipCode
    ,avg(MeanCarPrice) as [Mean Car Price]
    ,avg(AnnualHouseholdIncome) as [Mean Household Income]
    ,avg(NumberOfCarsOwned) as [Mean Number of Cars Owned]
    ,avg(CommuteTime) as [Mean Commute Time]
FROM
    OPENROWSET(
```

```
BULK
'https://mydlsaccountname.dfs.core.windows.net/mydlsfilesystem/NationalCarSurvey.csv',
    FORMAT = 'CSV',
    PARSER_VERSION = '2.0',
    HEADER_ROW=TRUE
) AS [result]
group by ZipCode;
```

7. Now that we have views encapsulating our data, we can easily incorporate the data into a Power BI data model along with the rest of the data from the *CarDealer* database. We'll do this in the next module.