

# 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.

Validation succeeded

\* Basics \* Security Networking Tags **Review + create**

**Product Details**

Azure Synapse Analytics workspace by Microsoft  
Serverless SQL est. cost/TB 5.00 USD  
[Terms of use](#) | [Privacy policy](#)

**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
Workspace name	(new) synapse-analytics-project
Data Lake Storage Gen2 account	(new) https://adls21.dfs.core.windows.net

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

## Step 2.

Microsoft Azure portal interface showing the "Create SQL Database" wizard. The browser address bar shows "portal.azure.com/#create/Microsoft.SQLDatabase". The page title is "Create SQL Database". The "Review + create" tab is selected.

**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

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

**Cost summary**

<b>Basic (Basic)</b>	
Cost per DTU (in USD)	0.98
DTUs selected	x 5
<b>ESTIMATED COST / MONTH</b>	<b>4.90 USD</b>

### Step 3.

Microsoft Azure portal interface showing the "Create a key vault" wizard. The browser address bar shows "portal.azure.com/#create/Microsoft.KeyVault". The page title is "Create a key vault". The "Review + create" tab is selected.

**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](#)

Microsoft Azure portal showing the 'keyvault-32 | Secrets' page. The page displays a notification: 'Creating the secret 'sqlpwd'. The secret 'sqlpwd' has been successfully created.'

The left sidebar shows the navigation menu with 'Secrets' selected. The main content area shows a table of secrets:

Name	Type	Status	Expiration date
sqlpwd		✓ Enabled	

Buttons at the top include: Generate/Import, Refresh, Restore Backup, View sample code, and Manage deleted secrets.

Give feedback

## Step 4.

Microsoft Azure portal showing the 'Synapse Analytics' page. The page displays a notification: 'Will be created Is\_HttpServer1 (Linked service) will be created when publishing.'

The left sidebar shows the navigation menu with 'Linked services' selected. The main content area shows a table of linked services:

Name	Type	Related	Annotations
Is_AzureDataLakeStorage1	Azure Data Lake Storage Gen2	0	
Is_AzureKeyVault1	Azure Key Vault	1	
Is_AzureSqlDatabase1	Azure SQL Database	0	
Is_HttpServer1	HTTP	0	
synapse-analytics-project-WorkspaceD...	Azure Synapse Analytics	0	
synapse-analytics-project-WorkspaceD...	Azure Data Lake Storage Gen2	0	

Buttons at the top include: Validate all, Publish all, and New.

## Step 5.

The screenshot shows the Microsoft Azure Synapse Analytics interface. The browser address bar indicates the URL: `web.azuresynapse.net/en/authoring/explore/workspace/dataset/ds_AzureSqlTable1?workspace=%2Fsubscriptions%2F80ea84e8-afce-4851-928a-9e2219724c69%2Fresou...`. The interface displays the 'Data' section with a 'Workspace' tab selected. A search bar is present, and a notification states: '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.'

The dataset 'ds\_AzureSqlTable1' is shown with an 'Azure SQL Database' icon. Below the dataset name, the 'Connection' tab is active, showing the following configuration:

- Linked service \***: `Is_AzureSqlDatabase1` (with 'Test connection', 'Edit', '+ New', and 'Learn more' links).
- Integration runtime \***: `AutoResolveIntegrationRuntime` (with an 'Edit' link).
- Table**: `@dataset().srcSchema` and `@dataset().srcTable` (with a 'Preview data' link).
- Enter manually**: ☒

The screenshot shows the Microsoft Azure Synapse Analytics interface. The browser address bar indicates the URL: `web.azuresynapse.net/en/authoring/explore/workspace/dataset/Is_DelimitedText1?workspace=%2Fsubscriptions%2F80ea84e8-afce-4851-928a-9e2219724c69%2Fresou...`. The interface displays the 'Data' section with a 'Workspace' tab selected. A search bar is present, and a notification states: '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.'

The dataset 'Is\_DelimitedText1' is shown with a 'DelimitedText CSV' icon. Below the dataset name, the 'Connection' tab is active, showing the following configuration:

- Linked service \***: `Is_AzureDataLakeStorage1` (with 'Test connection', 'Edit', '+ New', and 'Learn more' links).
- Integration runtime \***: `AutoResolveIntegrationRuntime` (with an 'Edit' link).
- File path**: `raw` / `@concat(dataset().targetSchema, '/', d...` / `@concat(dataset().targetTable, '_', utc...` (with 'Browse', 'Preview data', and 'Detect format' links).
- Compression type**: `Select...`
- Column delimiter**: `Comma (,)`
- Row delimiter**: `Default (\n or \r\n)`
- Encoding**: `Default(UTF-8)`
- Quote character**: `Double quote (")`

## Step 6.

Microsoft Azure | Synapse Analytics | synapse-analytics-project

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

Copy data

SQL DB to ADLS2

Source dataset: ds\_AzureSqlTable1

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

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

Copy data

SQL DB to ADLS2

Sink dataset: ls\_DelimitedText1

Dataset properties

Name	Value	Type
targetTable	@pipeline().parameters.Table	string
targetSchema	@pipeline().parameters.Schema	string

Copy behavior: Select...

Microsoft Azure | Synapse Analytics | synapse-analytics-project

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

Copy data

HTTP to ADLS2

Properties

General

Name: Pipeline HTTP to ADLS2

Description

Annotations

General Source Sink Mapping Settings User properties

Sink dataset: ls\_DelimitedText1

Dataset properties

Name	Value	Type
targetTable	@pipeline().parameters.Table	string
targetSchema	@pipeline().parameters.Schema	string

Copy behavior: Select...

Max concurrent connections

## After executing pipeline

Microsoft Azure

raw

Container

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

Location: raw / saleslt / customer

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size
[.]					
customer_2024-11-05T21:42:20.0124918Z	05/11/2024, 16:42:30	Hot (Inferred)		Block blob	212.88 KiB

Microsoft Azure

Search resources, services, and docs (G+)

cloud\_user\_p\_ed31a31d...  
PLURALSIGHT CLOUD

All services > Microsoft.Azure.SynapseAnalytics-20241105225658 | Overview > 1-650fce02-playground-sandbox > adls21 | Containers >

**raw**  
Container

Search

Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease Give feedback

**Overview**

Diagnose and solve problems

Access Control (IAM)

Settings

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

Location: raw / saleslt

Search blobs by prefix (case-sensitive)

Show deleted objects

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
[-]						...
address	05/11/2024, 23:02:08					...
customer	05/11/2024, 23:01:53					...
customeraddress	05/11/2024, 23:02:19					...
product	05/11/2024, 23:02:30					...
salesorderdetail	05/11/2024, 23:02:45					...
salesorderheader	05/11/2024, 23:03:00					...

## Step 7.

### Creating External Table for Customer Data

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

Microsoft Azure | Synapse Analytics | synapsehw1

Search

cloud\_user\_p\_ed31a31d@realhandsontabs.com  
PLURALSIGHT CLOUD

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

Synapse live Validate all Publish all

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.

Connect to Built-in Use database projectDB

Develop

Filter resources by name

SQL scripts

SQL script 1

```

1 CREATE EXTERNAL FILE FORMAT csvfile
2 WITH (FORMAT_TYPE = DELIMITEDTEXT,
3       FORMAT_OPTIONS(
4         FIELD_TERMINATOR = ',',
5         STRING_DELIMITER = '',
6         FIRST_ROW = 2,
7         USE_TYPE_DEFAULT = FALSE)
8 );
9
10
11 --Creating External Data Source
12 CREATE EXTERNAL DATA SOURCE ext_src_adls
13 WITH
14 ( LOCATION = 'abfss://raw@adls21.dfs.core.windows.net'
15 )
16
17 --Creating External Tables
18 CREATE EXTERNAL TABLE dbo.customerData (
19   CustomerID NVARCHAR(4000),
20   NameStyle BIT,
21   Title NVARCHAR(4000),
22   FirstName NVARCHAR(4000),
23   MiddleName NVARCHAR(4000),
24   LastName NVARCHAR(4000),
25   Suffix NVARCHAR(4000),
26   CompanyName NVARCHAR(4000),
27   SalesPerson NVARCHAR(4000),
28   EmailAddress NVARCHAR(4000),
29   Phone NVARCHAR(4000),
30   PasswordHash NVARCHAR(4000),
31   PasswordSalt NVARCHAR(4000),
32   rowguid NVARCHAR(4000),
33   ModifiedDate NVARCHAR(4000)
34 )
35 WITH (
36   FILE_FORMAT = csvfile,
37   DATA_SOURCE = ext_src_adls,
38   LOCATION = 'saleslt/customer/'
39 )
40
41 SELECT * from customerData

```

Results Messages

000009 Query executed successfully.

Properties

General Related (0)

Name \*  
SQL script 1

Description

Type  
sql script

Size  
0 bytes

Results settings per query

First 5000 rows (default)

All rows

**SQL script 1**

```

2 CREATE EXTERNAL FILE FORMAT csvFile
3 WITH (FORMAT_TYPE = DELIMITEDTEXT,
4     FORMAT_OPTIONS(
5         FIELD_TERMINATOR = ',',
6         STRING_DELIMITER = "'",
7         FIRST_ROW = 2,
8         USE_TYPE_DEFAULT = FALSE)
9 );
10
11 --Creating External Data Source
12 CREATE EXTERNAL DATA SOURCE ext_src_adls
13 WITH
14 ( LOCATION = 'abfss://raw@ad1s21.dfs.core.windows.net'
15 )
16
17 --Creating External Tables
18 CREATE EXTERNAL TABLE dbo.customerData (
19     CustomerID NVARCHAR(4000),
20     NameStyle BIT,
21     Title NVARCHAR(4000),
22     FirstName NVARCHAR(4000),
23     MiddleName NVARCHAR(4000),
24     LastName NVARCHAR(4000)
25 );

```

**Results**

CustomerID	NameStyle	Title	FirstName	MiddleName	LastName	Suffix	CompanyName	SalesPerson	EmailAddress
1	False	Mr.	Orlando	N.	Gee	(NULL)	A Bike Store	adventure-wor...	orlando@adv...
2	False	Mr.	Keith	(NULL)	Harris	(NULL)	Progressive Sp...	adventure-wor...	keith@adv...
3	False	Ms.	Donna	F.	Carreras	(NULL)	Advanced Bike ...	adventure-wor...	donna@adv...
4	False	Ms.	Janet	M.	Gates	(NULL)	Modular Cycle ...	adventure-wor...	janet1@adv...
5	False	Mr.	Lucy	(NULL)	Harrington	(NULL)	Metropolitan S...	adventure-wor...	lucy@adv...
6	False	Ms.	Rosmarie	J.	Carroll	(NULL)	Aerobic Exercis...	adventure-wor...	rosmarie@adv...

00:00:09 Query executed successfully.

## 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

**SQL script 1**

```

143 --Finding the gender distribution of customers
144 SELECT Title as Title, count(title) as noOfPeople from customerData
145 GROUP BY Title
146
147
148
149
150

```

**Results**

Title	noOfPeople
Mr.	490
Sra.	4
(NULL)	0
Ms.	340
Sr.	6

00:00:00 Query executed successfully.



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
```

The screenshot shows the Microsoft Azure Synapse Analytics interface. The query editor displays the following SQL query:

```
147 --What is the total bill amount for customers?
148 SELECT ca.CustomerID, a.AddressID, a.City, soh.TotalDue as BillAmount FROM Address as a
149 JOIN CustomerAddress as ca
150 ON a.AddressID = ca.AddressID
151 JOIN salesorderheader as soh
152 ON soh.CustomerID = ca.CustomerID
153 ORDER BY soh.totaldue DESC
154
```

The results are displayed in a table with the following columns: CustomerID, AddressID, City, and BillAmount. The results show 8 rows of data, with the highest bill amount being 119960.824 for CustomerID 29736 in Woolston.

CustomerID	AddressID	City	BillAmount
29736	659	Woolston	119960.824
30050	662	London	108597.9536
29546	635	London	98138.2131
29957	992	Union City	92663.5609
29796	642	Liverpool	86222.8072
29929	999	Fullerton	81834.9826
29932	637	Gloucestershire	70698.9922

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
```

The screenshot shows the Azure Synapse Analytics web interface. The query editor contains the following SQL code:

```

155 ---Which is the city that produces the highest sales?
156 SELECT a.City, SUM(soh.TotalDue) as BillAmount FROM Address as a
157 JOIN CustomerAddress as ca
158 ON a.AddressID = ca.AddressID
159 JOIN salesorderheader as soh
160 ON soh.CustomerID = ca.CustomerID
161 GROUP BY a.city
162 ORDER BY billamount DESC

```

The results are displayed in a table with two columns: City and BillAmount.

City	BillAmount
London	206736.1667
Woolston	119960.824
Union City	92663.5609
Liverpool	86222.8072
Fullerton	81834.9826
Gloucestershire	70698.9922
Sherman Oaks	63686.2708

The status bar indicates: 00:00:01 Query executed successfully.

#### 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

```

The screenshot shows the Azure Synapse Analytics web interface. The query editor contains the following SQL code:

```

164 ---What percentage of Revenue came from online ordering
165 SELECT soh.onlineorderflag, SUM(soh.TotalDue) as BillAmount FROM Address as a
166 JOIN CustomerAddress as ca
167 ON a.AddressID = ca.AddressID
168 JOIN salesorderheader as soh
169 ON soh.CustomerID = ca.CustomerID
170 GROUP BY soh.onlineorderflag
171 ORDER BY billamount DESC

```

The results are displayed in a table with two columns: onlineorderflag and BillAmount.

onlineorderflag	BillAmount
False	647261.8721
True	309041.7228

The status bar indicates: 00:00:01 Query executed successfully.

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
```

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
```