

ANALYZE DATA IN A LAKE DATABASE

**SUBMITTED BY
NITIN**



OBJECTIVES

This Lab aims to build essential data management and analysis skills within a lake database environment using Azure Synapse Analytics. I learned to set up cloud resources, create and load data tables, manipulate lake database tables, and perform data operations with Spark. The assignment enhanced my ability to work with large datasets in a data lake context, equipping me with practical knowledge for future data engineering roles.

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

14% Completed
Lab Time Left: 01:46:44

Lab Actions

4. Analyze data in a lake database

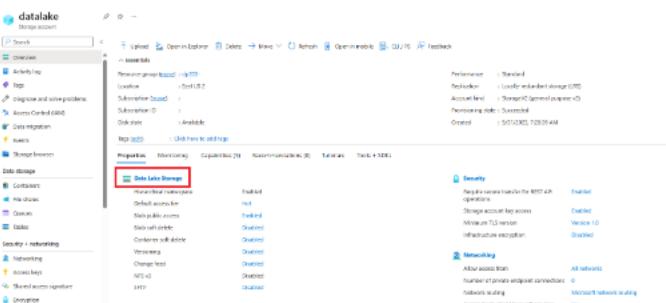
Modify container permissions:

After the deployment script has been completed, in the Azure portal, go to the **dp203-xxxxxx** resource group that it created, and notice that this resource group contains your Synapse workspace, a Storage account for your data lake, and an Apache Spark pool.

1. Select the **Storage account** for your data lake named **datalakexxxxxxxx**

2. Select the **Data Lake Storage** link to go to the **datalakexxxxxxxx** container.

3. Within the **datalakexxxxxxxx** container, select the **files** folder




Page: 4/14

Support

Microsoft Azure | Microsoft | datalakejhnxouv - Microsoft Azure

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

Copilot

Home > Resource groups > dp203-jhnxouv >

datalakejhnxouv

Storage account

Search

Upload Open in Explorer Delete Move Refresh Open in mobile ...

Overview

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

Essentials

Resource group ([move](#)) [dp203-jhnxouv](#) Performance Standard Location westus2 Replication Locally-redundant storage (LRS) Account kind StorageV2 (general purpose v2) Subscription ([move](#)) [XLabs-973](#) Provisioning state Succeeded Subscription ID 40911347-4086-47ba-a836-c020ead78f4d Disk state Available Creation time 11/3/2024, 5:28:55 PM Tags ([edit](#)) Add tags

Properties Monitoring Capabilities (5) Recommendations (0) Tutorials Tools + SDKs

Data Lake Storage

Hierarchical namespace Enabled

Security

Require secure transfer for REST API operations

5:33 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabManua... Microsoft Azure Portal | Microsoft Azure synapsejhnxouv - Microsoft Azure synapsejhnxouv - Azure Synaps +

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

21% Completed
Lab Time Left: 01:42:57

Lab Actions ▾

5. Analyze data in a lake database

Create a lake database:

- A lake database is a type of database that you can define in your workspace, and work with using the built-in serverless SQL pool.
- 1. Select your Synapse workspace, and in its **Overview** page, in the **Open Synapse Studio** card, select **Open** to open Synapse Studio in a new browser tab; signing in if prompted.
- 2. On the left side of Synapse Studio, use the **»** icon to expand the menu - this reveals the different pages within Synapse Studio that you'll use to manage resources and perform data analytics tasks.
- 3. On the **Data** page, view the **Linked** tab and verify that your workspace includes a link to your Azure Data Lake Storage Gen2 storage account.
- 4. On the **Data** page, switch back to the **Workspace** tab and note that there are no databases in your workspace.
- 5. In the **+** menu, select **Lake database** to open a new tab in which you can design your database schema (accepting the database templates terms of use if prompted).
- 6. In the **Properties** pane for the new database, change the **Name** to **RetailDB** and verify that the **Input folder**

Page: 5/14

Support

Microsoft Azure | Synapse Analytics ▶ Search

Synapse live ▾ Validate all Publish all

Home Data Develop Integrate Monitor Manage

Data

Workspace Linked

Filter resources by name

Azure Data Lake Storage Gen2 2

synapsejhnxouv (Primary - datalake...) (Attached Containers)

Select an item

Use the resource explorer to select or create a new item

5:37 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabMan... Microsoft Azure Portal | synapsejhnxouv - Microsoft Edge synapsejhnxouv - Azure New Tab

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

35% Completed
Lab Time Left: 01:27:50

Lab Actions

6. Analyze data in a lake database
Create a table:
Now that you have created a lake database, you can define its schema by creating tables.

Define the table schema

1. Switch back to the **RetailDB** tab for your database definition, and in the **+ Table** list, select **Custom**, and note that a new table named **Table_1** is added to your database.
2. With **Table_1** selected, in the **General** tab under the database design canvas, change the **Name** property to **Customer**.
3. Expand the **Storage settings for table** section and note that the table will be stored as delimited text in the **files/RetailDB/Customer** folder in the default data lake store for your Synapse workspace.
4. On the **Columns** tab, note that by default, the table contains one column named **Column_1**. Edit the column definition to match the following properties:

Name	Keys	Description	Nullability	Data type	Format / Length
CustomerId	PK <input checked="" type="checkbox"/>	Unique customer ID	<input type="checkbox"/>	long	

5. In the **+ Column** list, select **New column**, and modify the new column definition to add a **FirstName** column to the table as follows:

Name	Keys	Description	Nullability	Data type	Format / Length
CustomerId	PK <input checked="" type="checkbox"/>	Unique customer ID	<input type="checkbox"/>	long	
FirstName	PK <input type="checkbox"/>	Customer first name	<input type="checkbox"/>	string	256

6. Add more new columns until the table definition looks like this:

Page: 6/14

Synapse Analytics > synapsejhnxouv Search

Synapse live Validate all Publish all 1

RetailDB files

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.

Customer

Column_1 PK

See less

General Columns Relationships

Filter by keyword + Column Clone Convert type Delete

Name	Keys	Description	Nullability	Data type	Format / Length
Column_1	PK <input checked="" type="checkbox"/>	Unique customer ID	<input type="checkbox"/>	long	121 long
FirstName	PK <input type="checkbox"/>	Customer first name	<input type="checkbox"/>	string	256

Standard column (1)

Column_1 PK Unique customer ID Null long

Partition column (0)

5:52 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabMan... Microsoft Azure Portal | synapsejhnxouv - Micro... synapsejhnxouv - Azure New Tab

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

Name Keys Description Nullability Data type Format / Length

CustomerId	PK <input checked="" type="checkbox"/>	Unique customer ID	<input type="checkbox"/>	long	
------------	--	--------------------	--------------------------	------	--

5. In the + Column list, select New column, and modify the new column definition to add a FirstName column to the table as follows:

Name	Keys	Description	Nullability	Data type	Format / Length
CustomerId	PK <input checked="" type="checkbox"/>	Unique customer ID	<input type="checkbox"/>	long	
FirstName	PK <input type="checkbox"/>	Customer first name	<input type="checkbox"/>	string	256

6. Add more new columns until the table definition looks like this:

Name	Keys	Description	Nullability	Data type	Format / Length
CustomerId	PK <input checked="" type="checkbox"/>	Unique customer ID	<input type="checkbox"/>	long	
FirstName	PK <input type="checkbox"/>	Customer first name	<input type="checkbox"/>	string	256
LastName	PK <input type="checkbox"/>	Customer last name	<input checked="" type="checkbox"/>	string	256
EmailAddress	PK <input type="checkbox"/>	Customer email	<input type="checkbox"/>	string	256
Phone	PK <input type="checkbox"/>	Customer phone	<input checked="" type="checkbox"/>	string	256

7. When you've added all of the columns, publish the database again to save the changes.

8. In the Data pane on the left, switch back to the **Workspace** tab so you can see the **RetailDB** lake database. Then expand it and refresh its **Tables** folder to see the newly created **Customer** table.

Previous Exercise Next Exercise >

Page: 6/14

Synapse Analytics > synapsejhnxouv Search

Synapse live Validate all Publish all 1

RetailDB files

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.

Customer

Column_1 PK

FirstName

LastName

EmailAddress

Phone

See less ^

General Columns Relationships

Filter by keyword + Column Clone Convert type Delete

Name	Keys	Description	Nullability	Data type	Format / Length
Column_1	<input checked="" type="checkbox"/> PK	Unique customer ID	<input type="checkbox"/> Null	long	
FirstName	<input type="checkbox"/> PK	Customer first name	<input type="checkbox"/> Null	string	256
LastName	<input type="checkbox"/> PK	Customer last name	<input checked="" type="checkbox"/> Null	string	256
EmailAddress	<input type="checkbox"/> PK	Customer email	<input type="checkbox"/> Null	string	256
Phone	<input type="checkbox"/> PK	Customer phone	<input checked="" type="checkbox"/> Null	string	256

5:56 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabManual?la... Microsoft Azure Portal | Microsoft synapsejhnxouv - Microsoft Azure synapsejhnxouv - Azure Synaps

Instance ID:6746314 Xtremlabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

1. In the main pane, switch back to the **files** tab, which contains the file system with the **RetailDB** folder. Then open the **RetailDB** folder and create a new folder named **Customer** in it. This is where the **Customer** table will get its data.

2. Open the new **Customer** folder, which should be empty.

3. Download the **customer.csv** data file from <https://raw.githubusercontent.com/MicrosoftLearning/dp-203-azure-data-engineer/master/Allfiles/labs/04/data/customer.csv> and save it in a folder on your lab computer (it doesn't matter where). Then in the **Customer** folder in Synapse Explorer, use the **Upload** button to upload the **customer.csv** file to the **RetailDB/Customer** folder in your data lake.

Note: In a real production scenario, you would probably create a pipeline to ingest data into the folder for the table data. We're uploading it directly in the Synapse Studio user interface in this exercise for expediency.

4. In the **Data** pane on the left, on the **Workspace** tab, in the ... menu for the Customer table, select **New SQL script > Select TOP 100 rows**. Then, in the new SQL script 1 pane that has opened, ensure that the **Built-in** SQL pool is connected, and use the **Run** button to run the SQL code. The results should include first 100 rows from the **Customer** table, based on the data stored in the underlying folder in the data lake.

5. Close the **SQL script 1** tab, discarding your changes.

Previous Exercise Next Exercise >

Page: 7/14

Support

Search

Synapse Analytics > synapsejhnxouv Search

Synapse live Validate all Publish all 1

RetailDB files SQL script 1

Run Undo Publish Query plan Connect to Built-in ...

1 SELECT TOP (100) [Column_1]
2 ,[FirstName]
3 ,[LastName]
4 ,[EmailAddress]
5 ,[Phone]
6 FROM [RetailDB].[dbo].[Customer]

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

Size 111 bytes

Results settings per query ⓘ

First 5000 rows (default)
 All rows

Results Messages

View Table Chart Export results

Search

Column_1	FirstName	LastName	EmailAddress	Phone
1	Orlando	Gee	orlando0@adv...	245-555-0173

00:00:09 Query executed successfully.

6:31 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabManual?labInstanceId=...

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

Product name
IntroductionDate
ActualAbandonmentDate
ProductGrossWeight
ItemSku

On the toolbar in the **Columns** pane, select **Delete** to remove the selected columns. This should leave you with the following columns:

Name	Keys	Description	Nullability	Data type	Format / Length
ProductId	PK <input checked="" type="checkbox"/>	The unique identifier of a Product.	<input type="checkbox"/>	long	
ProductName	PK <input type="checkbox"/>	The name of the Product...	<input checked="" type="checkbox"/>	string	128
IntroductionDate	PK <input type="checkbox"/>	The date that the Product was introduced for sale.	<input checked="" type="checkbox"/>	date	YYYY-MM-DD
ActualAbandonmentDate	PK <input type="checkbox"/>	The actual date that the marketing of the product was discontinued...	<input checked="" type="checkbox"/>	date	YYY-MM-DD
ProductGrossWeight	PK <input type="checkbox"/>	The gross product weight.	<input checked="" type="checkbox"/>	decimal	18,8
ItemSku	PK <input type="checkbox"/>	The Stock Keeping Unit identifier...	<input checked="" type="checkbox"/>	string	20

Add a new column named **ListPrice** to the table as shown here:

Page: 8/14

Microsoft Azure Portal | Micros x synapsejhnxouv - Microsoft Az x synapsejhnxouv - Azure Synap x

web.azure-synapse.net/en/authoring/explore/workspace/databases/RetailDB?workspace=%2...

synapsejhnxouv

Search

Synapse live Validate all Publish all 1

RetailDB files

Table Map data Publish

ProductId PK
ProductName
IntroductionDate
ActualAbandonmentDate
ProductGrossWeight
ItemSku

General Columns Relationships

Filter by keyword

Column	Type	Description	Null	Length
ProductName	string	The name of the Product...	Null	128
IntroductionDate	date	The date that the Product was introduced for sale.	Null	YYYY-MM-DD
ActualAbandonmentDate	date	The actual date that the marketing of the product was discontinued...	Null	YYY-MM-DD
ProductGrossWeight	decimal	The gross product weight.	Null	18,8
ItemSku	string	The Stock Keeping Unit identifier...	Null	20

Partition column (0)

Support

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

ProductId	<input checked="" type="checkbox"/> PK	identifier of a Product.	<input type="checkbox"/>	long	
ProductName	<input checked="" type="checkbox"/> PK	The name of the Product...	<input checked="" type="checkbox"/>	string	128
IntroductionDate	<input checked="" type="checkbox"/> PK	The date that the Product was introduced for sale.	<input checked="" type="checkbox"/>	date	YYYY-MM-DD
ActualAbandonmentDate	<input checked="" type="checkbox"/> PK	The actual date that the marketing of the product was discontinued...	<input checked="" type="checkbox"/>	date	YYYY-MM-DD
ProductGrossWeight	<input checked="" type="checkbox"/> PK	The gross product weight.	<input checked="" type="checkbox"/>	decimal	18,8
ItemSku	<input checked="" type="checkbox"/> PK	The Stock Keeping Unit identifier...	<input checked="" type="checkbox"/>	string	20
ListPrice	<input checked="" type="checkbox"/> PK	The product price.	<input checked="" type="checkbox"/>	decimal	18,2

9. When you've modified the columns as shown above, publish the database again to save the changes.

10. In the **Data** pane on the left, switch back to the **Workspace** tab so you can see the **RetailDB** lake database. Then use the ... menu for its **Tables** folder to refresh the view and see the newly created **Product** table.

[Previous Exercise](#) [Next Exercise](#)

⊕ Page: 8/14 ⊕

[Support](#)

Microsoft Azure Portal | Microsoft Synapse JHNXOUV - Microsoft Azure Synapse JHNXOUV - Azure Synapse

web.azure-synapse.net/en/authoring/explore/workspace/databases/RetailDB?workspace=%2F%2Fsynapsejhnxouv

synapsejhnxouv

Search

Synapse live Validate all Publish all

RetailDB files

Table Map data Publish

Publishing completed Successfully published

ProductId PK

ProductName

IntroductionDate

ActualAbandonmentDate

ProductGrossWeight

ItemSku

ListPrice

See less

General Columns Relationships

+ Column Clone Convert type Delete

	<input type="checkbox"/> PK	The name of the Product,	<input checked="" type="checkbox"/> Null	abc string	128
	<input type="checkbox"/> PK	The date that the Product	<input checked="" type="checkbox"/> Null	date	YYYY-MM-DD
entDate	<input type="checkbox"/> PK	The actual date that the	<input checked="" type="checkbox"/> Null	date	YYYY-MM-DD
ght	<input type="checkbox"/> PK	The gross product	<input checked="" type="checkbox"/> Null	e ^x decimal	18,8
	<input type="checkbox"/> PK	The Stock Keeping Unit	<input checked="" type="checkbox"/> Null	abc string	20
	<input type="checkbox"/> PK	The product price	<input checked="" type="checkbox"/> Null	e ^x decimal	18,2

labs.xtremelabs.io/LabViewerConnection/DetachLabManual?la... Microsoft Azure Portal | Microsoft synapsejhnxouv - Microsoft Azure synapsejhnxouv - Azure Synaps...

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

62% Completed
Lab Time Left: 00:23:52

Lab Actions ▾

9. Analyze data in a lake database

Load data into the table's storage path:

- 1. In the main pane, switch back to the **files** tab, which contains the file system, and navigate to the **files/RetailDB** folder, which currently contains the **Customer** folder for the table you created previously.
- 2. In the **RetailDB** folder, create a new folder named **Product**. This is where the **Product** table will get its data.
- 3. Open the new **Product** folder, which should be empty.
- 4. Download the **product.csv** data file from <https://raw.githubusercontent.com/MicrosoftLearning/dp-203-azure-data-engineer/master/Allfiles/labs/04/data/product.csv> and save it in a folder on your lab computer (it doesn't matter where). Then in the **Product** folder in Synapse Explorer, use the **Upload** button to upload the **product.csv** file to the **RetailDB/Product** folder in your data lake.
- 5. In the **Data** pane on the left, on the **Workspace** tab, in the ... menu for the Product table, select **New SQL script > Select TOP 100 rows**. Then, in the new **SQL script 1** pane that has opened, ensure that the Built-in SQL pool is connected, and use the **Run** button to run the SQL code. The results should include first 100 rows from the **Product** table, based on the

⊕ Page: 9/14 ⊕

Synapse Analytics > synapsejhnxouv Search 3 6 ? 🔍 XLab-Vha-973@xtremelabs.us LAB DIRECTORY

Synapse live Validate all Publish all 1

RetailDB files SQL script 1

Run Undo Publish Query plan Connect to Built-in

```
1 SELECT TOP (100) [ProductId]
2 ,[ProductName]
3 ,[IntroductionDate]
4 ,[ActualAbandonmentDate]
5 ,[ProductGrossWeight]
6 ,[ItemSku]
7 ,[ListPrice]
8 FROM [RetailDB].[dbo].[Product]
```

Properties

General Related (0)

Name * SQL script 1

Description

Type .sql script

Size 167 bytes

Results settings per query ⓘ

First 5000 rows (default)

All rows

Results Messages

View Table Chart Export results

Search

ProductId	ProductName	IntroductionD...	ActualAbando...	ProductGross...
680	HL Road Frame...	2002-06-01	(NULL)	1016.04000000
706	HL Road Frame...	2002-06-01	(NULL)	1016.04000000

00:00:05 Query executed successfully.

Support

6:56 PM 11/3/2024 ENG IN

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

4. When the table has been created, note that it includes columns named **C1**, **C2**, and so on and that the data types have been inferred from the data in the folder. Modify the column definitions as follows:

Name	Keys	Description	Nullability	Data type	Format / Length
SalesOrderId	PK <input checked="" type="checkbox"/>	The unique identifier of an order.	<input type="checkbox"/>	long	
OrderDate	PK <input type="checkbox"/>	The date of the order.	<input type="checkbox"/>	timestamp	yyyy-MM-dd
LineItemId	PK <input checked="" type="checkbox"/>	The ID of an individual line item.	<input type="checkbox"/>	long	
CustomerId	PK <input type="checkbox"/>	The customer.	<input type="checkbox"/>	long	
ProductId	PK <input type="checkbox"/>	The product.	<input type="checkbox"/>	long	
Quantity	PK <input type="checkbox"/>	The order quantity.	<input type="checkbox"/>	long	

Note: The table contains a record for each individual item ordered, and includes a composite primary key comprised of **SalesOrderId** and **LineItemId**.

5. On the **Relationships** tab for the **SalesOrder** table, in the **+ Relationship** list, select **To table**, and then define the following relationship:

⊕ Page: 11/14 ⊕

Microsoft Azure Portal | synapsejhnxouv - Microsoft Edge | synapsejhnxouv - Azure

web.azure-synapse.net/en/authoring/explore/workspace/databases/RetailDB... 3 7 ?

synapsejhnxouv Search

Synapse live Validate all Publish all 1

RetailDB files

Table Map data Publish

SalesOrder

Column	Type	PK
SalesOrderId	long	PK
OrderDate	timestamp	
LineItemId	long	PK
CustomerId	long	
ProductId	long	
Quantity	long	

General Columns Relationships

Filter by keyword + Column Clone Convert type Delete

Standard column (6)

Column	Type	PK	Description	Null	Format / Length
SalesOrderId	long	PK	The unique identifier of	<input type="checkbox"/>	long
OrderDate	timestamp		The date of the order.	<input type="checkbox"/>	long
LineItemId	long	PK	line item.	<input type="checkbox"/>	long
CustomerId	long		The customer	<input type="checkbox"/>	long
ProductId	long		The product	<input type="checkbox"/>	long



labs.xtremelabs.io/LabViewerConnection/DetachLabManual?labInstanceGui... Instance ID:6746314

XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

Work with lake database tables:

Now that you have some tables in your database, you can use them to work with the underlying data.

Query tables using SQL

1. In Synapse Studio, select the **Develop** page.
In the **Develop** pane, in the + menu, select **SQL script**.

3. In the new **SQL script 1** pane, ensure the script is connected to the **Built-in** SQL pool and in the **User database** list, select **RetailDB**.

4. Enter the following SQL code:

Paste Content
Paste Content

```
SELECT o.SalesOrderID, c.EmailAddress, p.ProductName, o.Quantity
FROM SalesOrder AS o
JOIN Customer AS c ON o.CustomerId = c.CustomerId
JOIN Product AS p ON o.ProductId = p.ProductId
```

5. Use the **Run** button to run the SQL code.
The results show order details with customer and product information.

6. Close the **SQL script 1** pane, discarding your changes.

[Previous Exercise](#) [Next Exercise](#)

⊕ Page: 12/14 ⊕

Support

Microsoft Azure Portal | Microsoft Synapse Analytics - Microsoft Azure - synapsejhnxouv - Microsoft Azure Synapse - synapsejhnxouv - Azure Synapse

web.azure-synapse.net/en/authoring/analyze/sqlscripts/SQL%20script%201?workspace=%2F... 3 notifications

Synapse Analytics > synapse Search 3 notifications XLab-Vha-973@xtremelabs.us LAB DIRECTORY

Synapse live 1 notification

RetailDB files SQL script 1

Run Undo Publish Query plan Connect to Built-in

```
1 SELECT o.SalesOrderID, c.EmailAddress, p.ProductName, o.Quantity
2 FROM SalesOrder AS o
3 JOIN Customer AS c ON o.CustomerId = c.CustomerId
4 JOIN Product AS p ON o.ProductId = p.ProductId
```

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

Results Messages

View Table Chart Export results

Search

SalesOrderID	EmailAddress	ProductName	Quantity
71782	catherine0@adventure-works.com	Long-Sleeve Logo T-Shirt	8
71935	cory0@adventure-works.com	Road-350-W Yellow Jersey	2
71938	christopher1@adventure-works.com	Bike Wash - Disney Edition	6

00:00:06 Query executed successfully.

labs.xtremelabs.io/LabViewerConnection/DetachLabMa... Microsoft Azure Portal | Microsoft synapsejhnxouv - Microsoft Azure synapsejhnxouv - Azure Synaps +

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

89% Completed
Lab Time Left: 00:03:42

Lab Actions

13. Analyze data in a lake database

Insert data using Spark:

- 1. In the **Develop** pane, in the + menu, select **Notebook**.
- 2. In the new **Notebook 1** pane, attach the notebook to the **sparkxxxxxx*** Spark pool.
- 3. Enter the following code in the empty notebook cell:

```
%sql  
INSERT INTO `RetailDB`.`SalesOrder` VALUES (99999, CAST('2022-01-01' ASTimeStamp), 1, 6, 5, 1)
```
- 4. Use the ▶ button on the left of the cell to run it and wait for it to finish running. Note that it will take some time to start the Spark pool.
- 5. Use the + **Code** button to add a new cell to the notebook.
- 6. Enter the following code in the new cell:

```
%sql
```

Page: 13/14

Synapse Analytics > synapsejhnxouv

Synapse live Validate all Publish all 1

RetailDB files Notebook 1

Cancel all Undo Publish Outline Attach to sparkjhnxouv

Please wait a few minutes while your session starts.

1 %%sql
2 INSERT INTO `RetailDB`.`SalesOrder` VALUES (99999, CAST('2022-01-01' ASTimeStamp), 1, 6, 5, 1)

* 1 min 14 sec - Starting Apache Spark session

Properties

General Related (0)

Name * Notebook 1

Description

Type .ipynb notebook

Size 1,569 bytes

Notebook settings

Include cell output when saving

Enable unpublished notebook reference

Session

Configure session

Support

Search

7:16 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabMa... Microsoft Azure Portal | Microsoft synapsejhnxouv - Microsoft Azure Synapse synapsejhnxouv - Azure Synaps

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

AST('2022-01-01' AS TimeStamp), 1, 6, 5, 1)

4. Use the ▶ button on the left of the cell to run it and wait for it to finish running. Note that it will take some time to start the Spark pool.

5. Use the + **Code** button to add a new cell to the notebook.

6. Enter the following code in the new cell:

Paste Content
Paste Content

```
%sql  
SELECT * FROM `RetailDB`.`SalesOrder` WHERE SalesOrderId = 99999
```

7. Use the ▶ button on the left of the cell to run it and verify that a row for sales order 99999 was inserted into the **SalesOrder** table.

8. Close the **Notebook 1** pane, stopping the Spark session and discarding your changes.

Page: 13/14

Synapse Analytics > synapsejhnxouv

Synapse live Validate all Publish all 1

RetailDB files Notebook 1

Run all Undo Publish Outline Attach to sparkjhnxouv

Ready

[1] ✓ 3 min 17 sec in 2 min 46 sec 535 ms. Command executed in 31 sec 358 ms by XLab-Vha-973@xtremelabs.us

> Job execution Succeeded Spark 2 executors 8 cores

[View in monitoring](#) [Open Spark UI](#)

No data available

[2] 1 %sql
2 SELECT * FROM `RetailDB`.`SalesOrder` WHERE SalesOrderId = 99999

✓ 7 sec - Command executed in 6 sec 946 ms by XLab-Vha-973 on 7:19:12 PM, 11/03/24

> Job execution Succeeded Spark 2 executors 8 cores

[View in monitoring](#) [Open Spark UI](#)

View Table Chart Export results

SalesOrderId	OrderDate	LineItemId
99999	2022-01-01T00:00:00Z	1

Properties

General Related (0)

Name * Notebook 1

Description

Type .ipynb notebook

Size 3,473 bytes

Notebook settings

Include cell output when saving

Enable unpublished notebook reference

Session

[Configure session](#)

Support

7:19 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabMa... Microsoft Azure Portal | Microsoft Azure dp203-jhnxouv - Microsoft Azure

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

96% Completed
Lab Time Left: 00:00:28

Lab Actions

14. Analyze data in a lake database

Delete Azure resources:

- If you've finished exploring Azure Synapse Analytics, you should delete the resources you've created to avoid unnecessary Azure costs.
- 1. Close the Synapse Studio browser tab and return to the Azure portal.
- 2. On the Azure portal, on the Home page, select **Resource groups**.
- 3. Select the **dp203-xxxxxxxx** resource group for your Synapse Analytics workspace (not the managed resource group), and verify that it contains the Synapse workspace, storage account, and Spark pool for your workspace.
- 4. At the top of the **Overview** page for your resource group, select **Delete resource group**.
- 5. Enter the **dp203-xxxxxxxx** resource group name to confirm you want to delete it, and select **Delete**.

After a few minutes, your Azure Synapse workspace resource group and the managed workspace resource group associated with it will be deleted.

Page: 14/14

Support

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

Home > Resource groups > dp203-jhnxouv

Resource group

Search

Create Manage view Delete resource group Refresh Export to CSV Open query

... JSON View

Overview

Activity log Access control (IAM) Tags Resource visualizer Events Settings Cost Management Monitoring Automation Help

Essentials

Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Show hidden types No grouping List view

Name	Type	Location
datalakejhnxouv	Storage account	West US 2
sparkjhnxouv (synapsejhnxouv/sparkjhnxouv)	Apache Spark pool	West US 2
synapsejhnxouv	Synapse workspace	West US 2

Page 1 of 1

Give feedback

7:20 PM 11/3/2024 ENG IN

labs.xtremelabs.io/LabViewerConnection/DetachLabMa... Microsoft Azure Portal | Microsoft Azure dp203-jhnxouv - Microsoft Azure

Instance ID:6746314
XtremeLabs: DP-203T00-A-CEP [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database

96% Completed
Lab Time Left: EXPIRED

Lab Actions ▾

14. Analyze data in a lake database

Delete Azure resources:

- If you've finished exploring Azure Synapse Analytics, you should delete the resources you've created to avoid unnecessary Azure costs.
- Close the Synapse Studio browser tab and return to the Microsoft Azure Portal to delete the resource group.

Your lab is closing

5. Enter the **dp203-xxxxxx** resource group name to confirm you want to delete it, and select **Delete**.

After a few minutes, your Azure Synapse workspace resource group and the managed workspace resource group associated with it will be deleted.

Page: 14/14

Support

Microsoft Azure

Search resources, services, and docs (G+)

Copilot

dp203-jhnxouv Resource group

Search

Create Manage view Delete resource group Refresh Export to CSV Open query

Overview Essentials

Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Showing 1 to 2 of 2 records. Show hidden types No grouping List view

Name	Type	Location
sparkjhnxouv (synapsejhnxouv/sparkjhnxouv)	Apache Spark pool	West US 2
synapsejhnxouv	Synapse workspace	West US 2

Page 1 of 1

Give feedback

7:20 PM 11/3/2024 ENG IN

This training session brought to you by Southern Alberta Institute of Technology (XTP) Powered by



Welcome to XtremeLabs, Nitin Nitin



View Labs

Access Codes

History

Support

FAQs

Course DP-203T00-A-CEP: Data Engineering on Microsoft Azure

Go To Lesson

Lab Title: [DP-203T00-A-M01-CEP] Module 01: Explore Azure Synapse Analytics
Duration: 240 minutes
Status: **Completed and Relaunched**

Take Lab

Lab Title: [DP-203T00-A-M02-CEP] Module 02: Query files using a serverless SQL pool
Duration: 120 minutes
Status: **Completed and Relaunched**

Take Lab

Lab Title: [DP-203T00-A-M03-CEP] Module 03: Transform data using a serverless SQL pool
Duration: 120 minutes
Status: **Completed**

Take Lab

Lab Title: [DP-203T00-A-M04-CEP] Module 04: Analyze data in a lake database
Duration: 120 minutes
Status: **Completed**

Take Lab

Lab Title: [DP-203T00-A-M05-CEP] Module 05: Analyze data in a data lake with Spark
Duration: 120 minutes
Status: **Not Initiated**

Lab Title: [DP-203T00-A-M06-CEP] Module 06: Transform data using Spark in Synapse Analytics
Duration: 120 minutes
Status: **Not Initiated**



Search

