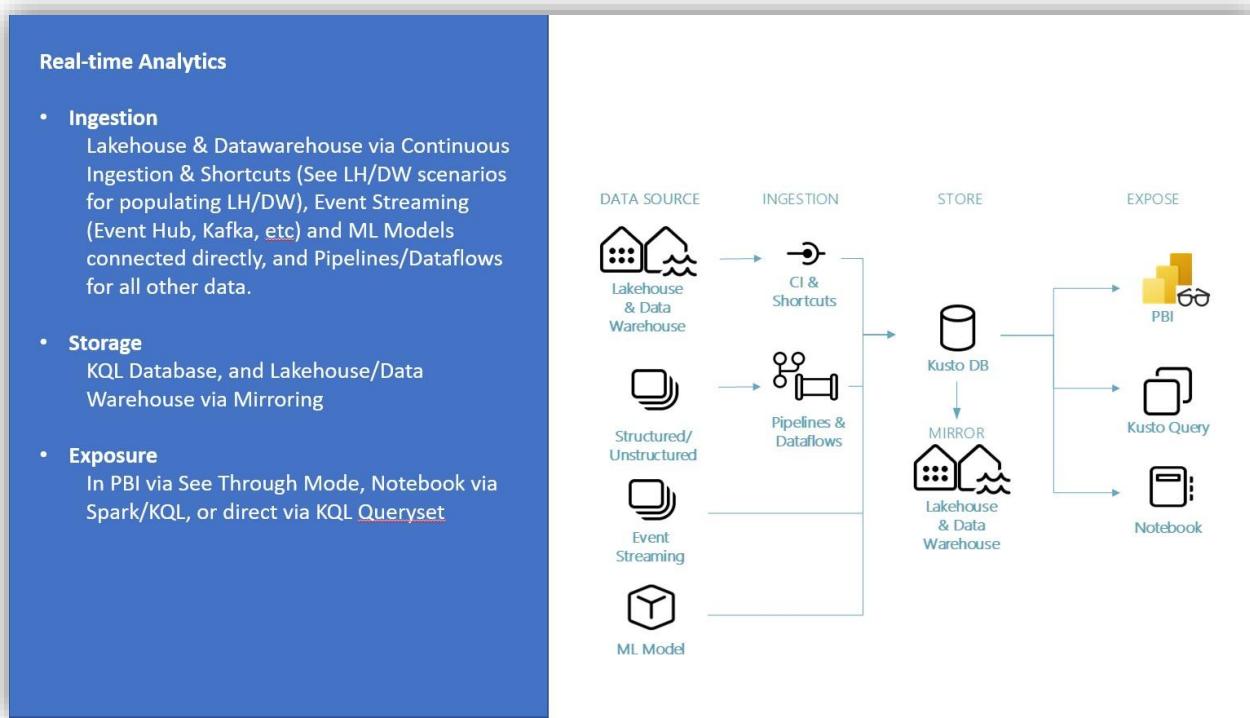


End-to-End Scenario Tutorial

Real-time Analytics

Published: July 2023

Updated : Nov 2024



Contents

Providing feedback.....	3
Introduction.....	5
Real-time Analytics.....	5
Scenario.....	6
Prerequisites.....	7
Module 1: Create a Fabric workspace.....	7
Module 2: Build your first Real-time Analytics solution in Fabric.....	9
Create a KQL Database.....	9
Create an Eventstream.....	11
Stream data from Eventstream to KQL Database	13
Explore data and build Power BI report.....	20
Module 3: Extending the real-time analytics solution	22
Get dimension data from Blob Storage.....	22
Query data.....	26
Explore data further in the KQL Queryset.....	27
Build Power BI report.....	30
Create OneLake shortcut	33
Module 4: Clean up resources	36

Introduction

What is Fabric?

Fabric provides a one-stop shop for all the analytical needs for every enterprise. It covers the complete spectrum of services including data movement, data lake, data engineering, data integration and data science, real time analytics, and business intelligence. With Fabric, there is no need to stitch together different services from multiple vendors. Instead, the customer enjoys an end-to-end, highly integrated, single comprehensive product that is easy to understand, onboard, create and operate. There is no other product on the market that offers the breadth, depth, and level of integration that Fabric offers. Additionally, Microsoft Purview is included by default in every tenant to meet compliance and governance needs.

To get an overview over the components and concepts of Fabric read [Fabric - Overview and Concepts](#).

The purpose of this tutorial

While many concepts in Fabric may be familiar to data and analytics professionals it can be challenging to apply those concepts in a new environment. This tutorial has been designed to walk step-by-step through an end-to-end scenario from data acquisition to data consumption to build a basic understanding of the Fabric UX, the various workloads and their integration points, and the Fabric professional and citizen developer experiences.

The tutorials are not intended to be a reference architecture, an exhaustive list of features and functionality, or a recommendation of specific best practices.

Real-time Analytics

Real-time Analytics is a portfolio of capabilities that provides an end-to-end analytics streaming solution across Fabric experiences. It supplies high velocity, low latency data analysis, and is optimized for time-series data, including automatic partitioning and indexing of any data format and structure, such as structured data, semi-structured (JSON), and free text.

Real-time Analytics delivers high performance when it comes to your increasing volume of data. It accommodates datasets as small as a few gigabytes or as large as several petabytes and allows you to explore data from different sources and a variety of data formats.

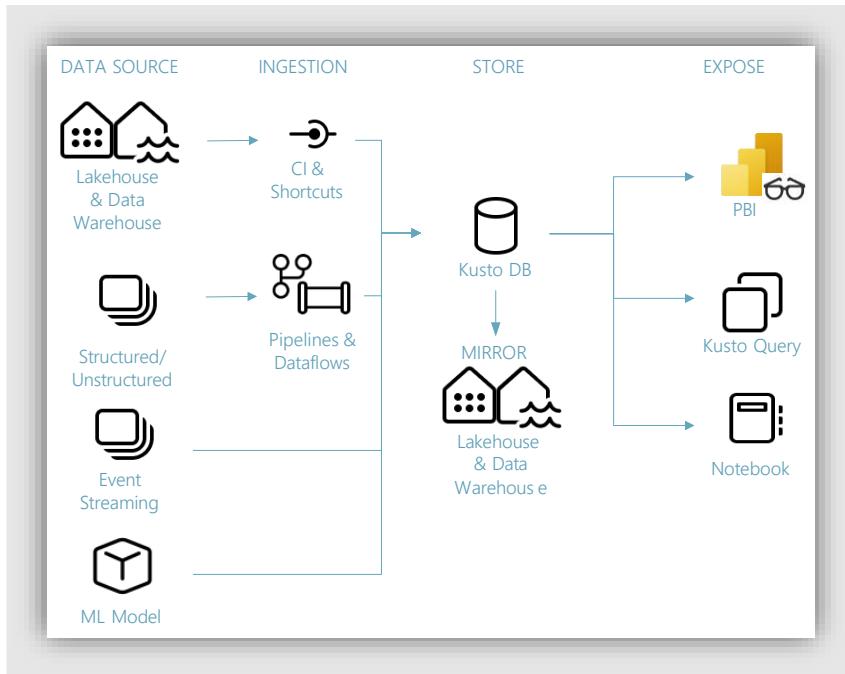


Figure 1: End-to-end Scenario - Real-time Analytics

Real-time Analytics includes integration with other Fabric experiences such as Lakehouse, Data Warehouse, Pipelines, Dataflows and Event Streaming on data sources and ingestion side. On the data exposition and consumption, Real-time analytics integrates with Power BI, and Notebooks.

You can use Real-time Analytics for a range of solutions, such as IoT analytics and log analytics, and in a number of scenarios including manufacturing operations, oil and gas, automotive, and more.

In this tutorial, you learn how to:

- Create a KQL Database
- Create Eventstream
- Stream data from Eventstream to KQL Database
- Check your data with sample queries
- Save queries as a KQL Queryset
- Create a Power BI report
- Create a OneLake shortcut

Scenario

This tutorial is based on a *sample on New York Yellow Taxi trip data*. The dataset contains trip records of New York's yellow taxis. The yellow taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. You'll use the streaming and query capabilities of Real-time Analytics to answer key questions about the trip statistics, taxi demand in the boroughs of New York and related insights.

Prerequisites

- Power BI Premium subscription. For more information, see [How to purchase Power BI Premium](#).
- Workspace

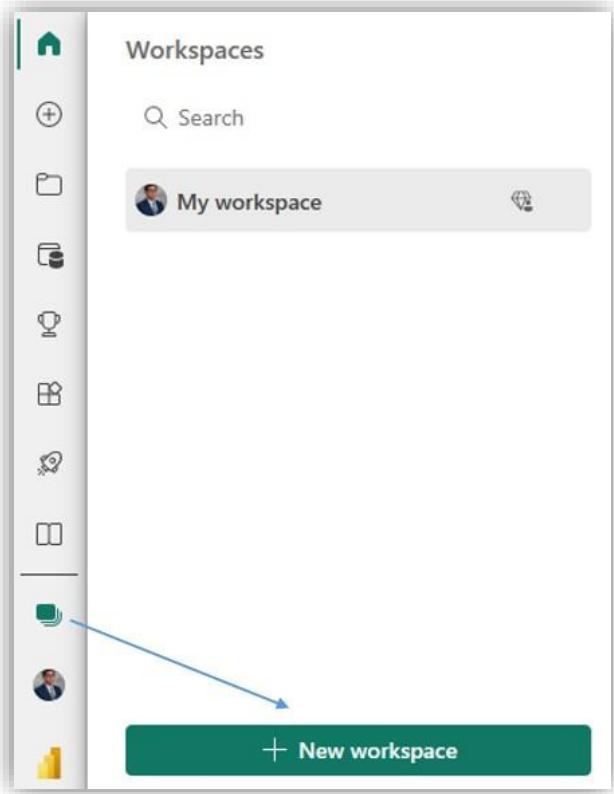
Module 1: Create a Fabric workspace

Before you can begin building the real-time analytics solution, you will need to enable Project Fabric on your Power BI tenant and create a workspace where you will build out the remainder of the tutorial. In this module, you will learn to:

- Create a Fabric workspace

In this step, you create a Fabric workspace in the Power BI service. The workspace will contain all the artifacts needed for real-time analytics including KQL Database, Eventstream, Power BI datasets, and reports.

1. Sign in to [Fabric](#).
2. Select **Workspaces > New Workspace**.



3. Fill out the **Create a workspace** form as follows:
 - a. **Name:** Enter *Fabric Real-time Analytics Tutorial*, and some characters for uniqueness.
 - b. **Description:** Optionally, enter a description for the workspace.

Create a workspace

Name *

This name is available

Description

Domain ⓘ

Learn more about workspace settings ⓘ

Workspace image



Upload

Reset

Advanced ^

Contact list * ⓘ

License mode ⓘ

Pro
Select Pro to use basic Power BI features and collaborate on reports, dashboards, and scorecards. To access a Pro workspace, users need Pro per-user licenses. [Learn more ⓘ](#)

Trial
Select the free trial per-user license to try all the new features and experiences in Microsoft Fabric for 60 days. A Microsoft Fabric trial license allows users to create Microsoft Fabric items and collaborate with others in a Microsoft Fabric trial capacity. Explore new capabilities in Power BI, Data Factory, Data Engineering, and Real-Time Analytics, among others. [Learn more ⓘ](#)

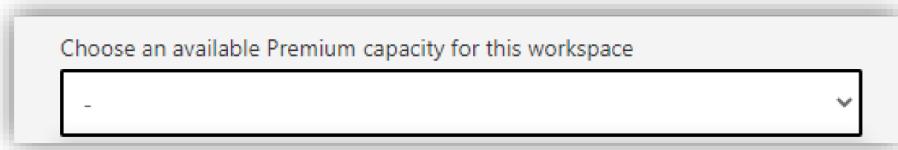
Premium per-user
Select Premium per-user to collaborate using Power BI Premium features, including paginated reports, dataflows, and datamarts. To collaborate and share content in a Premium per-user workspace, users need Premium per-user licenses. [Learn more ⓘ](#)

Premium capacity
Select premium capacity if the workspace will be hosted in a premium capacity. When you share, collaborate on, and distribute Power BI and Microsoft Fabric content, users in the viewer role can access this content without needing a Pro or Premium per-user license. [Learn more ⓘ](#)

Embedded
Select embedded if the workspace will be hosted in an Azure embedded capacity. ISVs and developers use Power BI Embedded to embed visuals and analytics in their applications. [Learn more ⓘ](#)

Apply **Cancel**

4. Expand the **Advanced** section.
5. Choose **Fabric capacity** in the **License Mode** section.
6. Choose a Fabric capacity you have access to.



7. Select **Apply**. The workspace will be created and opened.

Module 2: Build your first Real-time Analytics solution in Fabric

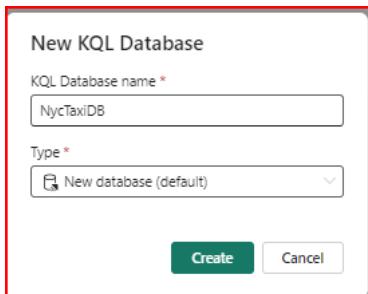
The intent of this module is to quickly build end to end journey of building a real-time analytics solution, ingesting streaming data from Eventstream and then using the KQL Database for creating a real-time refreshing Power BI report.

Create an Eventhouse

1. In the upper left corner of the Fabric Workspace home page, select **New item > Eventhouse**

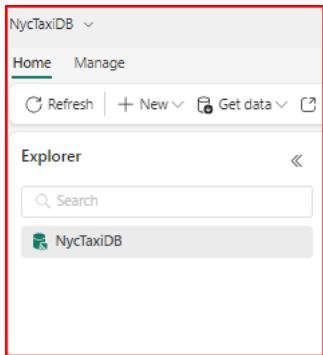
The screenshot shows the Fabric Real-time Analytics Tutorial workspace. On the left, there's a sidebar with 'New item', 'New folder', and 'Upload' buttons. The main area has a title 'New item' and a sub-section 'Select an item type'. A button 'All items' is highlighted with a red box. Below this, there's a section titled 'Store data' with a sub-section 'Organize, query, and store your ingested data in an easily retrievable format.' Six options are listed: 'Datamart (preview)', 'Eventhouse', 'Lakehouse', 'Sample warehouse', 'Semantic model', and 'SQL Database (preview)'. The 'Eventhouse' option is also highlighted with a red box. Each option has a description and a star icon below it.

2. Click on “+” within the new Eventhouse. On the **New KQL Database** dialog, enter a **unique name**.
3. Select **Create**.



4. When provisioning is complete the KQL database editor landing page will be shown.

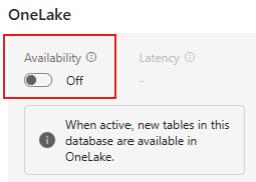
5. Select the **Database** in the Object tree.



6. Select **Explore your data**



7. Enable availability in Onelake (replace **NycTaxiDB** with your database's name).
- In the **Database details** card when NycTaxiDB is selected, toggle **Availability** to On.



In the pop-up window, click on the button **Turn On**.

Turn on OneLake availability

Only new data will be added to new and existing tables.

Apply to existing tables

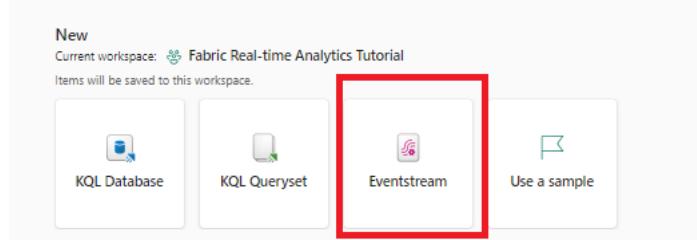
Turn on **Cancel**

Create an Eventstream

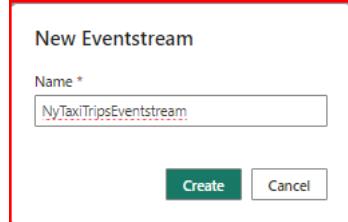
1. Return to the Real-Time Analytics home page. The **Home** icon directs you to the home page of the experience you're currently using.



2. In the **Real-time Analytics** section, select **Eventstream**

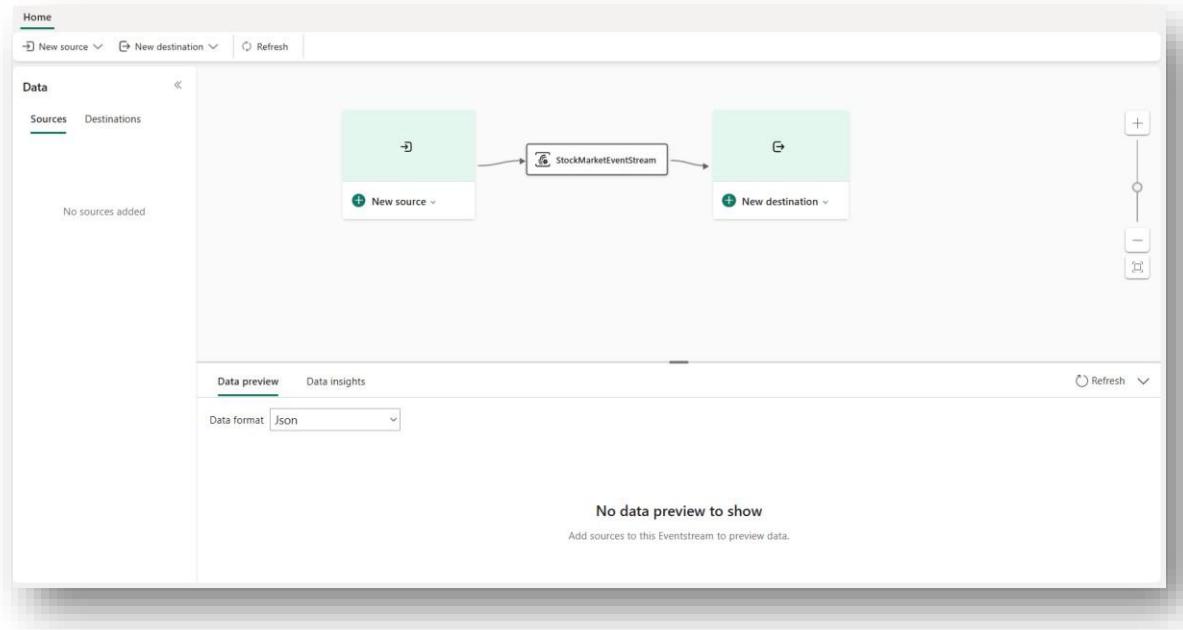


3. On the New Eventstream dialog, enter NyTaxiTripsEventstream as the name.



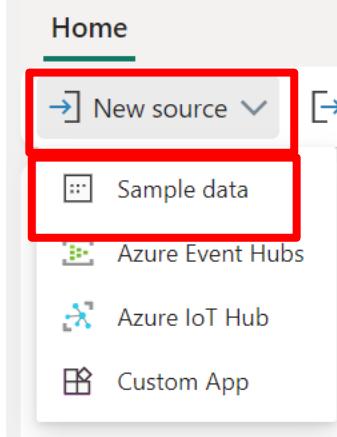
4. Select **Create**.

- When provisioning is complete, Eventstream landing page will be shown.



Stream data from Eventstream to KQL Database

- In the Eventstream authoring area, select **New source** and choose **Sample data**.



- Enter **nytaxitripsdatasource** as the Source Name on the right window, choose **Yellow Taxi(high data-rate)** and click on **Add**.

Sample data

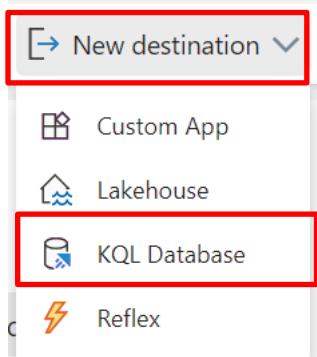
Source name *

nytaxitripsdatasource

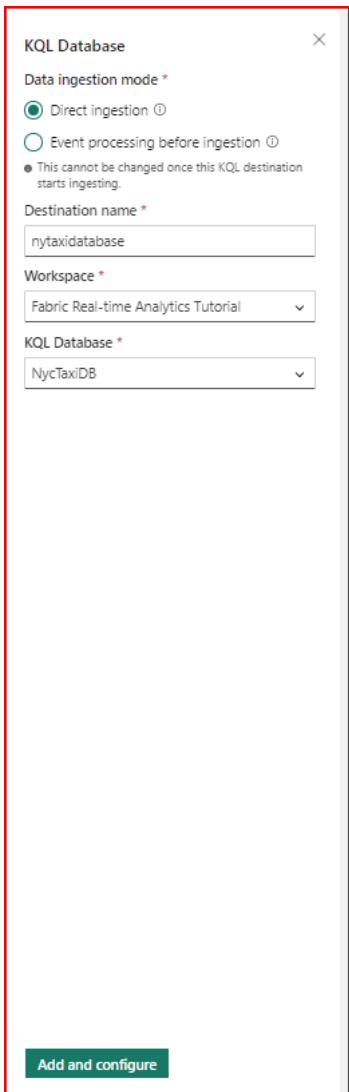
Sample data *

Yellow Taxi (high data-rate)

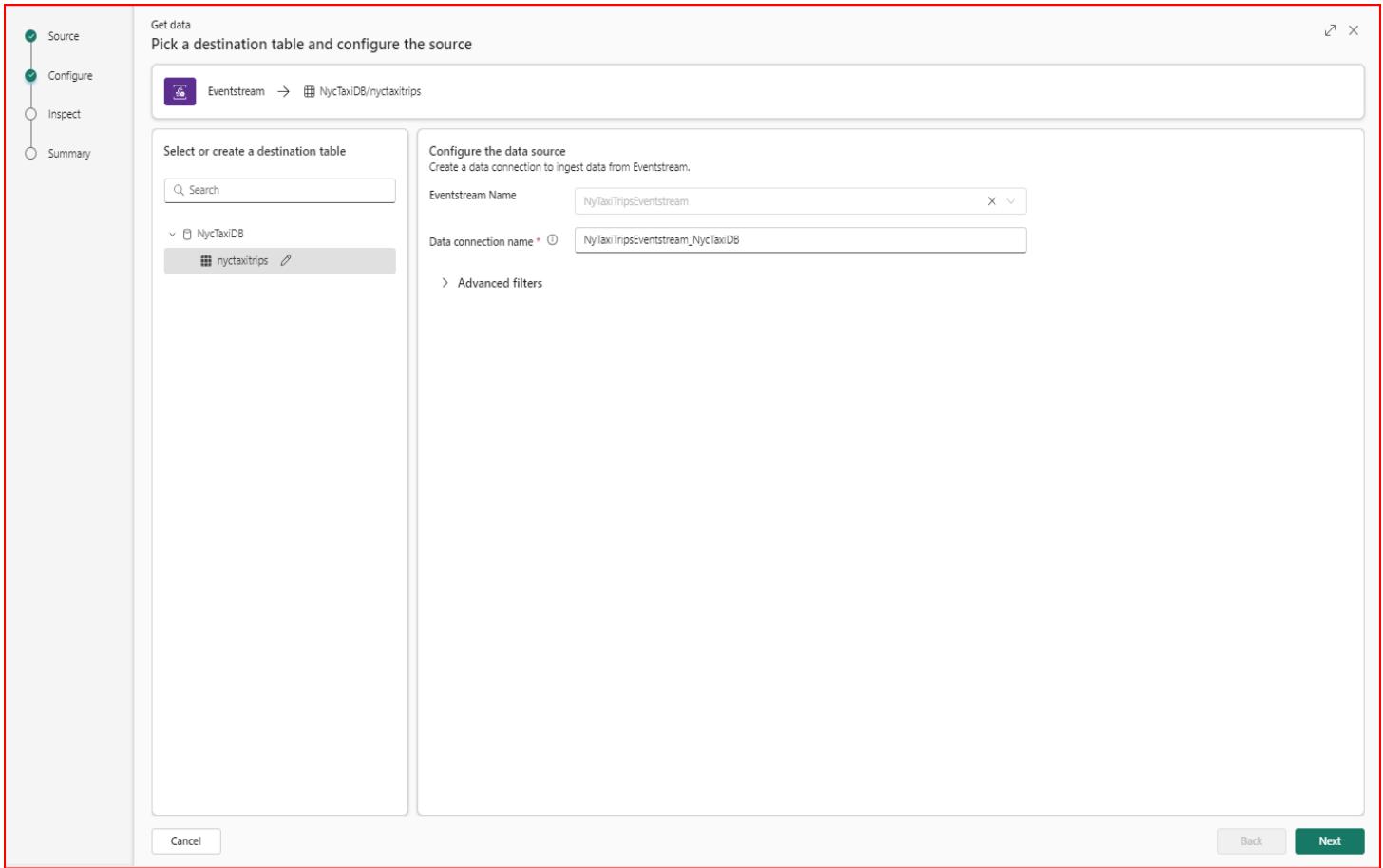
3. In the Eventstream authoring area, select **New destination** and choose **KQL Database**.



4. Select Data ingestion Mode as **Direct Ingestion**, Enter **nytaxidatabase** as the destination name, choose your Fabric workspace from the Workspace dropdown and then choose your KQL Database that you created above and Click on **Add and Configure**



5. In the Get Data Screen , select **New table** and enter **nyctaxitrips** as the table name. Then press on **Next**.



6. In the **Inspect** tab, choose **JSON** as the Data format dropdown.

The screenshot shows the 'Inspect' tab interface. On the left, there are four tabs: 'Source', 'Configure', 'Inspect' (which is selected and highlighted in green), and 'Summary'. The main area is titled 'Get data' and 'Inspect the data'. It shows an 'Eventstream' source connected to 'NyCTaxiDB/nyctaxitrips'. Below this, there are two tabs: 'Preview data' (selected) and 'Command viewer'. A message indicates 'Data sample found.' and '23 events found. 23 events match the selected settings. Open for more details.' The data preview table has columns: VendorID (long), tpep_pickup_datetime (datetime), tpep_dropoff_datetime (datetime), passenger_count (real), trip_distance (real), RatecodeID (real), store_and_fwd_flag (string), PULocationID (long), DOLocationID (long), and payment_type (long). The data is as follows:

VendorID (long)	tpep_pickup_datetime (datetime)	tpep_dropoff_datetime (datetime)	passenger_count (real)	trip_distance (real)	RatecodeID (real)	store_and_fwd_flag (string)	PULocationID (long)	DOLocationID (long)	payment_type (long)
2	2022-06-01 00:00:58.0000	2022-06-01 00:38:56.0000	1	19.79	2	N	132	62	1
2	2022-06-01 00:01:54.0000	2022-06-01 00:23:44.0000	1	6.48	1	N	79	74	1
1	2022-06-01 00:02:38.0000	2022-06-01 00:12:37.0000	0	2.1	1	N	238	41	1
1	2022-06-01 00:02:51.0000	2022-06-01 00:05:55.0000	2	0.6	1	N	140	140	3
2	2022-06-01 00:03:24.0000	2022-06-01 00:42:21.0000	1	18.44	2	N	132	68	1
1	2022-06-01 00:03:43.0000	2022-06-01 00:06:29.0000	1	0.5	1	N	113	79	2

On the right, a dropdown menu for 'Format' is open, showing options: JSON (selected), CSV, TSV, SCSV, TSVE, SOHSV, PSV, and TX1. Buttons for 'Edit columns' and 'Advanced' are also present.

7. After choosing JSON as the Data format, data preview will refresh and show the data in strongly typed columns.

This screenshot shows the same 'Inspect' tab interface as the previous one, but the data preview now displays strongly typed columns. The columns are explicitly named with their data types: VendorID (long), tpep_pickup_datetime (datetime), tpep_dropoff_datetime (datetime), passenger_count (real), trip_distance (real), RatecodeID (real), store_and_fwd_flag (string), PULocationID (long), DOLocationID (long), and payment_type (long). The data is identical to the previous preview.

8. In this step, we will change data types of multiple columns.

- a. Select **Edit Columns** on the top of screen For the column **VendorID** column name, **Change data type**, and then choose **int**.

- b. Change datatype to **long** for the columns passenger_count, PULocationID, DOLocationID and payment_type, if they are not already.
- c. Change datatype to **real** for the following columns: extra, mta_tax, tolls_amount, improvement_surcharge, congestion_charge, airport_fee, trip_distance, fare_amount, tip_amount, total_amount, if they are not already.

After changing the datatypes of the above columns , Click on **Apply**

The screenshot shows a 'Edit columns' interface with a sidebar on the left containing 'Source', 'Configure', 'Inspect' (which is selected), and 'Summary' buttons. The main area displays a table of columns with the following data:

Column name	Type	Source	Mapping transformation	Sample data
VendorID	int	VendorID		2
tpep_pickup_datetime	datetime	tpep_pickup_datetime		2022-06-01T00:59:15Z
tpep_dropoff_datetime	datetime	tpep_dropoff_datetime		2022-06-01T01:06:50Z
passenger_count	long	passenger_count		1
trip_distance	real	trip_distance		1.93
RatecodeID	real	RatecodeID		1
store_and_fwd_flag	string	store_and_fwd_flag		N
PULocationID	long	PULocationID		90
DOLocationID	long	DOLocationID		231
payment_type	long	payment_type		1
fare_amount	real	fare_amount		8
extra	real	extra		0.5
mta_tax	real	mta_tax		0.5
tip_amount	real	tip_amount		2.95
tolls_amount	real	tolls_amount		6.55
improvement_surcharge	real	improvement_surcharge		0.3
total_amount	real	total_amount		14.75
congestion_surcharge	real	congestion_surcharge		2.5

At the bottom right of the interface are 'Back' and 'Apply' buttons.

9. Click on Finish

Get data

Source Configure Inspect Summary

Inspect the data

Eventstream → NycTaxiDB/nyctaxitrips

Preview data Command viewer Format: JSON ▾ Edit columns Advanced ▾

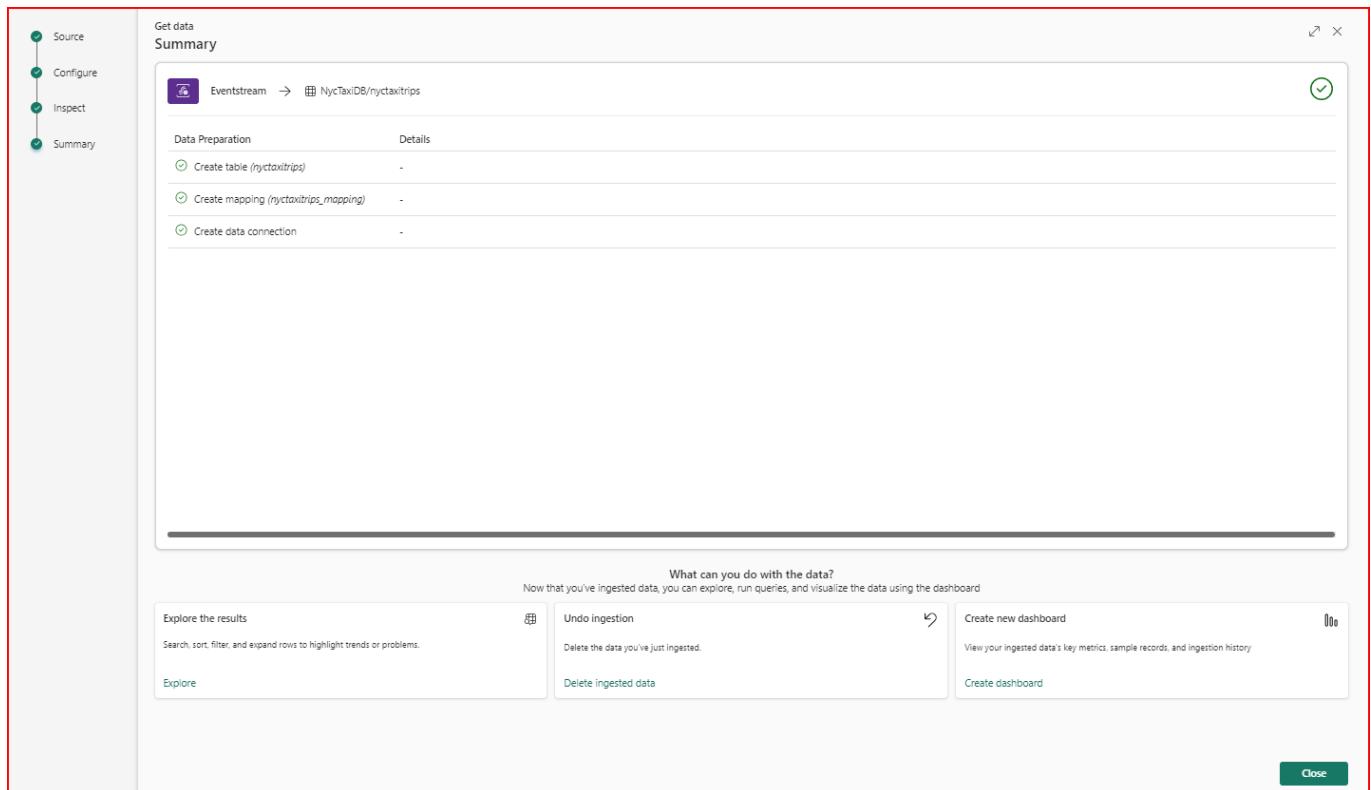
Data sample found. Fetch more data Discard and fetch new data

23 events found. 23 events match the selected settings. Open for more details.

VendorID (int)	tpep_pickup_datetime (datetime)	tpep_dropoff_datetime (datetime)	passenger_count (long)	trip_distance (real)	RatecodeID (real)	store_and_fwd_flag (string)	PULocationID (long)	DOLocationID (long)	payment_type (long)
2	2022-06-01 00:20:25.0000	2022-06-01 00:26:29.0000	2.04	1	N	162	237	1	
2	2022-06-01 00:20:25.0000	2022-06-01 00:44:11.0000	6.67	1	N	234	89	1	
2	2022-06-01 00:22:01.0000	2022-06-01 00:28:18.0000	2.36	1	N	142	166	1	
2	2022-06-01 00:22:37.0000	2022-06-01 00:32:55.0000	3.09	1	N	148	209	1	
1	2022-06-01 00:23:02.0000	2022-06-01 00:48:28.0000	7	1	N	249	61	1	
2	2022-06-01 00:26:20.0000	2022-06-01 00:35:52.0000	2.57	1	N	144	66	1	
2	2022-06-01 00:27:09.0000	2022-06-01 00:31:58.0000	1.41	1	N	137	79	1	
2	2022-06-01 00:28:18.0000	2022-06-01 00:31:25.0000	3.83	1	N	113	163	1	
1	2022-06-01 00:28:24.0000	2022-06-01 00:30:45.0000	0.4	1	N	230	230	1	
2	2022-06-01 00:31:16.0000	2022-06-01 00:40:22.0000	1.83	1	N	151	152	1	
2	2022-06-01 00:33:50.0000	2022-06-01 00:48:51.0000	4.78	1	Y	146	79	1	
1	2022-06-01 00:38:30.0000	2022-06-01 00:42:06.0000	1.7	1	N	239	230	1	
2	2022-06-01 00:36:18.0000	2022-06-01 00:40:27.0000	0.52	1	N	263	262	1	
2	2022-06-01 00:38:54.0000	2022-06-01 01:03:38.0000	4.59	1	N	231	230	1	
2	2022-06-01 00:39:51.0000	2022-06-01 01:02:57.0000	16.54	2	N	132	233	1	
1	2022-06-01 00:46:31.0000	2022-06-01 01:08:27.0000	6.5	1	N	234	7	2	
1	2022-06-01 00:46:49.0000	2022-06-01 01:09:23.0000	4.2	1	N	230	7	2	
2	2022-06-01 00:50:10.0000	2022-06-01 01:07:11.0000	6.35	1	N	79	74	1	
2	2022-06-01 00:56:37.0000	2022-06-01 01:05:01.0000	2.31	1	N	163	262	1	
2	2022-06-01 00:57:16.0000	2022-06-01 01:12:10.0000	9.19	1	N	264	107	1	
1	2022-06-01 00:58:37.0000	2022-06-01 01:11:55.0000	3.1	1	N	158	233	1	
2	2022-06-01 00:59:15.0000	2022-06-01 01:06:50.0000	1.93	1	N	90	231	1	

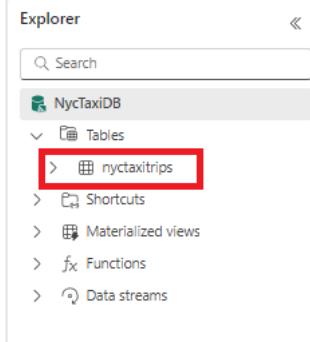
Cancel Back Finish

10. In the **Summary Tab, Continuous ingestion from Event Stream established** window, all steps will be marked with green check marks when the data connection is successfully created. Once click on **Publish**, the data from Eventstream will begin streaming automatically into your table. Click on **Close**

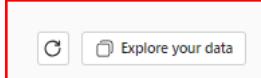


Explore data and build Power BI report

1. Navigate to your Fabric workspace, select your KQL Database.
2. In the object tree, select the table **nytaxitrips**



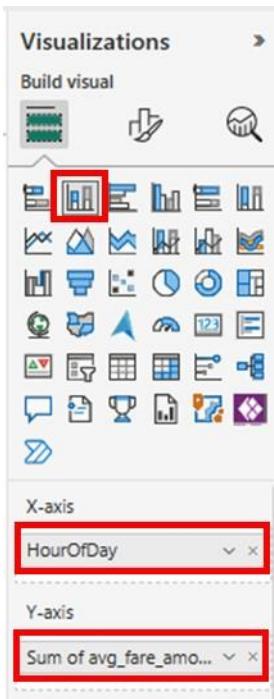
3. In the top right corner, select **Explore your data**.



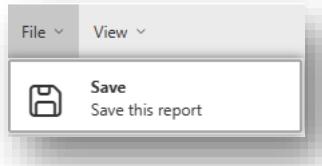
4. Paste the following query in the query editor and select **Run**

```
//Calculate average fare amount by the hour of the day.  
nytaxitrips  
| summarize avg(fare_amount) by HourOfDay = hourofday(tpep_dropoff_datetime)
```

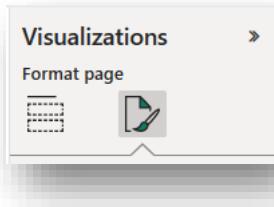
5. Select **Power BI** button. Empty PowerBI report editing window will open.
6. Select **Stacked Column Chart** in the Visualizations pane. From the **Data** pane, drag **HourOfDay** field to X-axis and **avg_fare_amount** to Y-axis



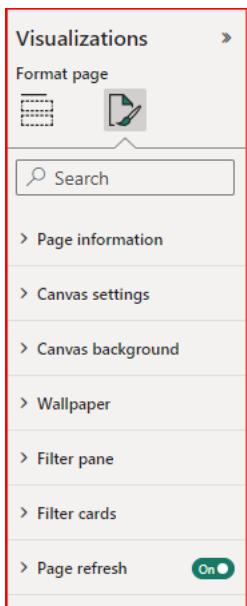
7. Select **File** menu and then **Save** in the top left corner



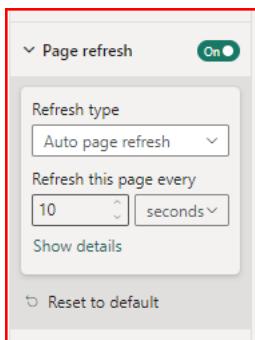
8. Enter **nyctaxitripstats** in the **Name your file in Power BI** field, choose your workspace.
9. Once the report is saved, click on the link **Open the file in Power BI to view, edit, and get a shareable link.**
10. Click on **Edit** button to edit the Power BI report.
11. Click on an Empty Space in the Canvas , Choose **Format page**



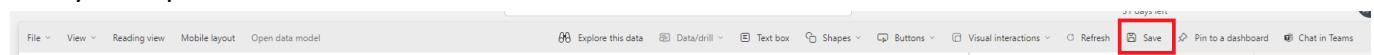
12. Toggle **Page Refresh** to **On**



13. Expand **Page Refresh** and set the refresh interval to 10 seconds. Please note: Refresh interval will be limited by the Admin interval. Refresh interval can only be greater than or equal to the Admin interval.



14. Save your report.



With this tutorial, you have now built an auto-refreshing Power BI report that is querying streaming data arriving in KQL Database from Eventstream.

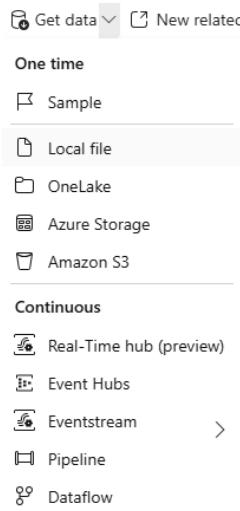
Module 3: Extending the real-time analytics solution

Get dimension data from Azure Storage

In this module, you are going to ingest Location available in a Azure storage. This data contains additional information on the pick-up locations and drop-off locations used in the trips dataset. Real-time analytics reads and ingests data directly from the blob storage without requiring any other intermediary service.

1. Navigate to KQL Database **NycTaxiDB**

2. From within the KQL Database, select **Get Data > Local File**.



Destination tab

In the **Destination tab**, **Database** is auto populated with the name of the selected KQL database.

3. Under **Table**, make sure that **New table** is selected, and enter ***locations*** as your table name.

The screenshot shows the 'Get data' configuration interface. On the left, a sidebar has 'Source' selected. The main area is titled 'Get data' and 'Pick a destination table and configure the source'. It shows 'Local file' selected as the source, pointing to 'NyTaxiDB/locations'. On the left, under 'Select or create a destination table', 'NyTaxiDB' is expanded, showing 'locations' selected. On the right, under 'Add up to 1,000 files', a table lists a single file: 'ny-yellow-taxi-location-info.csv' with status 'OK', size '15.5 KB', and a delete icon. At the bottom are 'Cancel', 'Back', and 'Next' buttons, with 'Next' being green.

- Instructor: Please provide the file in advance available under the data folder of the labFile : **ny-yellow-taxi-location-info.csv** click on + icon and click **Next** :

4. In the **Inspect tab**, click on **Finish**

- Source
- Configure
- Inspect
- Summary

Get data ↔ X

Inspect the data

Local file → NycTaxiDB/locations

Preview data Command viewer Edit columns ...

LocationID (long)	Borough (string)	Zone (string)	service_zone (string)	Latitude (real)	Longitude (real)
	Borough	Zone	service_zone		
1	EWR	Newark Airport	EWR	40.6895	-74.1745
2	Queens	Jamaica Bay	Boro Zone	40.6097	-73.8233
3	Bronx	Allerton/Pelham Gardens	Boro Zone	40.8667	-73.8478
4	Manhattan	Alphabet City	Yellow Zone	40.7258	-73.9795
5	Staten Island	Arden Heights	Boro Zone	40.5563	-74.1737
6	Staten Island	Arrochar/Fort Wadsworth	Boro Zone	40.5964	-74.0671
7	Queens	Astoria	Boro Zone	40.7644	-73.9235
8	Queens	Astoria Park	Boro Zone	40.779	-73.9229
9	Queens	Auburndale	Boro Zone	40.757	-73.7898
10	Queens	Baisley Park	Boro Zone	40.6781	-73.7855
11	Brooklyn	Bath Beach	Boro Zone	40.594224	-74.006131
12	Manhattan	Battery Park	Yellow Zone	40.703282	-74.017027
13	Manhattan	Battery Park City	Yellow Zone	40.711932	-74.016869
14	Brooklyn	Bay Ridge	Boro Zone	40.625801	-74.030621
15	Queens	Bay Terrace/Fort Totten	Boro Zone	40.791211	-73.776978
16	Queens	Bayside	Boro Zone	40.76314	-73.77125
17	Brooklyn	Bedford	Boro Zone	40.687216	-73.941776
18	Bronx	Bedford Park	Boro Zone	40.8701	-73.885691
19	Queens	Bellerose	Boro Zone	40.733014	-73.722938

Cancel Back Finish

5. Summary tab

In the **Data ingestion is in progress** window, all steps will be marked with green check marks when the data has been successfully ingested. The data from Azure Storage will begin streaming automatically into your table.

The screenshot shows the 'Get data' interface in the Azure Data Explorer. On the left, a sidebar lists 'Source', 'Configure', 'Inspect', and 'Summary'. The 'Summary' tab is selected, displaying a summary of the ingestion process.

Summary

Local file → NycTaxiDB/locations

1 blobs, 1 succeeded, 0 failed (checkmark)

Data Preparation		Details
(checkmark) Create table (<i>locations</i>)	-	
(checkmark) Create mapping (<i>locations_mapping</i>)	-	
(checkmark) Data queuing	-	

Blob name	Status	Details
https://uk8virtualdms.blob.core.windows.net/tr...	(checkmark) Successfully ingested	-

What can you do with the data?

Now that you've ingested data, you can explore, run queries, and visualize the data using the dashboard

Explore the results grid icon
 Search, sort, filter, and expand rows to highlight trends or problems.
[Explore](#)

Undo ingestion undo icon
 Delete the data you've just ingested.
[Delete ingested data](#)

Create new dashboard bar chart icon
 View your ingested data's key metrics, sample records, and ingestion history.
[Create dashboard](#)

Close

Now that you've got data in your database, click on **Close**. You're going to check your data with sample queries.

Query data

In the following step, you'll use the advanced data analysis capabilities of Kusto Query language to query your telemetry data.

1. Select **Explore your data** on the right-hand side of your database editor.

The screenshot shows the Azure Data Explorer interface for the 'nyctaxitrips' table. It displays various metrics such as Row count (62,900), Last ingestion (Today, this minute), and Size (3.2 MB Compressed size, 10.97 MB Original size). On the right, there's a 'Mappings' section with a clipboard icon and a note 'No items to show'. The 'Explore your data' button is highlighted with a red box.

Note: The numbers in the screen capture above may look different in your database editor page.

2. Let's take a look at the data itself. Paste the following query in **Explore your data** window to take 10 random records from your data.

Query to be pasted →

*nyctaxitrips
/ take 10*

The screenshot shows the 'Explore your data' window with the following Kusto Query Language (KQL) query:

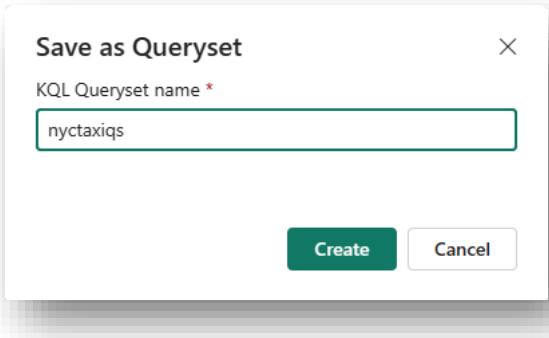
```
1 //*****  
2 // Here are two articles to help you get started with KQL:  
3 // KQL reference guide - https://aka.ms/KQLguide  
4 // SQL - KQL conversions - https://aka.ms/sqlcheatsheet  
5 //*****  
6  
7 // Use "take" to view a sample number of records in the table and check the data.  
8 nyctaxitrips  
9 | take 10  
10
```

3. Select **Run**.

Table 1 Stats

VendorID	tpep_pickup_datetime	tpep_dropoff_datetime	passenger_count	trip_distance	RatecodeID	store_and_fwd_flag
> 2	2022-06-06 18:00:04.0000	2022-06-06 18:09:47.0000	1	2.21	1.0	N
> 2	2022-06-06 18:05:30.0000	2022-06-06 18:25:48.0000	1	3.46	1.0	N
> 2	2022-06-06 18:26:37.0000	2022-06-06 18:33:04.0000	1	1.02	1.0	N
> 2	2022-06-06 18:33:18.0000	2022-06-06 18:37:46.0000	2	0.93	1.0	N
> 2	2022-06-06 18:35:16.0000	2022-06-06 18:55:46.0000	1	6.42	1.0	N
> 2	2022-06-06 18:37:57.0000	2022-06-06 18:53:17.0000	1	9.95	1.0	N
> 2	2022-06-06 18:41:01.0000	2022-06-06 18:49:22.0000	1	1.17	1.0	N
> 2	2022-06-06 18:51:12.0000	2022-06-06 19:04:42.0000	1	4.35	1.0	N
> 2	2022-06-06 18:55:50.0000	2022-06-06 18:56:59.0000	2	0.27	1.0	N
> 2	2022-06-06 18:55:50.0000	2022-06-06 18:56:59.0000	2	0.27	1.0	N

4. Select **Save as KQL Query Set** to save this and future queries for later use.
5. Under **KQL Queryset name**, enter *nyctaxiqs*, then select **Create**.



'Explore your data' enables you to run some quick queries to understand your data. This query can be saved as a KQL Queryset and persisted in the workspace as an item. Query set autosaves the queries as you type them and lets you resume from the point where you had stopped. In the next module, you will work with the KQL Queryset. Saving the quick query as KQL Query Set will automatically open your **KQL Queryset** with the queries that you wrote in the query editor.

Explore data further in the KQL Queryset

In this module, you are going to write queries using *Kusto Query Language* to explore the data that you have ingested from the Event hub and blob storage. Kusto Query Language is a powerful tool to explore your data and discover patterns, identify anomalies and outliers, create statistical modeling, and more. The query uses schema entities that are organized in a hierarchy similar to SQLs: databases, tables, and columns. Kusto query is a read-only request to process data and return results. The request is stated in plain text, using a data-flow model that is easy to read, author, and automate. Kusto queries are made of one or more query statements. You are going to write some simple Kusto queries to get familiar with the language and discover its power and simplicity.

Run the following queries in the new KQL Queryset you have created. Copy/paste each query into your environment, select the query and then select **Run**.

1. The following query returns the top 10 pickup locations in New York City for Yellow Taxis.

Query to be
pasted →

```
//Top 10 pickup locations
nytaxitrips
| summarize count() by PULocationID
| top 10 by count_
```

Table 1		
PULocationID		count_
> 237	61,883	
> 236	55,365	
> 132	55,057	
> 161	47,341	
> 162	40,191	
> 142	40,128	
> 186	37,716	
> 170	37,533	
> 230	35,370	

Note: Result of the query may not exactly match the screenshot provided as you are ingesting streaming data.

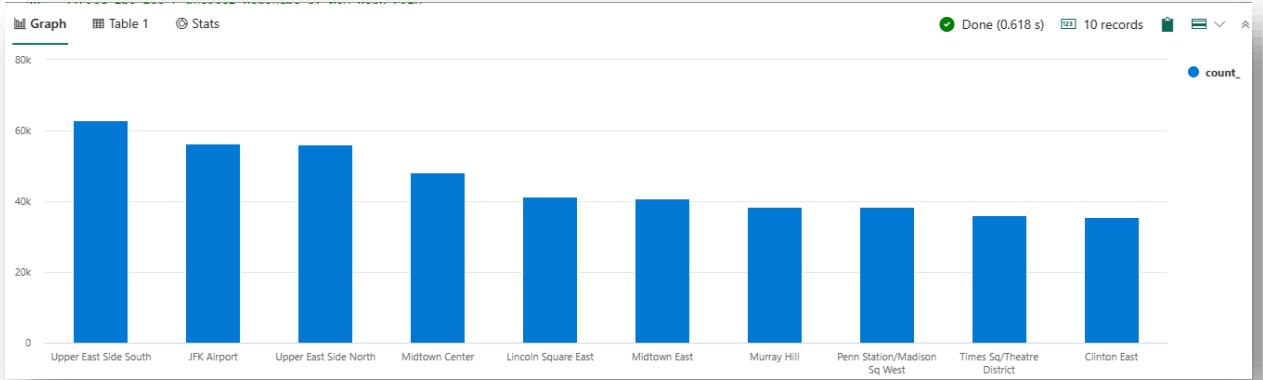
2. We will run the same query as in the previous step with an addition of looking up the corresponding zones of the top 10 pickup locations by using the 'locations' table.

Query to be
pasted →

```
//For the same top 10 locations, lookup the NYC zones --> Top 10 zones
nytaxitrips
| lookup (locations) on $left.PULocationID == $right.LocationID
| summarize count() by Zone
| top 10 by count_
| render columnchart
```

Table 1		
Zone		count_
> Upper East Side South	62,942	
> JFK Airport	56,417	
> Upper East Side North	56,052	
> Midtown Center	48,263	
> Lincoln Square East	41,236	
> Midtown East	40,895	
> Murray Hill	38,479	
> Penn Station/Madison Sq West	38,324	
> Times Sq/Theatre District	36,167	
> Clinton East	35,622	

Note: Result of the query may not exactly match the screenshot provided as you are ingesting streaming data.

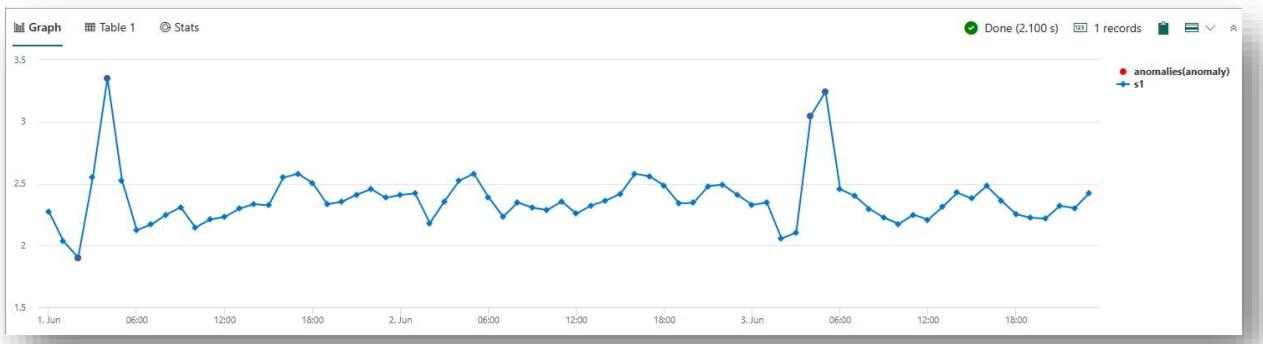


3. Let's check anomalies in the tips that have been given by the customers in the Manhattan borough. Hover over the red dots to see the values.

Query to be pasted →

```
//Find anomalies in the tips given by the customers
nyctaxitrips
| lookup (locations) on $left.PULocationID==$right.LocationID
| where Borough == "Manhattan"
| make-series s1 = avg(tip_amount) on tpep_pickup_datetime from datetime(2022-06-01) to datetime(2022-06-04) step 1h
| extend anomalies = series_decompose_anomalies(s1)
| render anomalychart with (anomalycolumns=anomalies)
```

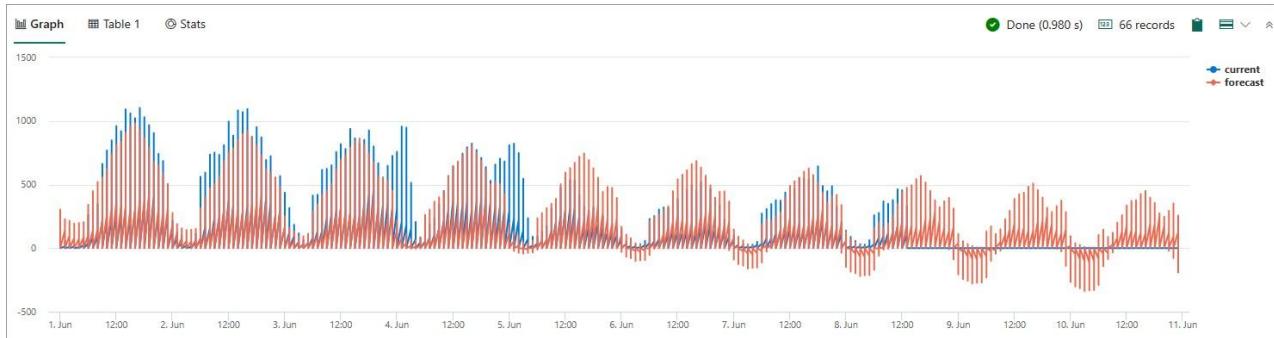
Note: Result of the query may not exactly match the screenshot provided as you are ingesting streaming data.



4. To ensure that the sufficient taxis are plying in the Manhattan borough, let's forecast the number of taxis needed per hour.

```
//Forecast the number of trips that will begin from Manhattan to line up the taxis in that borough
nyctaxitrips
| lookup (locations) on $left.PULocationID==$right.LocationID
| where Borough == "Manhattan"
| make-series s1 = count() on tpep_pickup_datetime from datetime(2022-06-01) to datetime(2022-06-08)+3d
step 1h by PULocationID
| extend forecast = series_decompose_forecast(s1, 24*3)
| render timechart
```

Note: Result of the query may not exactly match the screenshot provided below as you are ingesting streaming data.



Build Power BI report

A Power BI report is a multi-perspective view into a dataset, with visuals that represent findings and insights from that dataset.

1. Continuing the in same queryset, paste the following query. The output of this query will be used as the dataset for building the Power BI report.

```
//Find the total number of trips that started and ended at the same location
nyctaxitrips
| where PULocationID == DOLocationID
| lookup (locations) on $left.PULocationID==$right.LocationID
| summarize count() by Borough, Zone, Latitude, Longitude
```

2. Select the query and then select **Power BI** button.

The screenshot shows the Kusto Query Editor interface. At the top, there's a navigation bar with 'Home' and 'Manage' tabs, followed by buttons for 'Run', 'Recall', 'Save', 'Copy query', and 'Build Power BI report'. The 'Build Power BI report' button is highlighted with a red box. Below the navigation is a database dropdown set to 'NycTaxiDB' and a search bar. The main area contains a code editor with the following Kusto query:

```

28 nyctaxitrips
29 | where PUlocationID == $OlocationID
30 | lookup (locations) on $left.PUlocationID==$right.LocationID
31 | summarize count() by Borough, Zone, Latitude, Longitude
32
33

```

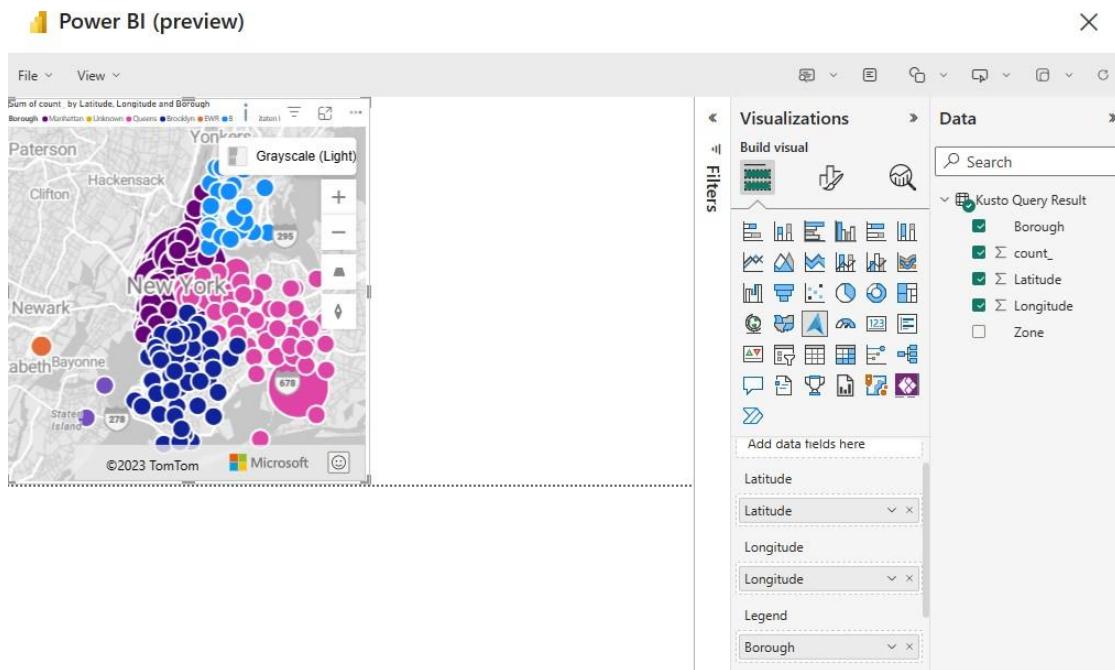
Power BI report editor will open with the result of the query available as a table with the name **Kusto Query Result**.

Note:

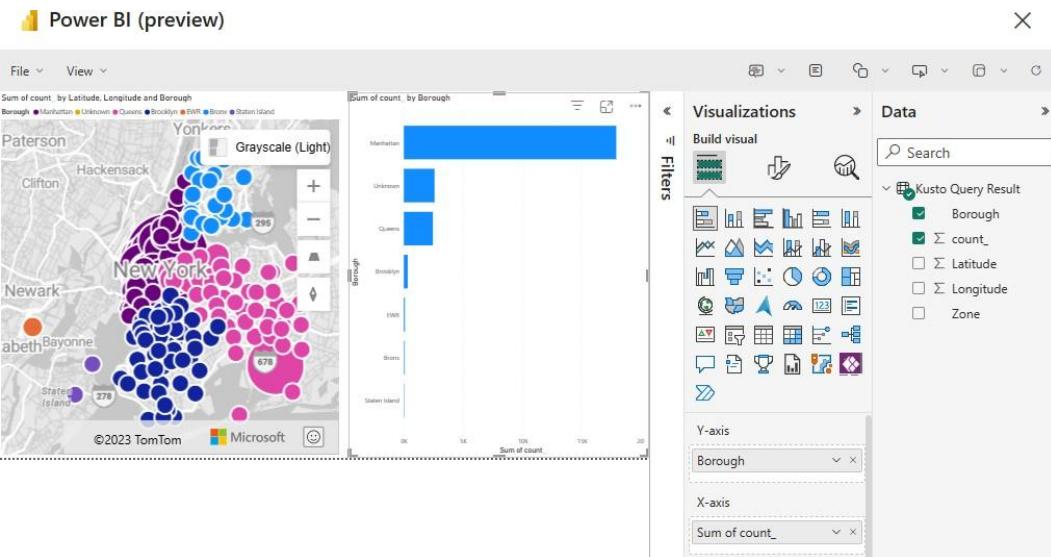
- (a) When you build a report, a dataset is created and saved in your workspace. You can create multiple reports from a single dataset.
- (b) Result of the query may not exactly match the screenshot provided below as you are ingesting streaming data.

If you delete the dataset, your reports will also be removed.

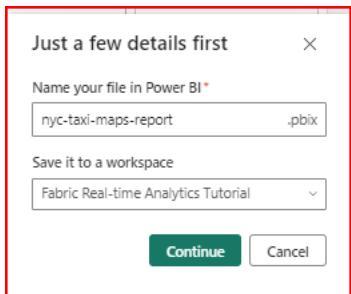
3. In the report editor, choose **Azure Maps** as the visual, drag **Latitude** field to Latitude, **Longitude** field to Longitude, **Borough** field to Legend, and **count_** field to Size. (For the report to display properly, tenant administrator also needs to enable **Use Azure Maps visual** in the **Tenant settings**).



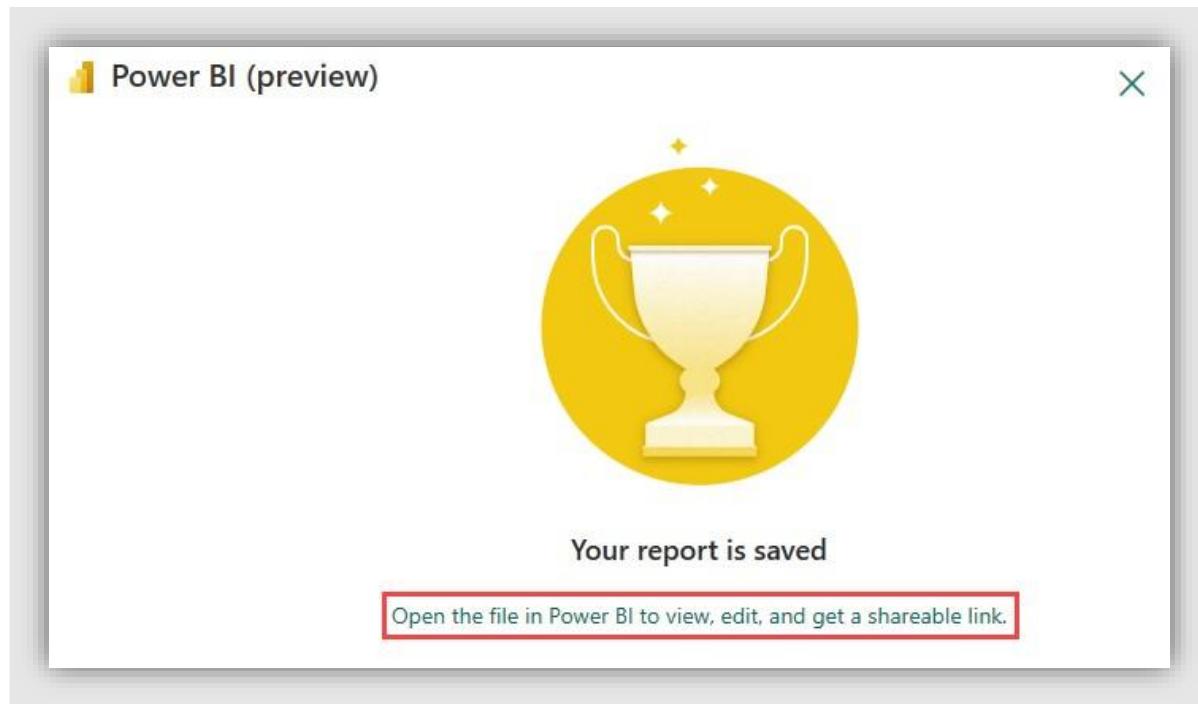
4. Add a **Stacked Bar Chart** to the canvas. Drag **Borough** field to Y-Axis and **count_** to the X-axis



5. Click File > Save
6. Under **Name your file in Power BI**, enter *nyc-taxi-maps-report*.



7. Select the workspace in which you want to save this report. The report can be saved in a different workspace than the one you started in.
8. Select **Open the file in Power BI to view, edit, and get a shareable link** to view and edit your report.



Create OneLake shortcut

Now that you have finished exploring your data, you may want to access the underlying data from other Fabric experiences.

OneLake is a single, unified, logical data lake for Fabric to store lakehouses, warehouses and other items. Shortcuts are embedded references within OneLake that point to other files' store locations. The embedded reference makes it appear as though the files and folders are stored locally but in reality; they exist in another storage location. Once you create a shortcut, you can access your data in all of Fabric's experiences. Shortcuts can be updated or removed from your items, but these changes will not affect the original data and its source.

1. Select **Create** in the **Navigation pane**.

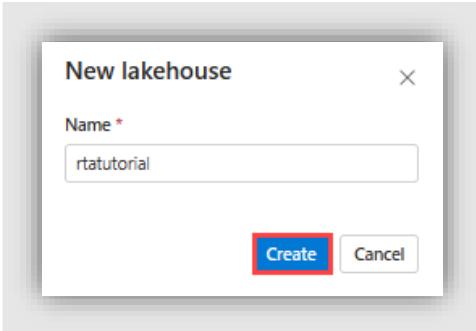


2. Under **Data engineering**, select **Lakehouse**.

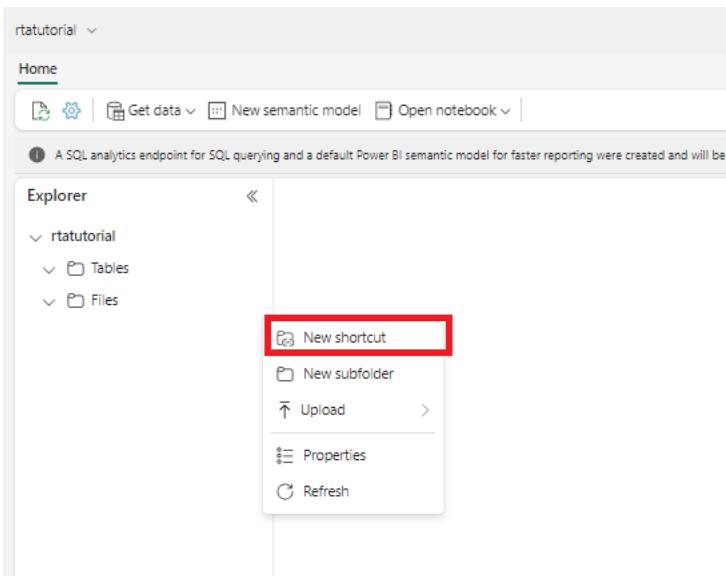
Data Engineering
Create a lakehouse and operationalize your workflow to build, transform, and share your data estate.

- Lakehouse** Store big data for cleaning, querying, reporting, and sharing.
- Notebook** Explore data and build machine learning solutions with Apache Spark applications.
- Environment (Preview)** Set up shared libraries, Spark compute settings, and resources for notebooks and Spark job definitions.
- Spark Job Definition** Define, schedule, and manage your Apache Spark jobs for big data processing.

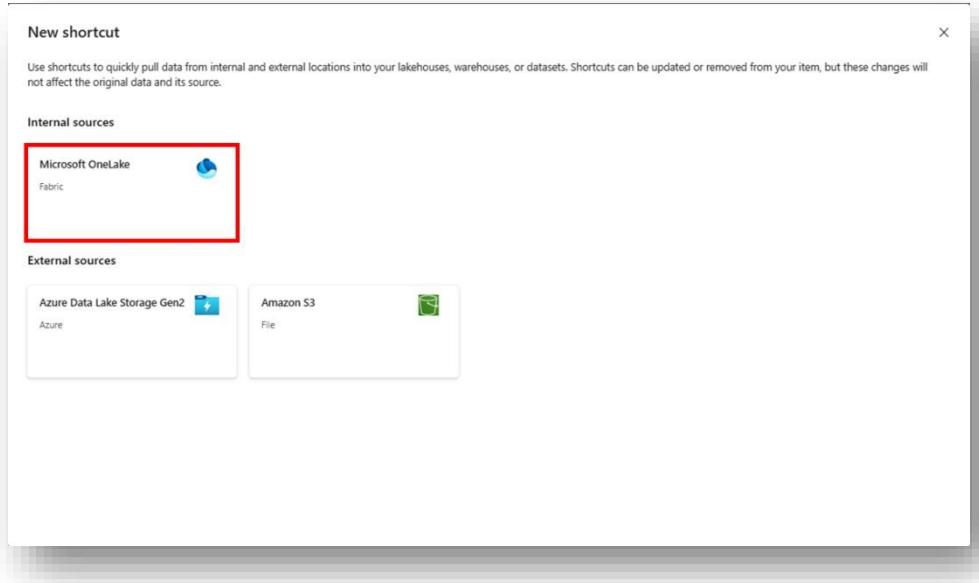
- Enter *rstatutorial* as your Lakehouse name, then select **Create**.



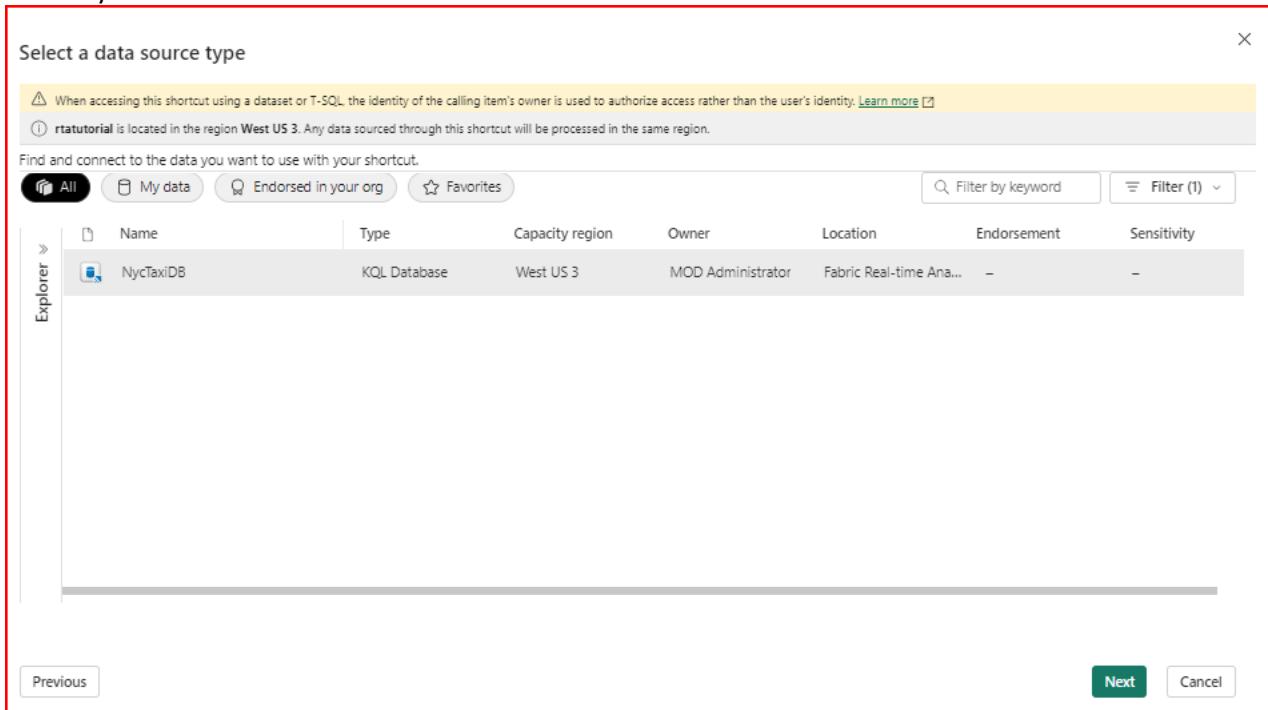
- In Lakehouse Explorer, navigate to **Files**, select “...”, select **New Shortcut** from the menu.



- Under **Internal sources**, select **OneLake**.



- In **Select a data source type**, Filter KQL Database and select **your KQL Database**, then select **Next** to connect the data to your shortcut.



- To connect the table with the data from Eventstream, select **>** to expand the tables in the left-hand pane, then select the table titled **nyctaxitrips**.

New shortcut

When accessing this shortcut using a dataset or T-SQL, the identity of the calling item's owner is used to authorize access rather than the user's identity. [Learn more](#)

rtatutorial is located in the region West US 3. Any data sourced through this shortcut will be processed in the same region.

Find and connect to the data you want to use with your shortcut.

Select the table and files that you want to include in your shortcut.

OneLake

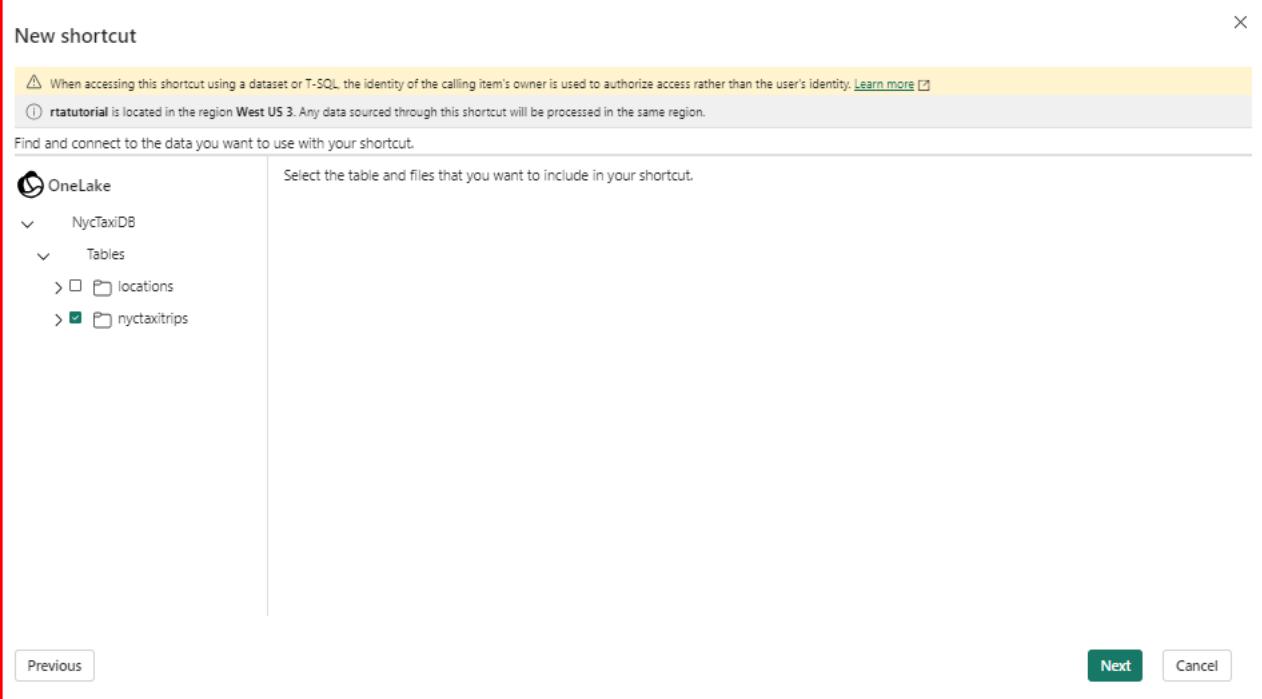
NycTaxiDB

Tables

locations

nyctaxitrips

Previous Next Cancel



4. Select **Create** to create the shortcut. The Lakehouse will automatically refresh. Click on Close

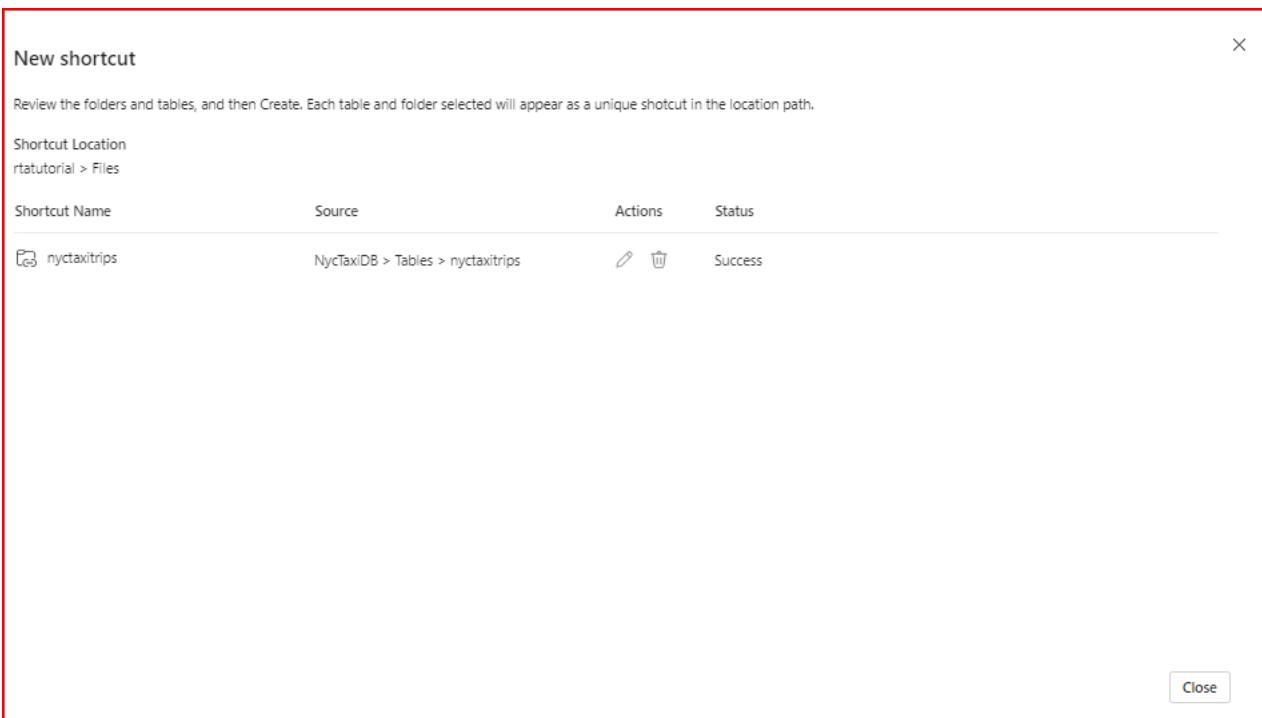
New shortcut

Review the folders and tables, and then Create. Each table and folder selected will appear as a unique shortcut in the location path.

Shortcut Location
rtatutorial > Files

Shortcut Name	Source	Actions	Status
nyctaxitrips	NycTaxiDB > Tables > nyctaxitrips		Success

Close



The Lakehouse shortcut has been created. Without additional management, you now have one logical copy of your data in the Lakehouse that you can use in other Fabric experiences, such as Notebooks and Spark jobs. It is important to note that there is a latency to OneLake Availability as of when the parquet files will be seen after shortcut is created. The lag often depends on the size of the streaming data.

Module 4: Clean up resources

You can delete individual reports, eventstreams, KQL databases, KQL querysets, and other items or remove the entire workspace.

1. Select Fabric Real-time Analytics Solution Tutorial in the left-hand navigation menu to return to the workspace artifact view



2. Below the workspace name and description at the top of the workspace header, select Workspace settings.

A screenshot of a workspace header. At the top left is the workspace name "Fabric Real-time Analytics Tutorial". To its right are several buttons: "+ New", "Upload", "Create deployment pipeline", "Create app", "Manage access", and "Workspace settings". A red rectangular box highlights the "Workspace settings" button.

3. Select **Other > Remove this workspace**

A screenshot of the "Workspace settings" dialog box. On the left, there's a sidebar with sections like "About", "Premium", "Azure connections", "System storage", "Git integration", "OneLake", and "Other". The "Other" section is highlighted with a red rectangular box. On the right, there's a main area titled "Delete this workspace and all data and items in it?" with a warning message about permanent deletion after 7 days. Below this is a green button labeled "Remove this workspace", which is also highlighted with a red rectangular box.

4. Select **Delete** on the warning.