Adilson Cesar

Follow 165 Followers About

Using Azure Data Lake to copy data from CSV file to a SQL database

Adilson Cesar Nov 15, 2018 · 5 min read

Great, but what is Azure Data Lake?

Azure Data Lake includes all the capabilities required to make it easy for developers, data scientists, and analysts to store data of any size, shape, and speed, and do all types of processing and analytics across platforms and languages. It removes the complexities of ingesting and storing all of your data while making it faster to get up and running with batch, streaming, and interactive analytics.

A sources. The service does not impose limits on account sizes, file sizes, or the amount of data that can be stored in a data lake.

> *Pricing for Azure Data Lake is dependent upon numerous variables, including storage capacity, the number of analytics units (AUs) per minute. As of this writing, the Azure Data Lake Store service is priced at $0.039 per GB per month for* pay as you go*, with capacity-based discounts up to 33% for monthly commitments. The Azure Pricing Calculator can help customers determine exact data lake costs.*

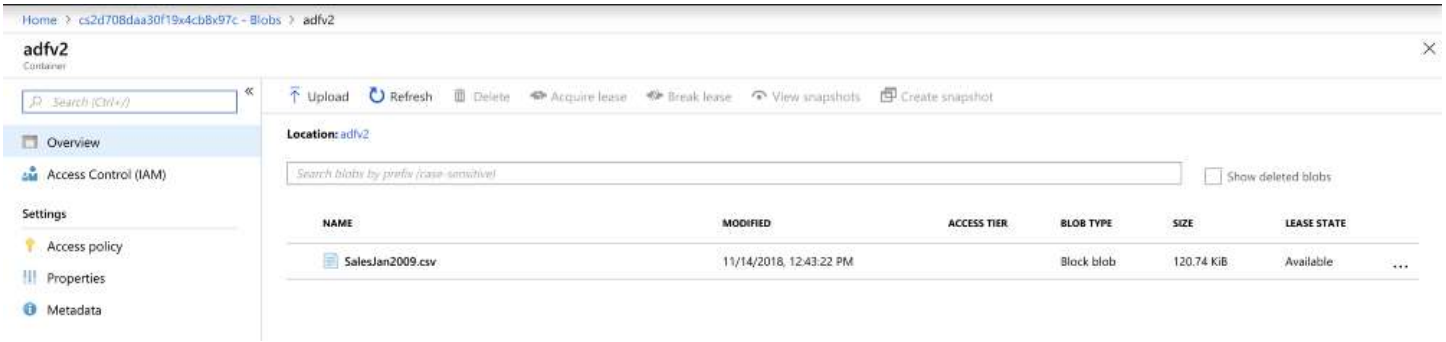
In this tutorial, you use the Azure portal to create a data factory. Then, you use the Copy Data tool to create a pipeline that copies data from CSV file data to a SQL database.

Prerequisites

- Azure storage account: Use Blob storage as the *source* data store. If you don't have an Azure storage account, see the instructions in Create a storage account.
- Azure SQL Database: Use a SQL database as the *sink* data store. If you don't have a SQL database, see the instructions in Create a SQL database.
- The Sales Jan 2009 file contains some sales transactions. You can download CSV data here.

Prepare your Blob storage and your SQL database tables by performing these steps.

- Create a container named **adfv2** and upload the CSV file to the *container*. You can use various tools to perform these tasks, such as Azure Storage Explorer.

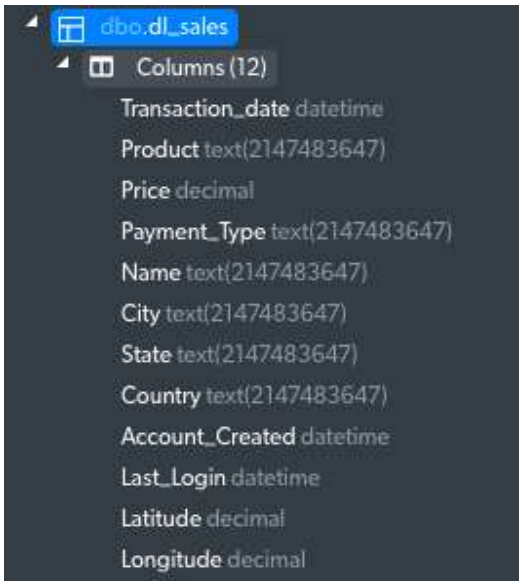


Blob Storage folder

Use the following SQL script to create a database named **dbo.adl_db** in your SQL database:

Also, create a sink SQL table:

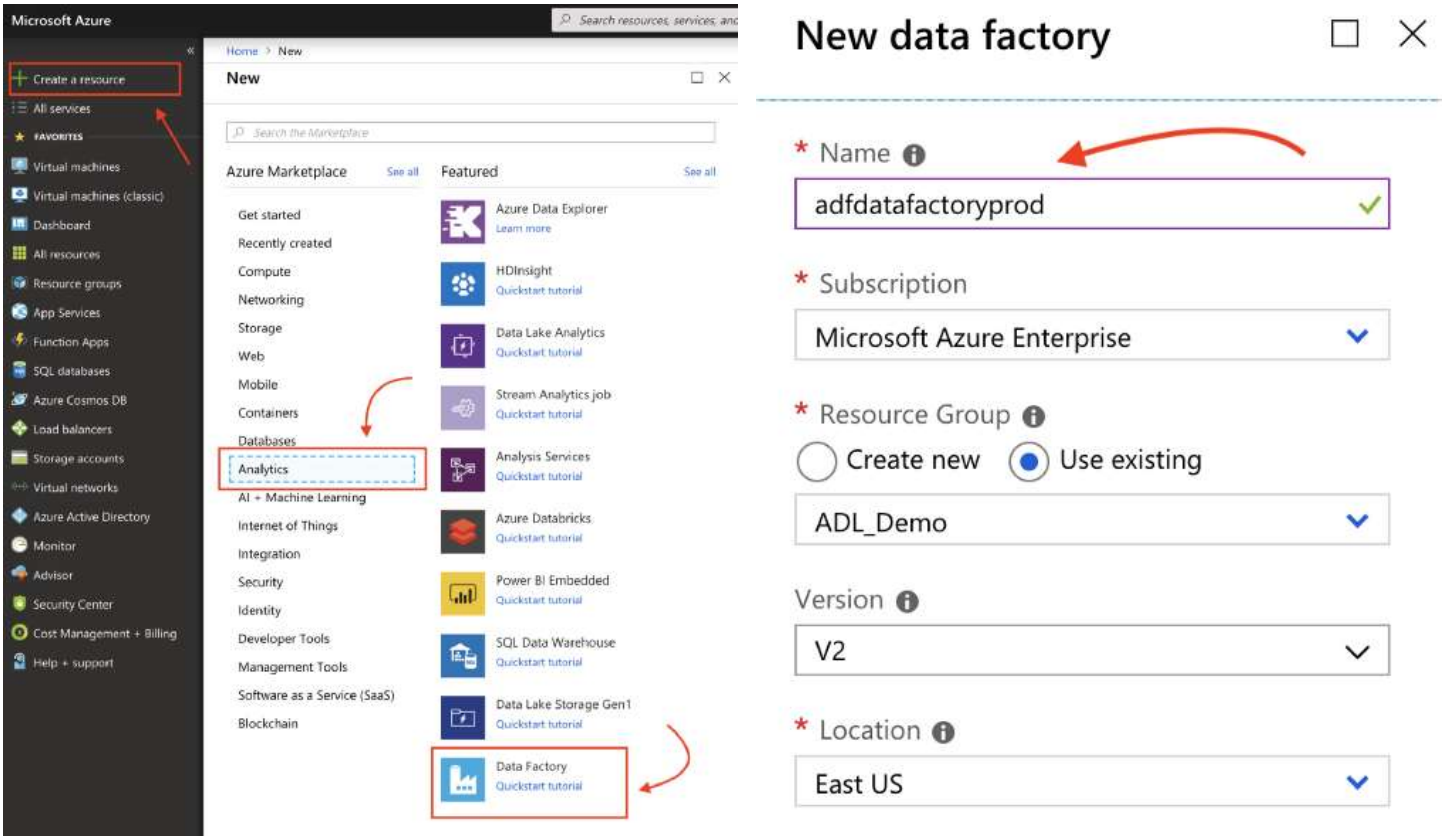
```
CREATE TABLE dl_sales (
Transaction_date DATETIME,
Product TEXT,
Price DECIMAL,
Payment_Type TEXT,
Name TEXT,
City TEXT,
State TEXT,
Country TEXT,
Account_Created DATETIME,
Last_Login DATETIME,
Latitude DECIMAL,
Longitude DECIMAL,
)
```



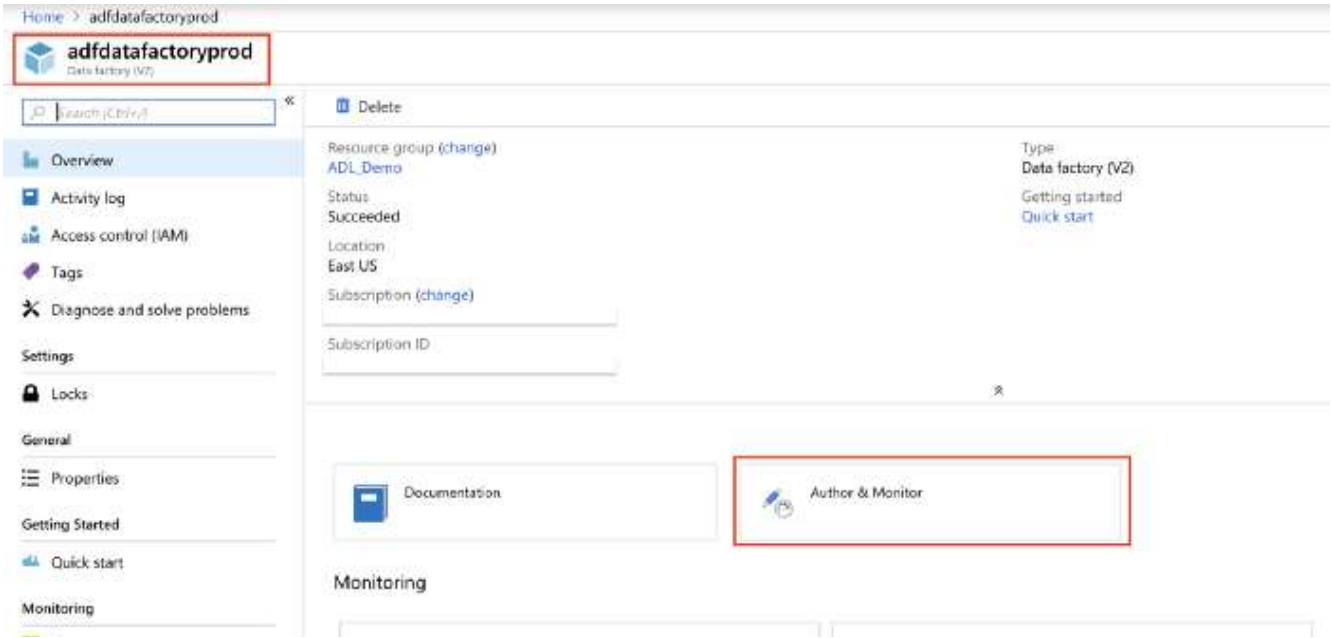
SQL Database created!

Let's create a data factory.

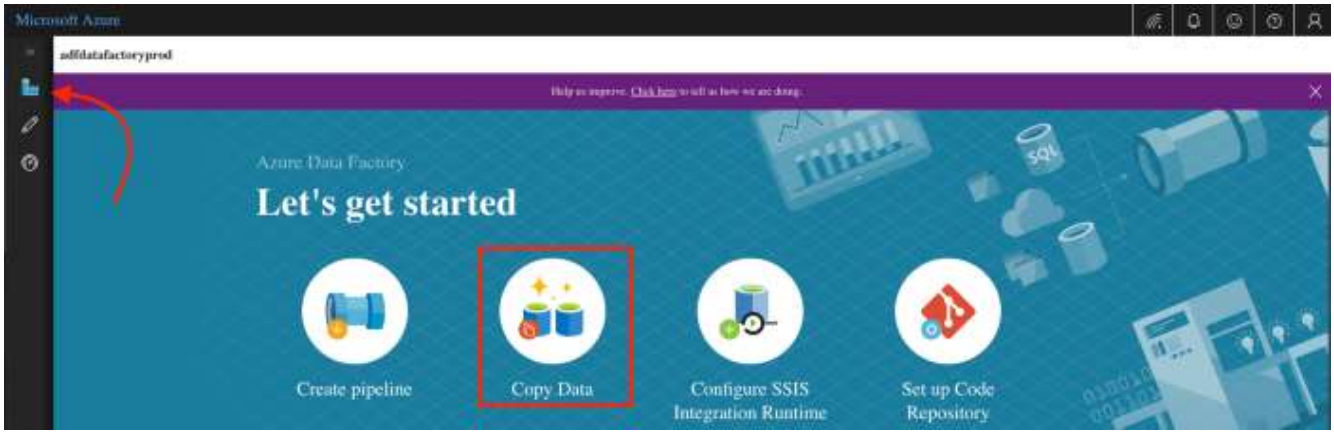
> *Azure Data Factory, a cloud data integration service, to compose data storage, movement, and processing services into automated data pipelines.*
> *This platform for these kinds of scenarios below. It is a* cloud-based data integration service that allows you to create data-driven workflows in the cloud that orchestrate and automate data movement and data transformation.



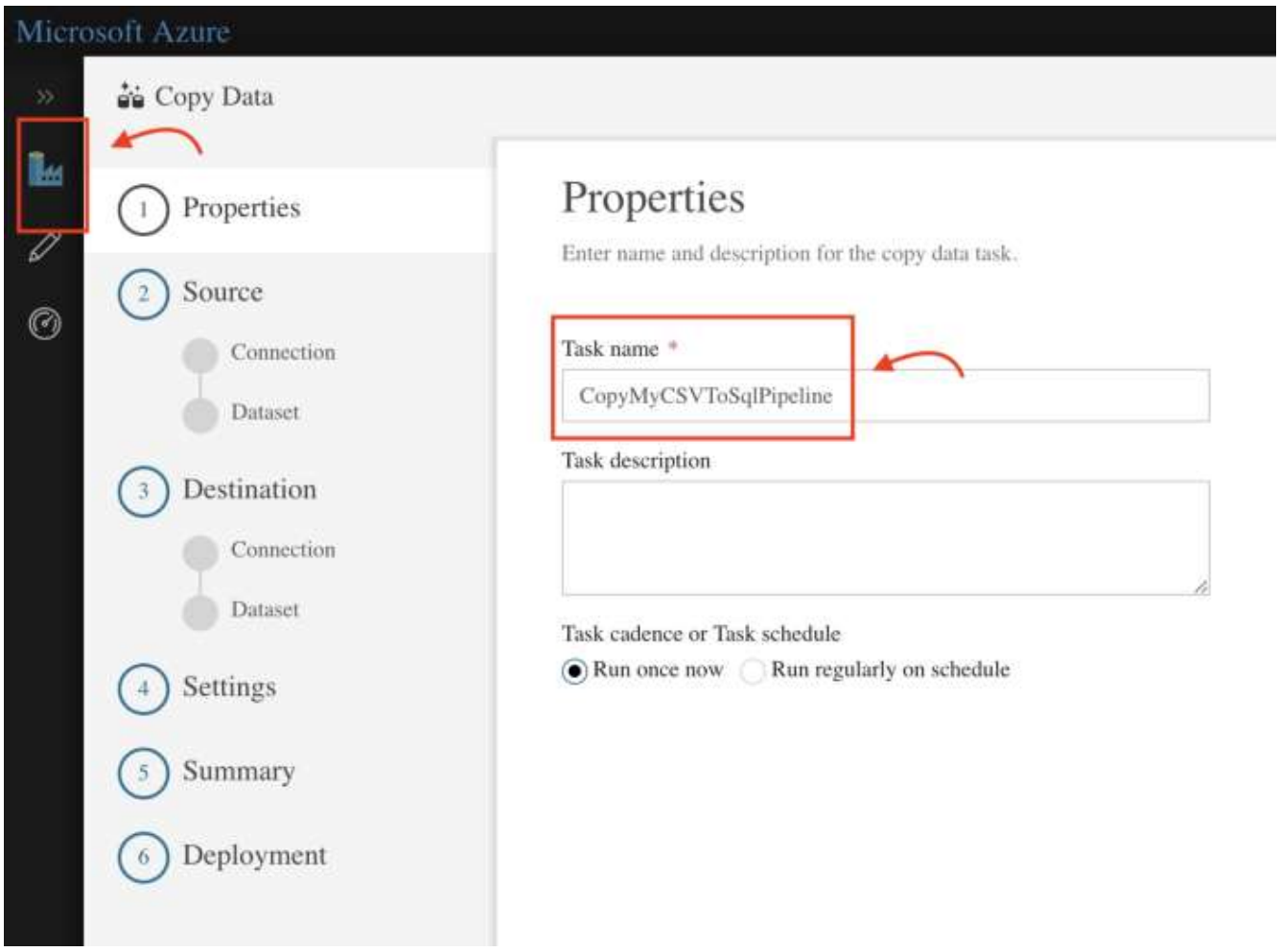
On the left menu, Create a resource + New > Data + Analytics > Data Factory. On the New data factory page, under Name, enter **adfdatafactoryprod.**

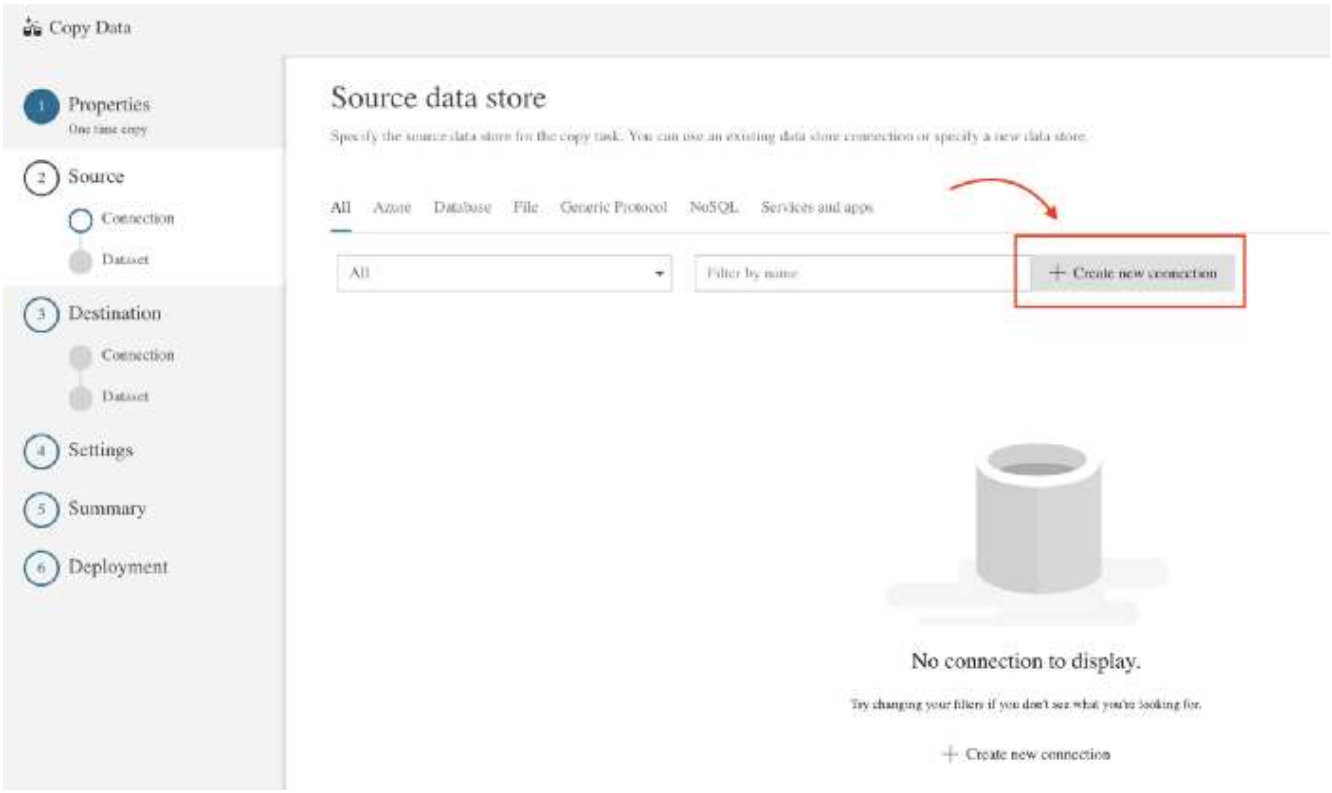


Great!! Now you can launch the Azure Data Factory user interface (UI) in a separate tab, select the **Author & Monitor tile.**

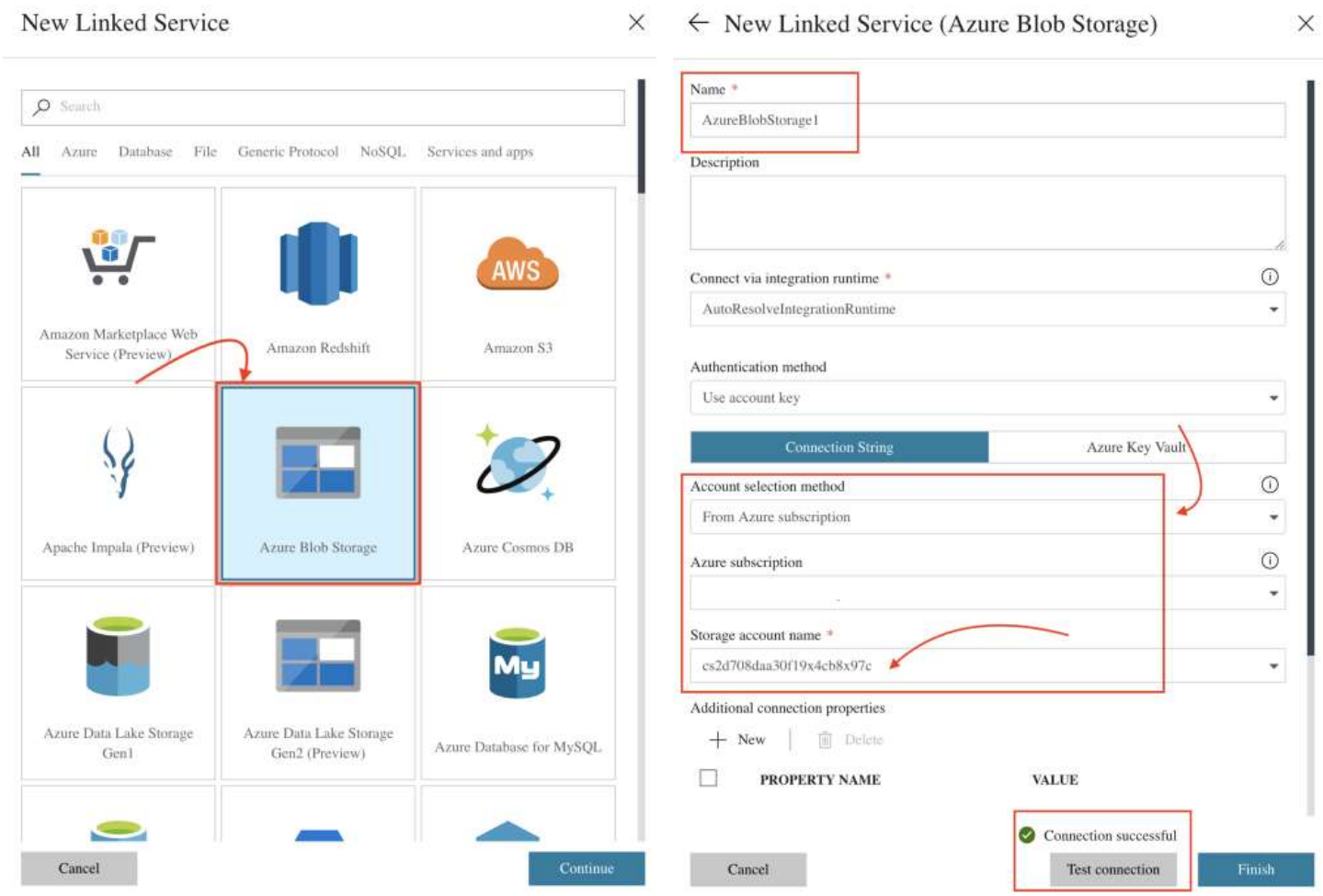


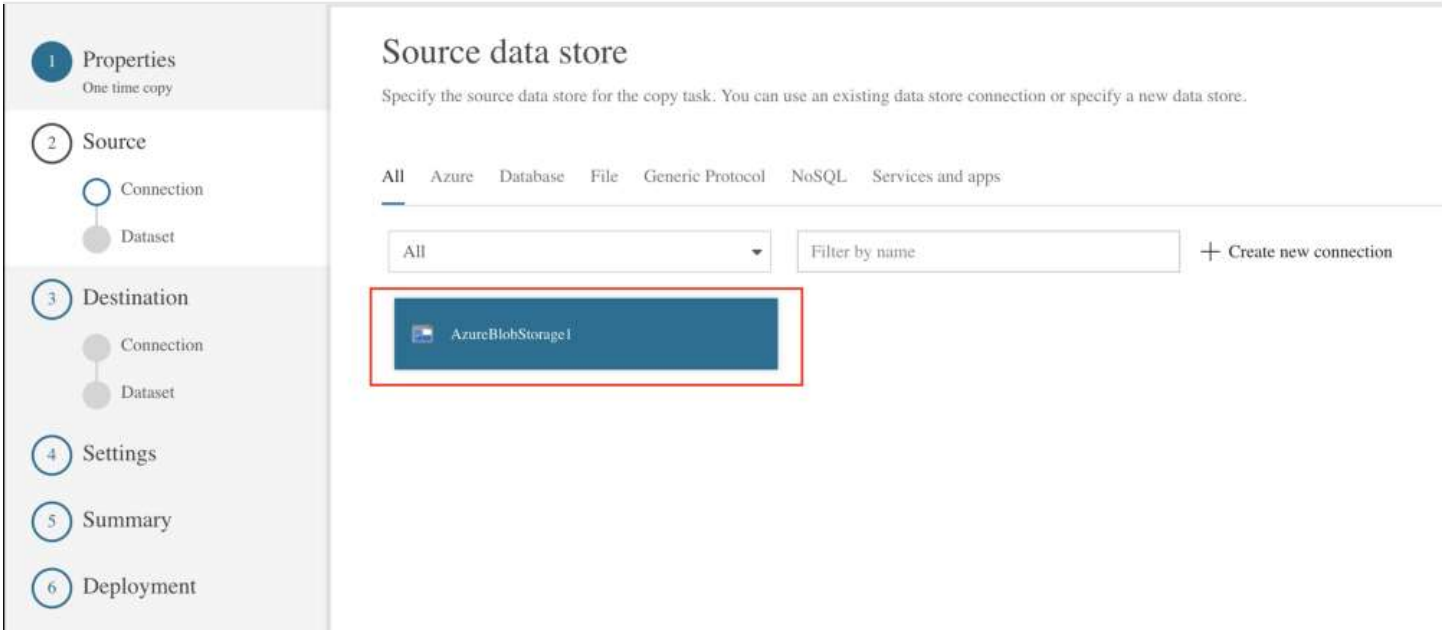
On the Let's get started page and Click **Copy Data.**



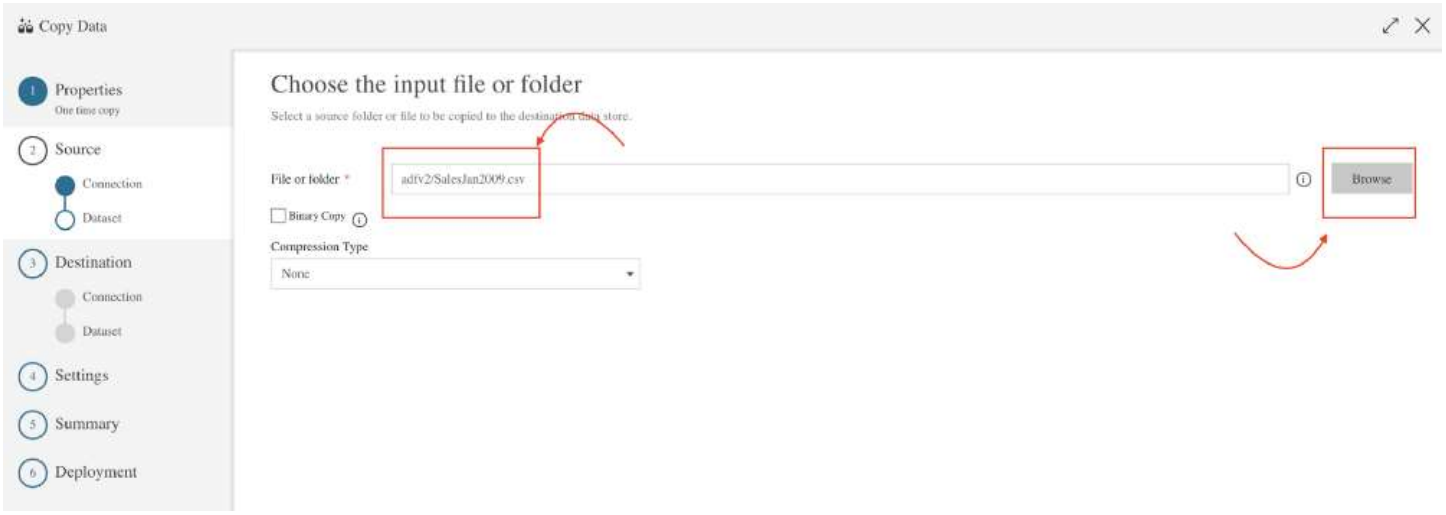


Create new connection

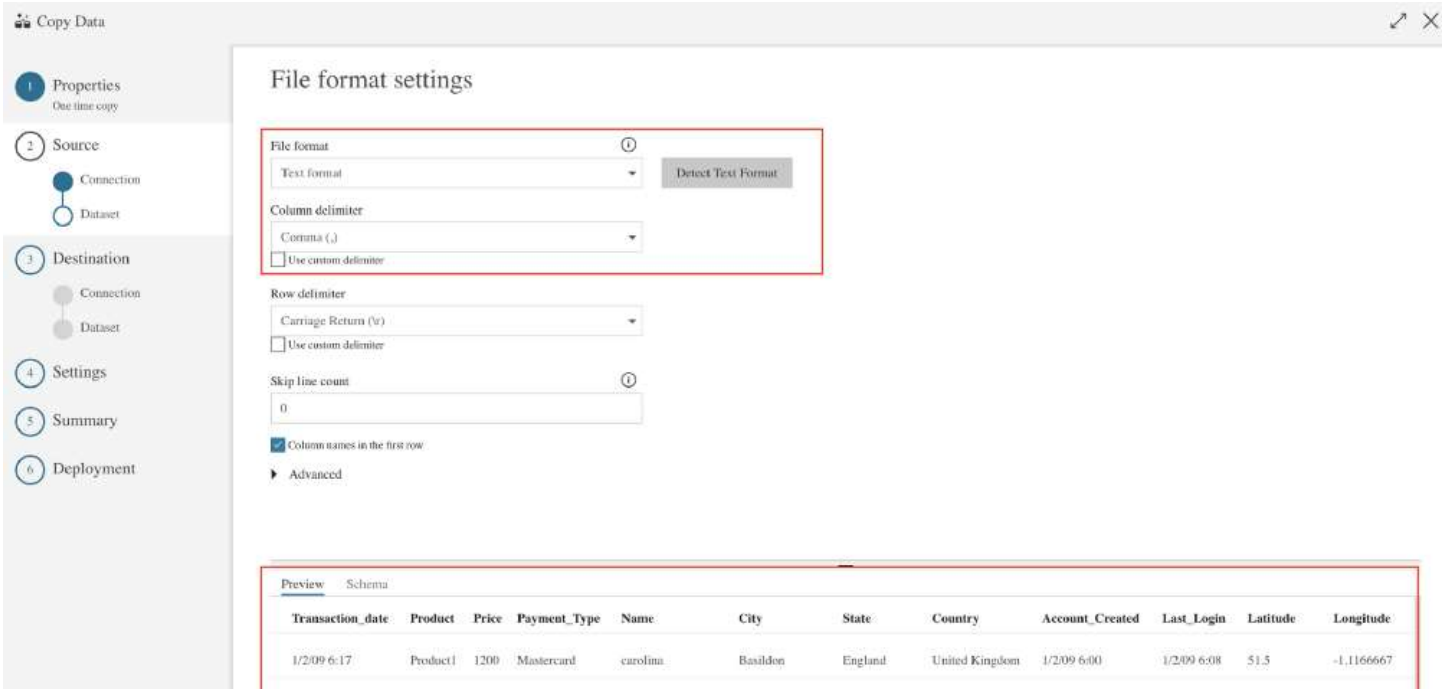




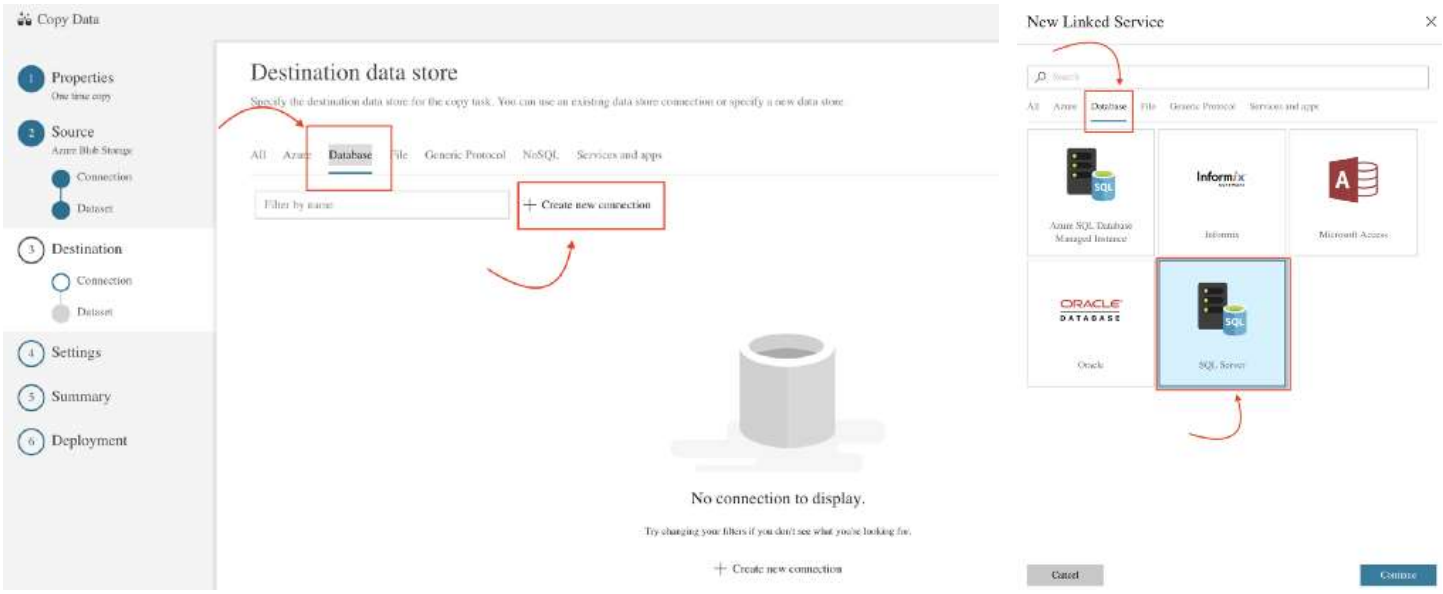
Blob Storage connection created!



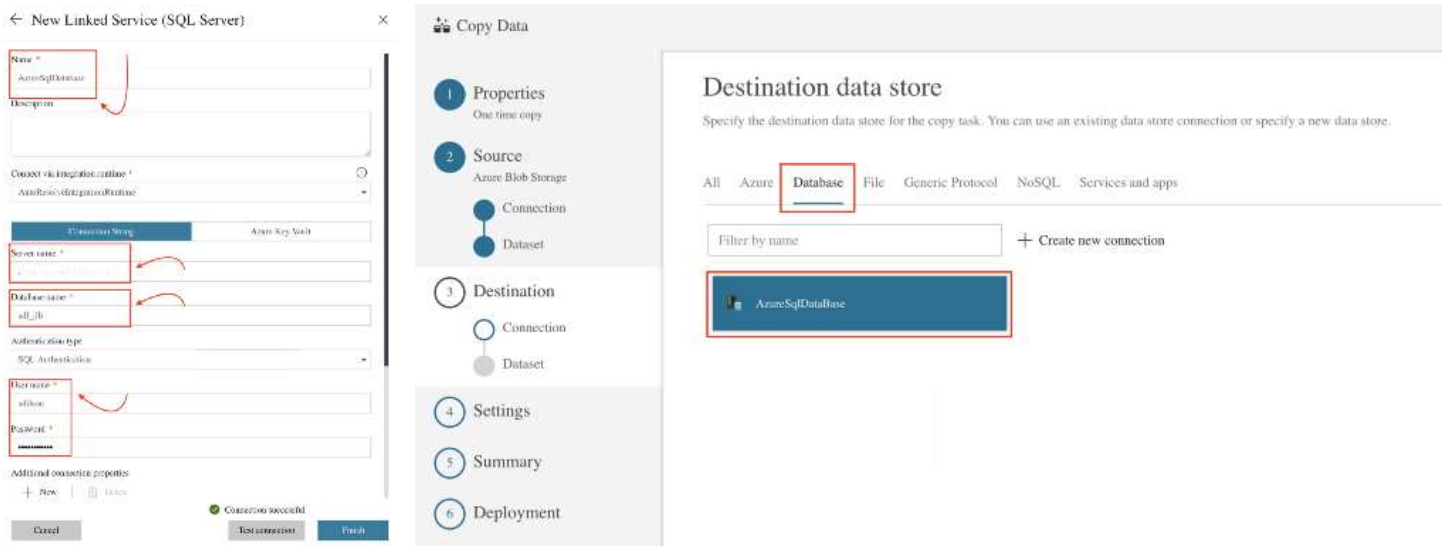
Click in Browse and choose your CSV File



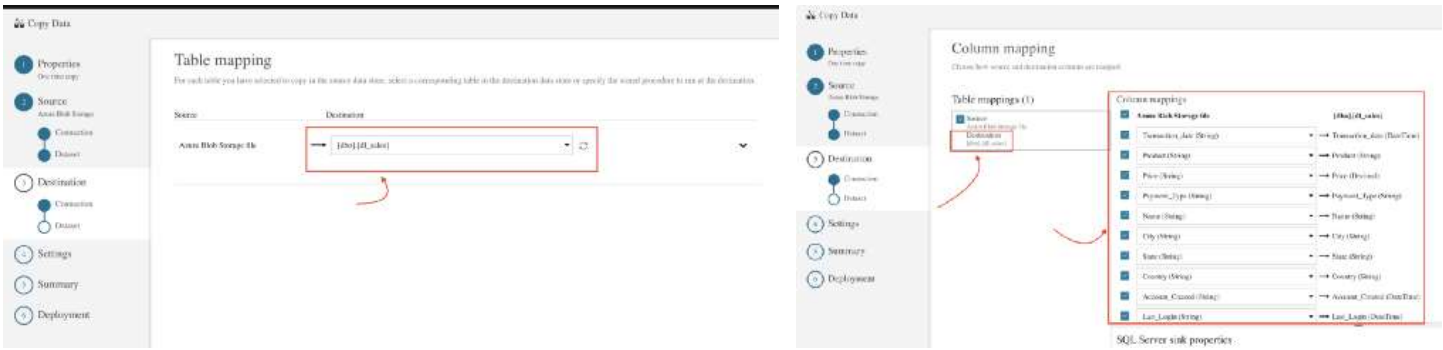
Checking Dataset



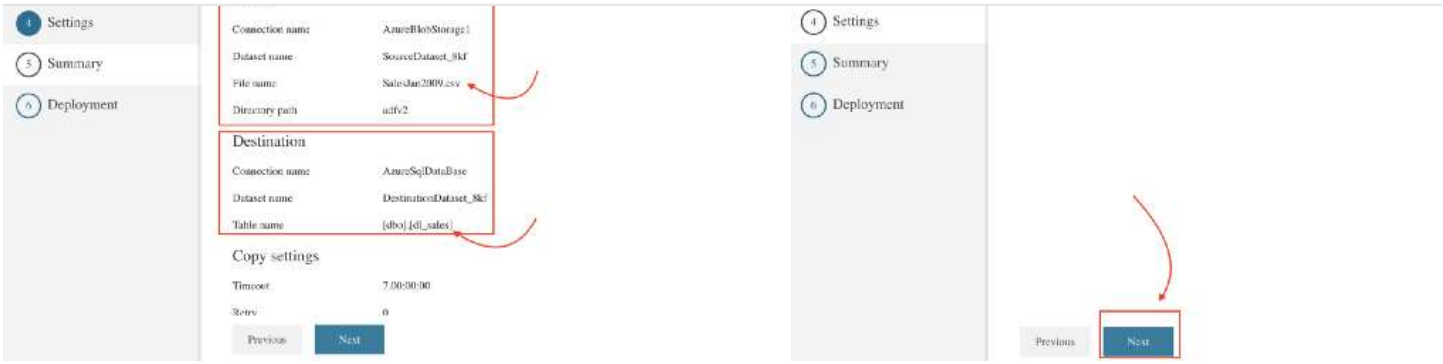
SQL Database Destination and create new connection

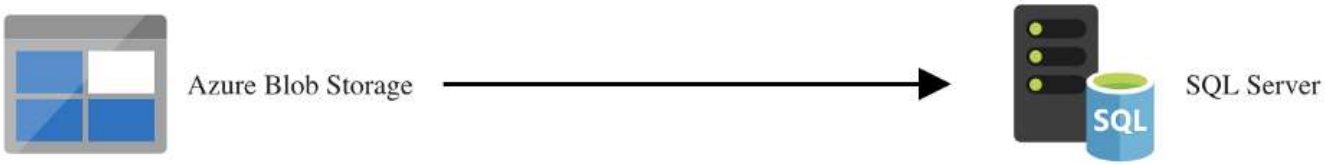


Left Choose Name, SQL Server Name, Database Name and DB credentials — Right Destination created!

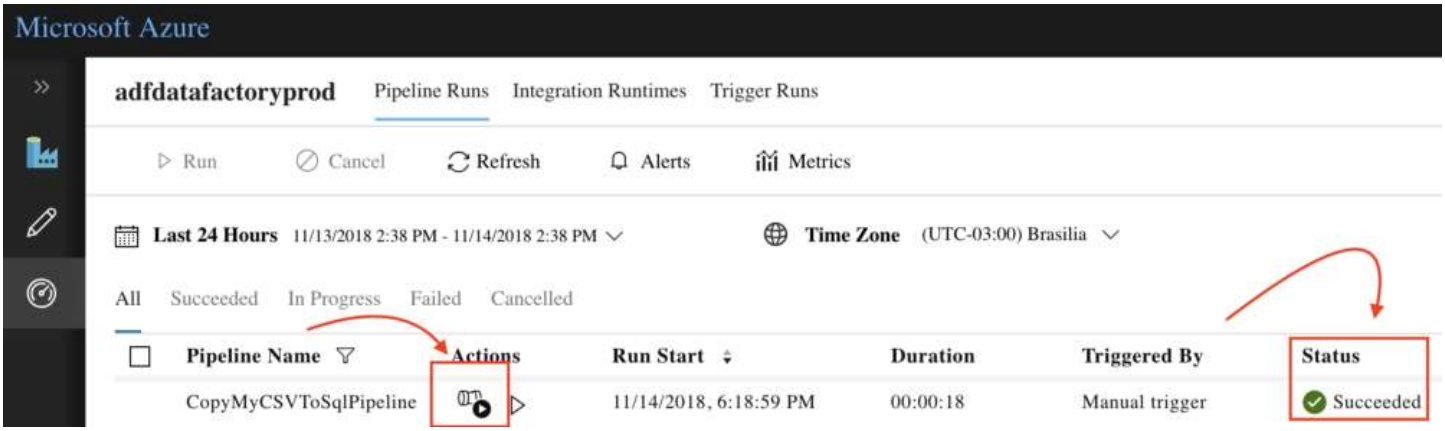


Choose your SQL Database





On the Deployment page, select Monitor to monitor the pipeline (task).



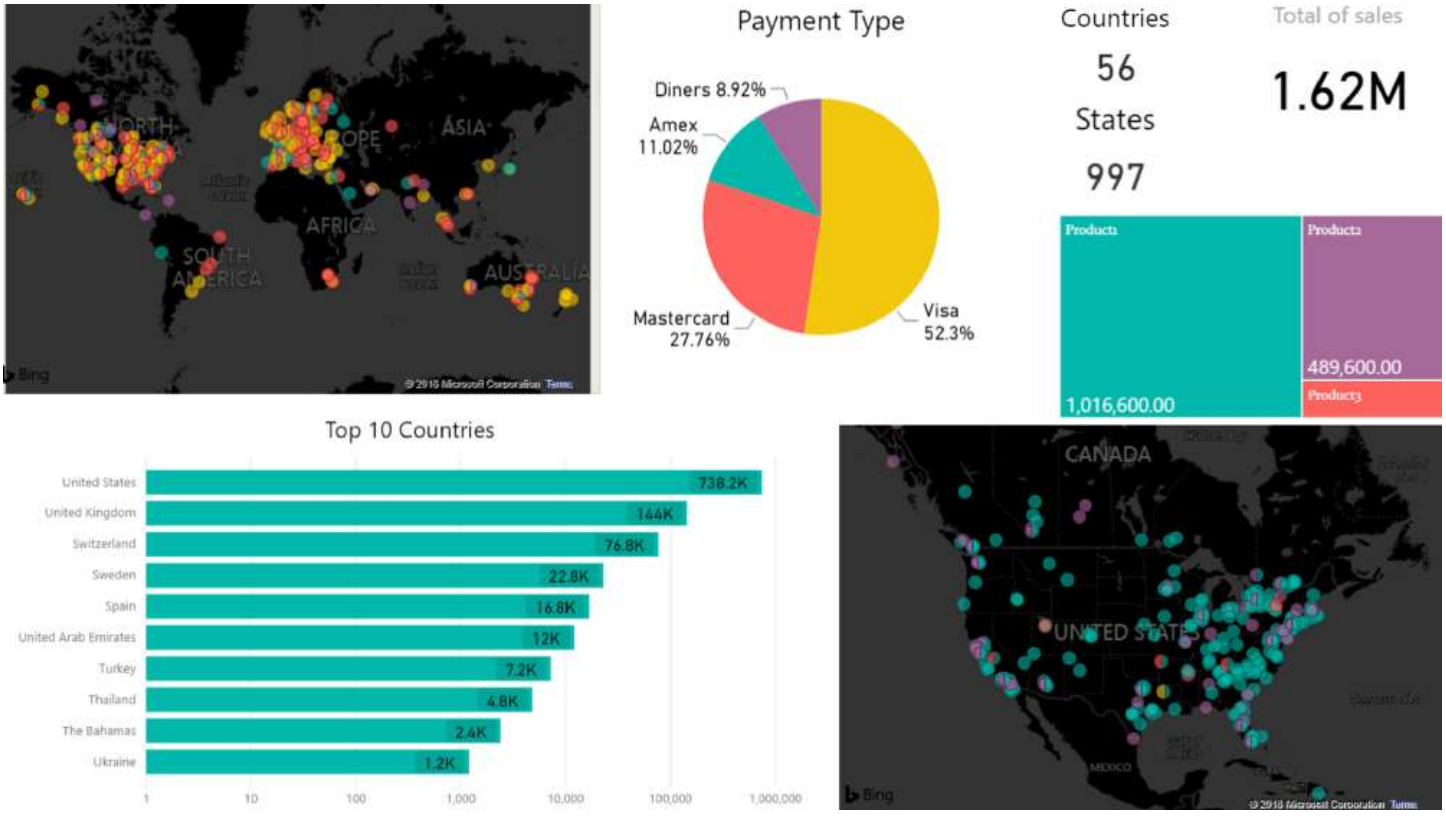
Notice that the Monitor tab on the left is automatically selected. The Actions column includes links to view activity run details and to rerun the pipeline. Select Refresh to refresh the list.

Well Done!! Verify that the data is inserted into the dbo.dl_sales table in your SQL

#	Transaction_date	Product	Price	Payment_Type	Name	City	State	Country	Account_Created	Last_Login	Latitude	Longitude
1	2009-01-02 06:17:00.0	Product1	1200	Mastercard	carolina	Basildon	England	United Kingd...	2009-01-02 06:00:00.0	2009-01-02 06:08:00.0	51	-1
2	2009-01-02 04:53:00.0	Product1	1200	Visa	Betina	Parkville	MO	United States	2009-01-02 04:42:00.0	2009-01-02 07:49:00.0	39	-94
3	2009-01-02 13:08:00.0	Product1	1200	Mastercard	Federica e A...	Astoria	OR	United States	2009-01-01 16:21:00.0	2009-01-03 12:32:00.0	46	-123
4	2009-01-03 14:44:00.0	Product1	1200	Visa	Gouya	Echuca	Victoria	Australia	2005-09-25 21:13:00.0	2009-01-03 14:22:00.0	-36	144
5	2009-01-04 12:56:00.0	Product2	3600	Visa	Gerd W	Cahaba Heig...	AL	United States	2008-11-15 15:47:00.0	2009-01-04 12:45:00.0	33	-86
6	2009-01-04 13:19:00.0	Product1	1200	Visa	LAURENCE	Mickleton	NJ	United States	2008-09-24 15:19:00.0	2009-01-04 13:04:00.0	39	-75
7	2009-01-04 20:11:00.0	Product1	1200	Mastercard	Fleur	Peoria	IL	United States	2009-01-03 09:38:00.0	2009-01-04 19:45:00.0	40	-89
8	2009-01-02 20:09:00.0	Product1	1200	Mastercard	adam	Martin	TN	United States	2009-01-02 17:43:00.0	2009-01-04 20:01:00.0	36	-88
9	2009-01-04 13:17:00.0	Product1	1200	Mastercard	Renee Elisab...	Tel Aviv	Tel Aviv	Israel	2009-01-04 13:03:00.0	2009-01-04 22:10:00.0	32	34

\o/

Also, don't forget to configure your PowerBI with SQL Database data.



Sample of data from SQL

Conclusion

Azure Data Lake is an important new part of Microsoft's ambitious cloud offering. With Data Lake, Microsoft provides service to store and analyze data of any size at an affordable cost. In related posts, we will learn more about Data Lake Store, Data Lake Analytics, and HDInsight.