Data Engineering in the Cloud

Build Aggregates and Output Them to Synapse Analytics

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- After the real-time streaming is completed, we build aggregates and output these to Synapse Analytics using Stream Analytics windowing functions
- This lab is identical to the last lab except for the queries in and

Lab

- Step 1. Create a datalake (storage account with hierarchical namespace enabled)
  - To lower the cost, you may choose `Redundancy` as `LRS`
  - Create a container in the storage account for the Synapse to use

- Step 2. Create an Azure Synapse workspace and
  - Create a dedicates SQL pool
  - Create a table in the SQL pool with the following SQL code

```
CREATE TABLE [dbo].[Orders]
(
[OrderID] varchar(10),
[Quantity] int,
[UnitPrice] decimal(5,2),
[DiscountCategory] varchar(10)
)
```

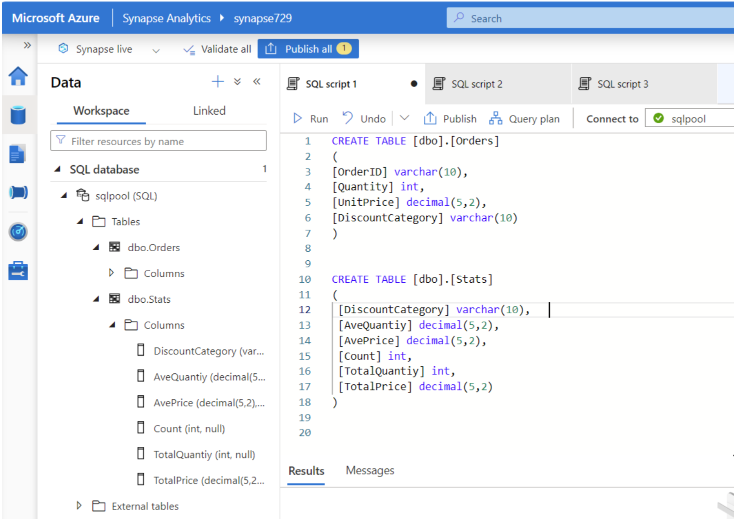
Lab

- Create another table in the SQL pool with the following SQL code

```
CREATE TABLE [dbo].[Stats]
(
 [DiscountCategory] varchar(10),
 [AveQuantiy]  decimal(5,2),
 [AvePrice]  decimal(5,2),
 [Count] int,
 [TotalQuantiy] int,
 [TotalPrice]  decimal(5,2)
)
```

Lab

- That is, we created two tables for the database



- SQL database 1
  - sqlpool (SQL)
    - Tables
      - dbo.Orders
        - Columns
      - dbo.Stats
        - Columns
          - DiscountCategory (var...
          - AveQuantiy (decimal(5...
          - AvePrice (decimal(5,2),...
          - Count (int, null)
          - TotalQuantiy (int, null)
          - TotalPrice (decimal(5,2...
    - External tables

```sql
CREATE TABLE [dbo].[Orders]
(
[OrderID] varchar(10),
[Quantity] int,
[UnitPrice] decimal(5,2),
[DiscountCategory] varchar(10)
)


CREATE TABLE [dbo].[Stats]
(
 [DiscountCategory] varchar(10),
 [AveQuantiy] decimal(5,2),
 [AvePrice] decimal(5,2),
 [Count] int,
 [TotalQuantiy] int,
 [TotalPrice] decimal(5,2)
)
```

Lab

- Step 3. Search for `Event Hubs` to create an event hubs namespace
  - Create a Event hub in the event hubs namespace

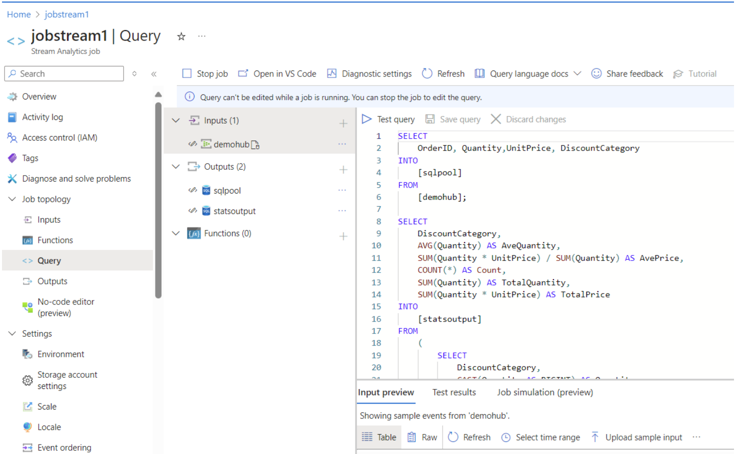
- Step 4. Create a Stream Analytics job (with **two outputs** shown below)
  - Add the storage account
  - Add an input under `Job Topology` with input from the Event hub
  - Add one output under `Job Topology` with output of input data to the Synapse Analytics
  - Add the 2nd output under `Job Topology` with output of the statistics of the input data to the Synapse Analytics
  - Click on the "Query" option under `Job topology` with code in the next slide
- Data types (Azure Stream Analytics): https://learn.microsoft.com/en-us/stream-analytics-query/data-types-azure-stream-analytics

Lab

```
SELECT
    OrderID, Quantity,UnitPrice, DiscountCategory
INTO
    [sqlpool]
FROM
    [demohub];
```

Lab

Lab



Query can't be edited while a job is running. You can stop the job to edit the query.

```
SELECT
    OrderID, Quantity,UnitPrice, DiscountCategory
INTO
    [sqlpool]
FROM
    [demohub];

SELECT
    DiscountCategory,
    AVG(Quantity) AS AveQuantity,
    SUM(Quantity * UnitPrice) / SUM(Quantity) AS AvePrice,
    COUNT(*) AS Count,
    SUM(Quantity) AS TotalQuantity,
    SUM(Quantity * UnitPrice) AS TotalPrice
INTO
    [statsoutput]
FROM
    (
        SELECT
            DiscountCategory,
```

Showing sample events from 'demohub'.

Lab

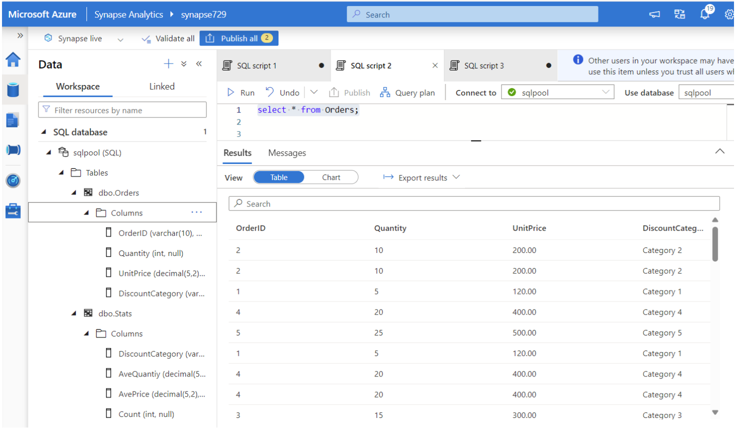
- Step 5: Start the Stream Analytics job

- Step 6: Go to Event Hubs Namespace to generate real-time data or use other applications to generate data

```
[
    {
        "OrderID": "1",
        "Quantity": 5,
        "UnitPrice": 120,
        "DiscountCategory": "Category 1"
    },
    {
        "OrderID": "2",
        "Quantity": 10,
        "UnitPrice": 200,
        "DiscountCategory": "Category 2"
    },
    {
        "OrderID": "3",
        "Quantity": 15,
        "UnitPrice": 300,
        "DiscountCategory": "Category 3"
    },
    {
        "OrderID": "4",
        "Quantity": 20,
        "UnitPrice": 400,
        "DiscountCategory": "Category 4"
    },
    {
        "OrderID": "5",
        "Quantity": 25,
        "UnitPrice": 500,
        "DiscountCategory": "Category 5"
    }
]
```

Lab

- Step 7. Go to Synapse check the results.



```
select * from Orders;
```

OrderID	Quantity	UnitPrice	DiscountCateg...
2	10	200.00	Category 2
2	10	200.00	Category 2
1	5	120.00	Category 1
4	20	400.00	Category 4
5	25	500.00	Category 5
1	5	120.00	Category 1
4	20	400.00	Category 4
4	20	400.00	Category 4
3	15	300.00	Category 3

License