Student Name: Iliadis Viktoras
Student ID: 8180026

Big Data Management Systems
Assignment 4 - Data Streams

All the code and files can be found in the following repository: repository, similar to the previous assignments. Although some of the outputs will not be showcased in this report due to their size, they will be included in the deliverable to demonstrate the query results.

Configuration

To set up and configure Azure Stream Analytics, I followed the instructions provided in the slides. For the purpose of this presentation, I will briefly explain some of the steps I undertook:

- **Create a trial account at Azure**: I visited the Azure website and signed up for a trial account, providing my necessary details like email address, phone number, and credit card information.
- **Setup an Event Hub**: In the Azure portal, I created a new Event Hub under the Azure Event Hubs namespace. I provided a unique name for the Event Hub and configured its properties according to the instructions.
- **Generate a Security Access Signature**: I downloaded the RedDog tool from GitHub and used it to generate a Security Access Signature (SAS) for my Event Hub
- **Edit Generator.html and update the CONFIG variables**: I opened the Generator.html file in a text editor (VS Code) and updated the CONFIG variables with my SAS.
- **Feed the Event Hub with the use of Generator.html**: I opened the Generator.html file in a web browser and clicked the "Send Data" button to feed data into my Event Hub.
- **Setup a Storage account**: In the Azure portal, I created a new Storage account, providing a unique name and configuring its properties according to the instructions.
- **Upload the Reference Data files to your storage account**: I uploaded the necessary Reference Data files (Customer.json, Atm.json, Area.json) to my Storage account using the Azure portal.
- **Setup a Stream Analytics Job**: In the Azure portal, I created a new Stream Analytics job.
- **Use the Event Hub + Reference Data Files as Input**: In the Stream Analytics job, I added 4 separate inputs, 1 that uses the Event Hub and 1 for each Reference Data file from my Storage account.
- **Create a Blob Storage Output**:In the Stream Analytics job, I added an output that writes data to a Blob Storage in my Storage account.

Queries

Below are the queries one to eight :

Query 1

Query code:

```
SELECT
    SUM(Amount) AS TotalAmount
INTO
    output
FROM
    input
WHERE
    Type = 0
    AND ATMCode = 21
GROUP BY
    SlidingWindow(minute, 10)
```

Output

```
[{"TotalAmount":884}]
```

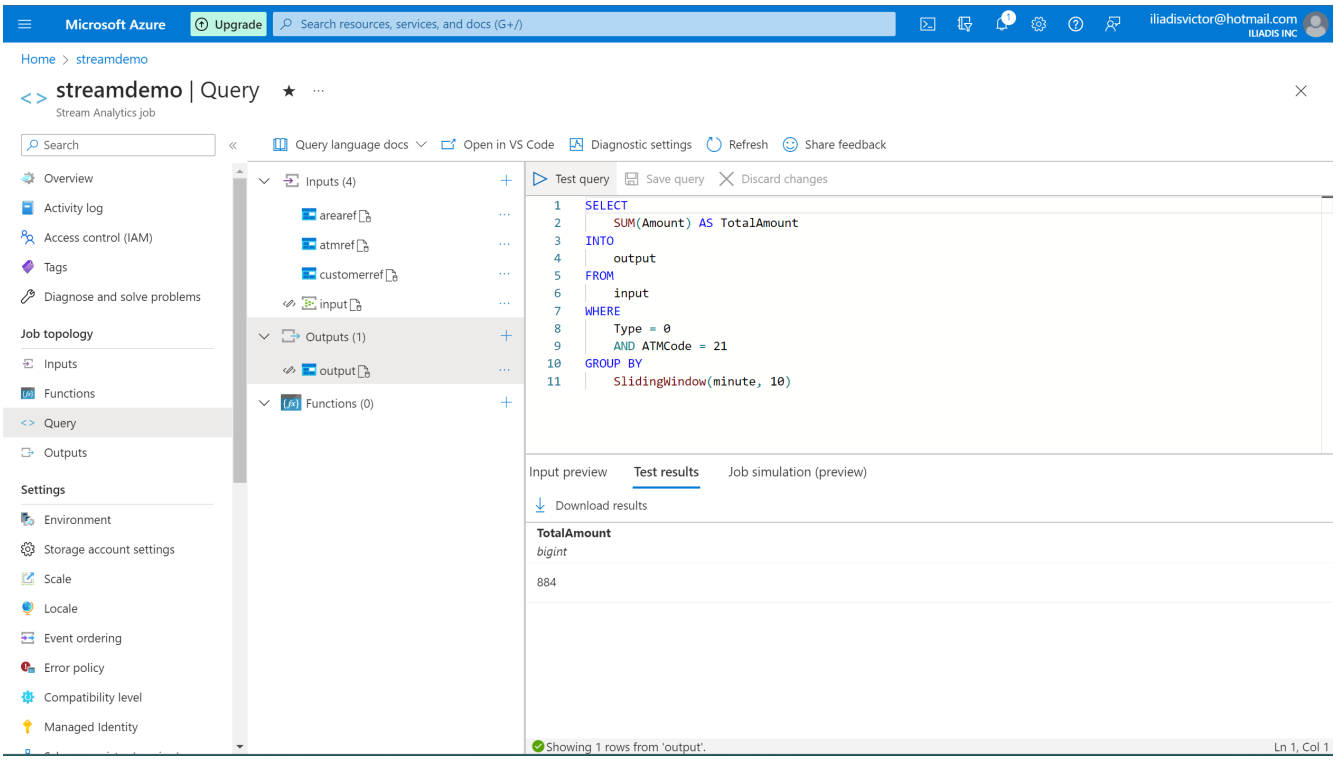

Figure 1: Query 1

Query 2

Query code:

```
SELECT
    SUM(Amount) AS TotalAmount
INTO
    output
FROM
    input
WHERE
    Type = 1
    AND ATMCode = 21
GROUP BY
    TumblingWindow(hour, 1)
```

Output

```
[{"TotalAmount":1025}]
```

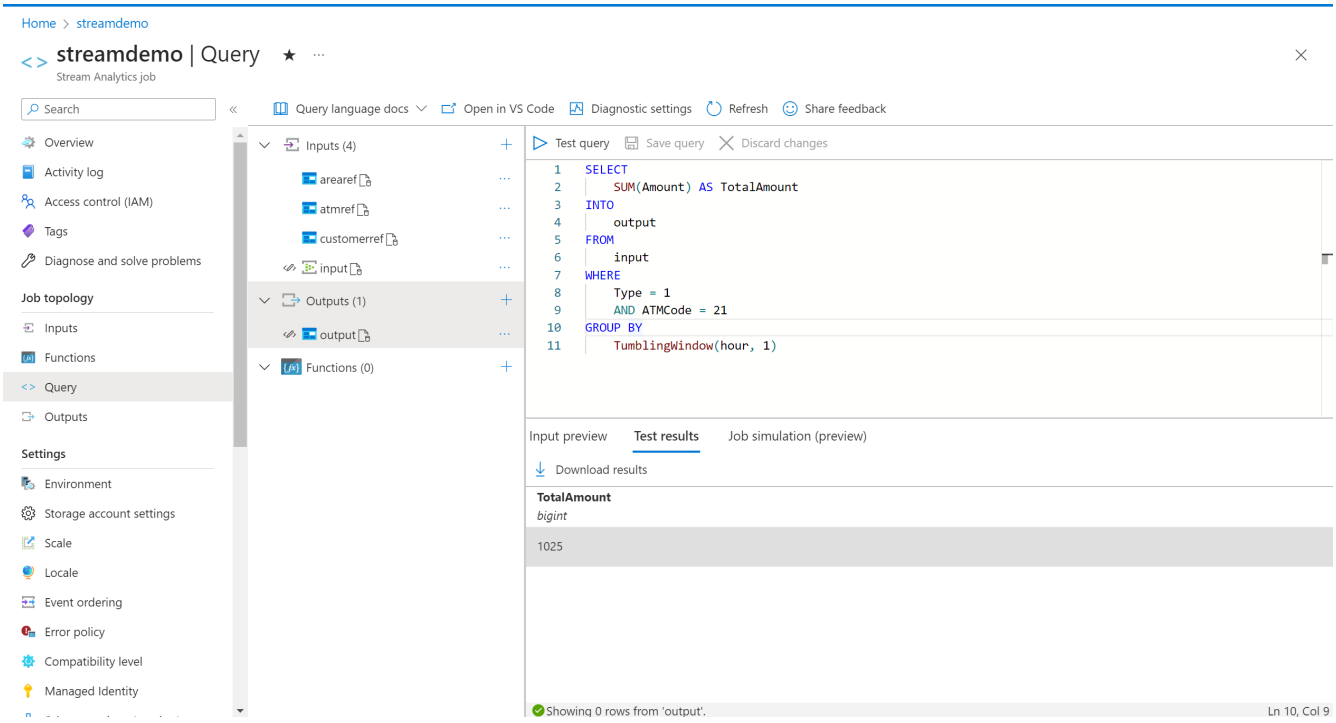

Figure 2: Query 2

Query 3

Query code:

```
SELECT
    SUM(Amount) AS TotalAmount
INTO
    output
FROM
    input
WHERE
    Type = 1
    AND ATMCode = 21
GROUP BY
    HoppingWindow(minute, 60, 30)
```

Output

```
[{"TotalAmount":1025},{"TotalAmount":1025}]
```

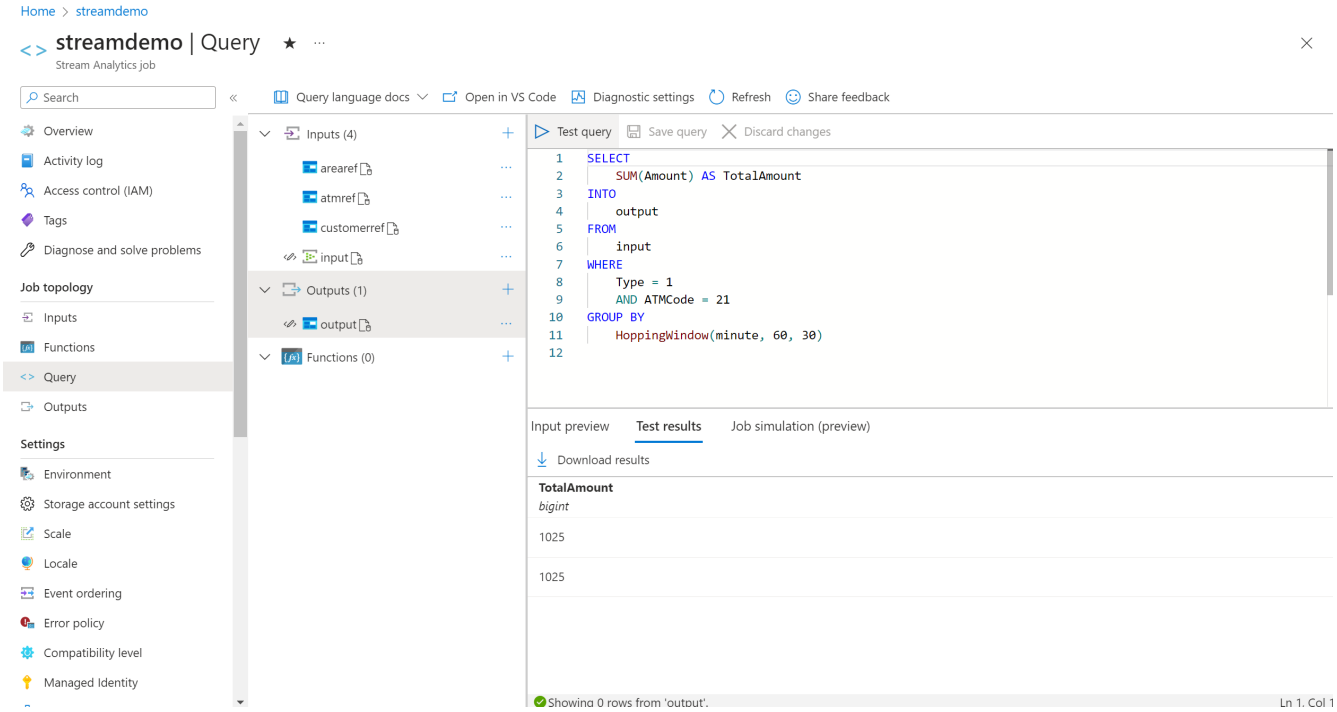

Figure 3: Query 3

Query 4

Query code:

```
SELECT
    ATMCode,
    SUM(Amount) AS TotalAmount
INTO
    output
FROM
    input
WHERE
    Type = 1
GROUP BY
    ATMCode,
    SlidingWindow(hour, 1)
```

Output


```
[{"ATMCode":20,"TotalAmount":1650},
{"ATMCode":15,"TotalAmount":2338},
{"ATMCode":18,"TotalAmount":1359},
{"ATMCode":16,"TotalAmount":358},
{"ATMCode":19,"TotalAmount":1760},
{"ATMCode":10,"TotalAmount":2384},
{"ATMCode":17,"TotalAmount":537},
{"ATMCode":13,"TotalAmount":688},
{"ATMCode":12,"TotalAmount":784},
{"ATMCode":21,"TotalAmount":1025},
{"ATMCode":4,"TotalAmount":16},
{"ATMCode":8,"TotalAmount":42},
{"ATMCode":7,"TotalAmount":18},
{"ATMCode":11,"TotalAmount":44}]
```


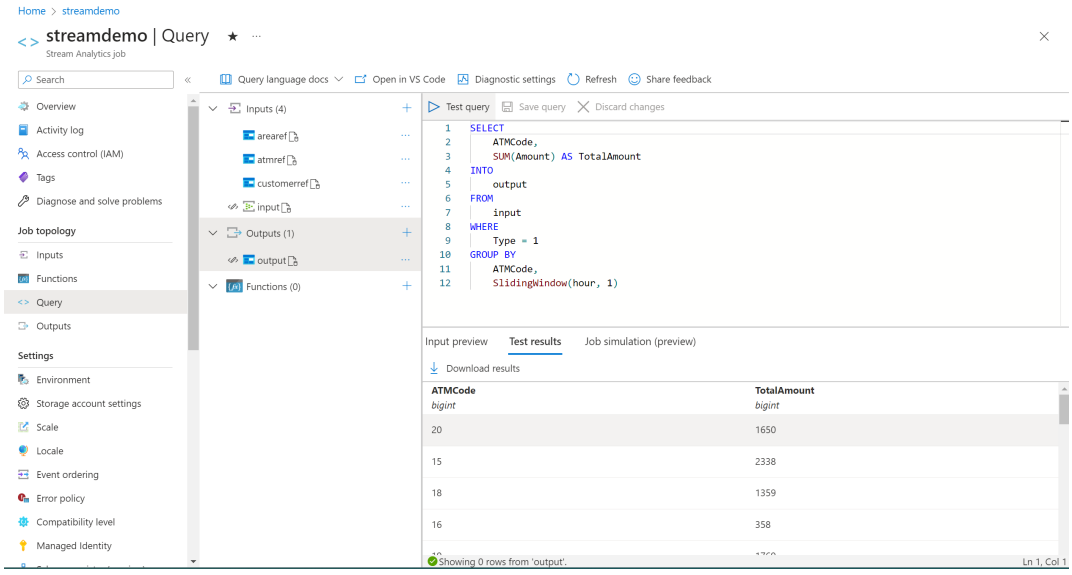

Figure 4: Query 4

Query 5

Query code:

```
SELECT
    atmref.area_code AS AreaCode,
    SUM(input.Amount) AS TotalAmount
INTO
    output
FROM
    input
JOIN
    atmref
ON
    input.ATMCode = atmref.atm_code
WHERE
    input.Type = 1
GROUP BY
    atmref.area_code,
    TumblingWindow(hour, 1)
```

Output


```
[{"AreaCode":8,"TotalAmount":44},
{"AreaCode":5,"TotalAmount":2806},
{"AreaCode":19,"TotalAmount":16},
{"AreaCode":14,"TotalAmount":18},
{"AreaCode":1,"TotalAmount":3126},
{"AreaCode":3,"TotalAmount":537},
{"AreaCode":2,"TotalAmount":2292},
{"AreaCode":7,"TotalAmount":518},
{"AreaCode":11,"TotalAmount":2384},
{"AreaCode":4,"TotalAmount":1947},
{"AreaCode":13,"TotalAmount":42},
{"AreaCode":10,"TotalAmount":688},
{"AreaCode":9,"TotalAmount":784}]
```

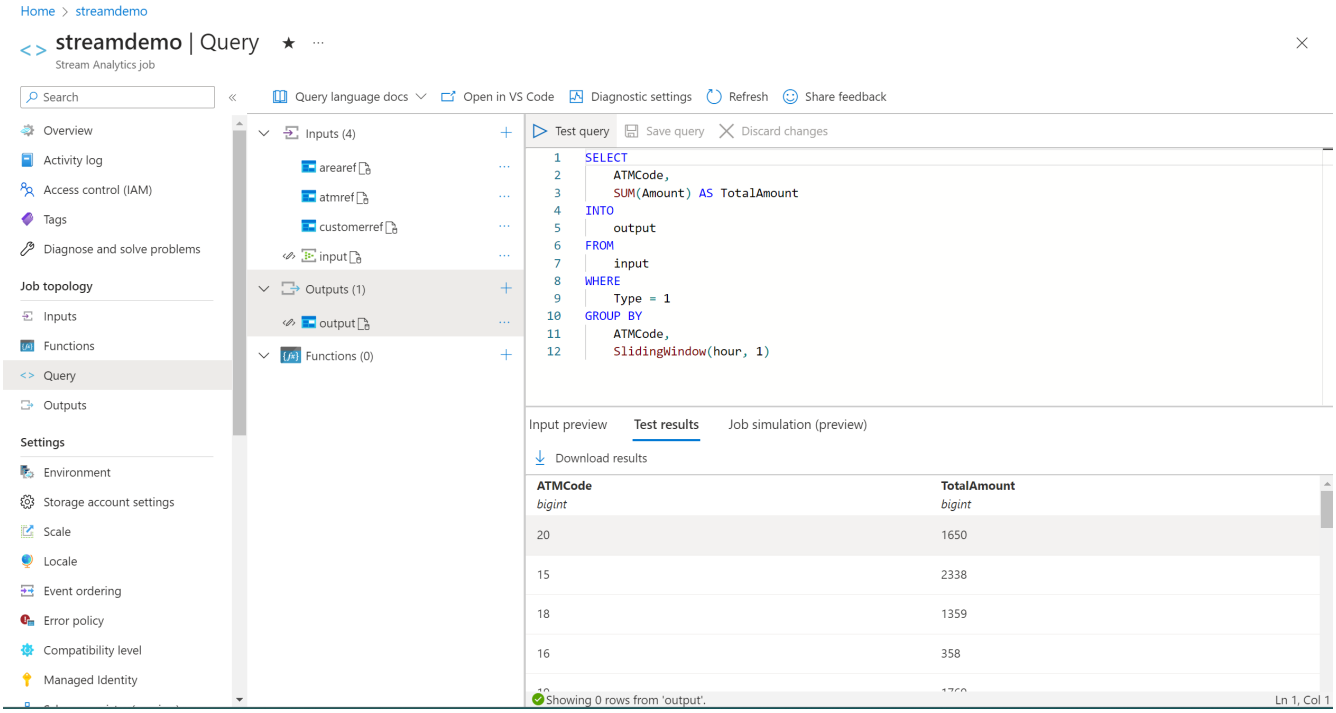

Figure 5: Query 5

Query 6

Query code:

```
SELECT
    arearef.area_city AS City,
    customerref.gender AS Gender,
    SUM(input.Amount) AS TotalAmount
INTO
    output
FROM
    input
JOIN
    customerref
ON
    input.CardNumber = customerref.card_number
JOIN
    atmref
ON
    input.ATMCode = atmref.atm_code
JOIN
    arearef
ON
    atmref.area_code = arearef.area_code
GROUP BY
    arearef.area_city,
    customerref.gender,
    TumblingWindow(hour, 1)
```

Output


```
[
  {"City":"Schaumburg","Gender":"Female","TotalAmount":4307},
  {"City":"Baltimore","Gender":"Male","TotalAmount":627},
  {"City":"Omaha","Gender":"Female","TotalAmount":931},
  {"City":"Tacoma","Gender":"Male","TotalAmount":21},
  {"City":"Memphis","Gender":"Male","TotalAmount":2913},
  {"City":"Tacoma","Gender":"Female","TotalAmount":1207},
  {"City":"Greeley","Gender":"Male","TotalAmount":45},
  {"City":"Memphis","Gender":"Female","TotalAmount":1132},
  {"City":"Vancouver","Gender":"Female","TotalAmount":17},
  {"City":"Springfield","Gender":"Male","TotalAmount":3728},
  {"City":"Canton","Gender":"Female","TotalAmount":45},
  {"City":"Canton","Gender":"Male","TotalAmount":1625},
  {"City":"Schaumburg","Gender":"Male","TotalAmount":1383},
  {"City":"Vancouver","Gender":"Male","TotalAmount":1027},
  {"City":"Springfield","Gender":"Female","TotalAmount":1959},
  {"City":"Greeley","Gender":"Female","TotalAmount":44},
  {"City":"Omaha","Gender":"Male","TotalAmount":3231},
  {"City":"Dayton","Gender":"Male","TotalAmount":14},
  {"City":"Baltimore","Gender":"Female","TotalAmount":576}
]
```

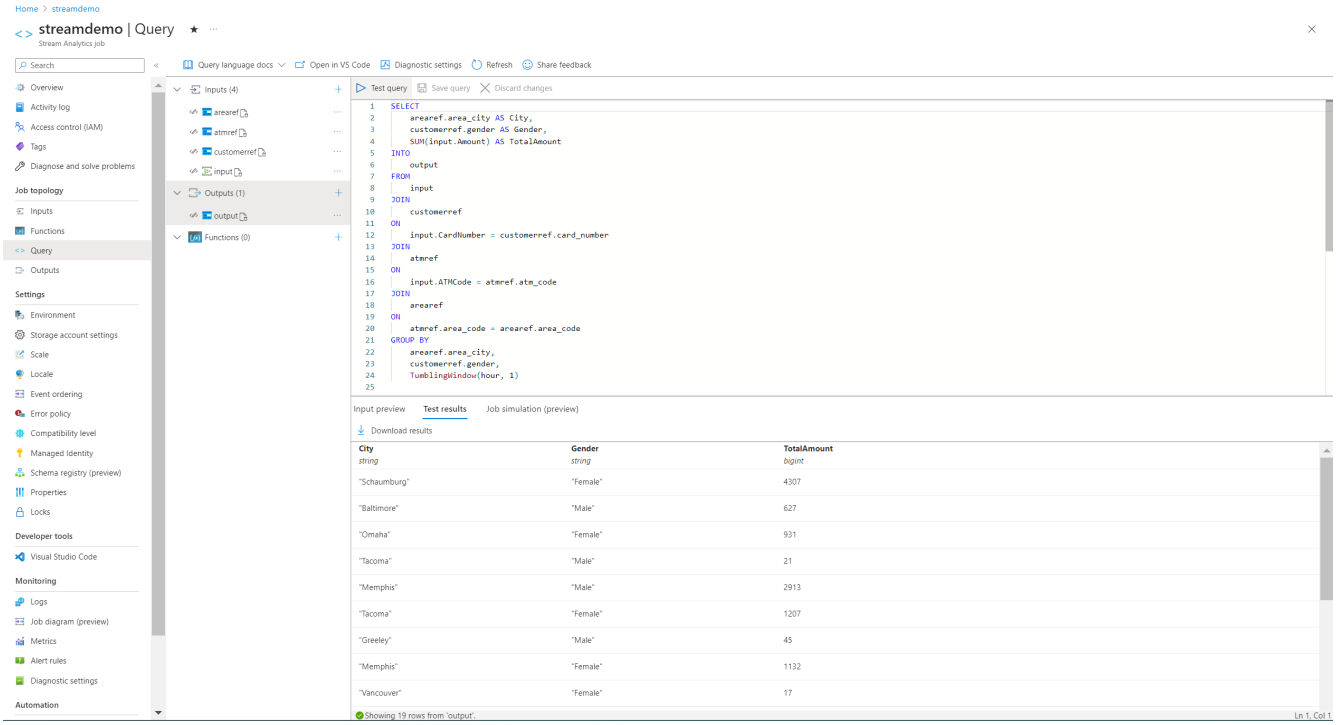

Figure 6: Query 6

Query 7

Query code:

```
SELECT
    1 AS Alert
INTO
    output
FROM
    (
        SELECT
            CardNumber,
            COUNT(*) AS TransactionCount
        FROM
            input
        WHERE
            Type = 1
        GROUP BY
            CardNumber,
            SlidingWindow(hour, 1)
    ) AS Transactions
WHERE
    TransactionCount >= 2
```

Output(20 Alerts)


```
[[{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
```

```
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1}]]
........
```


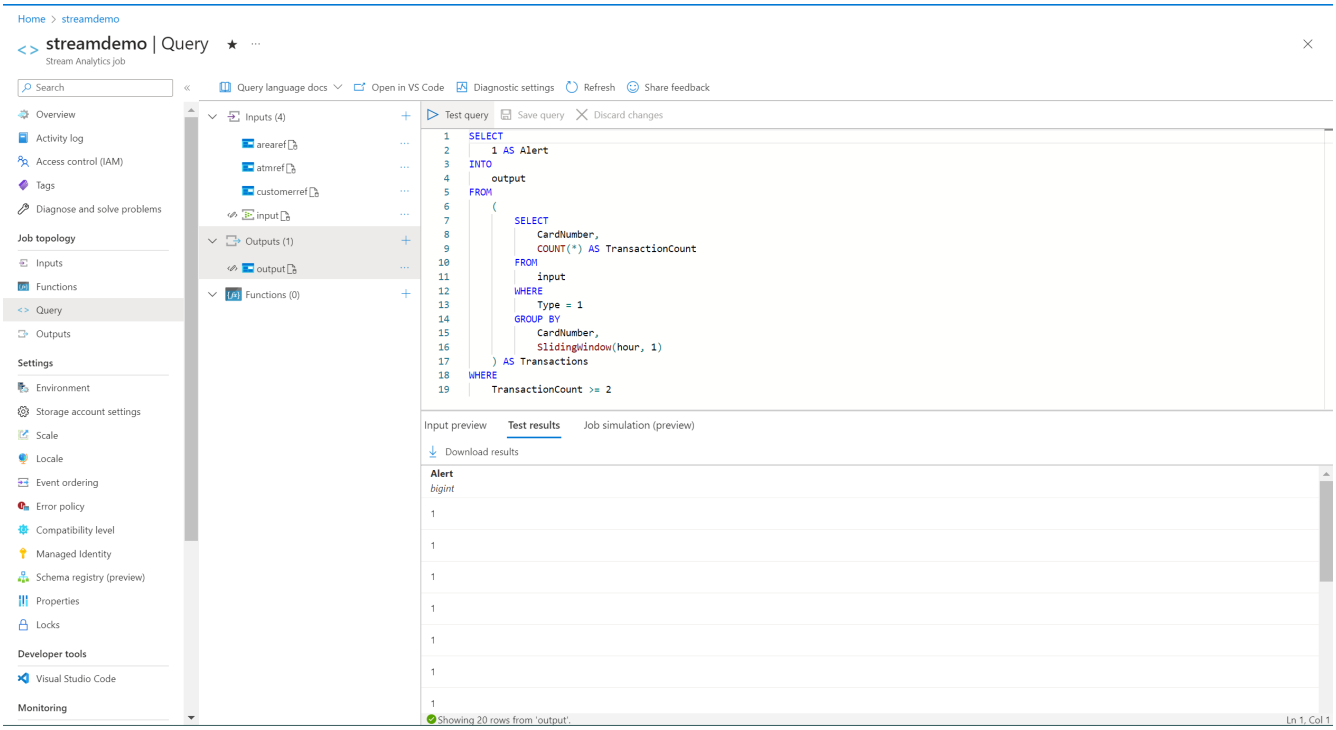

Figure 7: Query 7

Query 8

Query code:

```
SELECT
    1 AS Alert
INTO
    output
FROM
    input
JOIN
    atmref
ON
    input.ATMCode = atmref.atm_code
JOIN
    customerref
ON
    input.CardNumber = customerref.card_number
WHERE
    atmref.area_code <> customerref.area_code
```

Output: (879 alerts)


```
[{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
...
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1},
{"Alert":1}]
.........
```

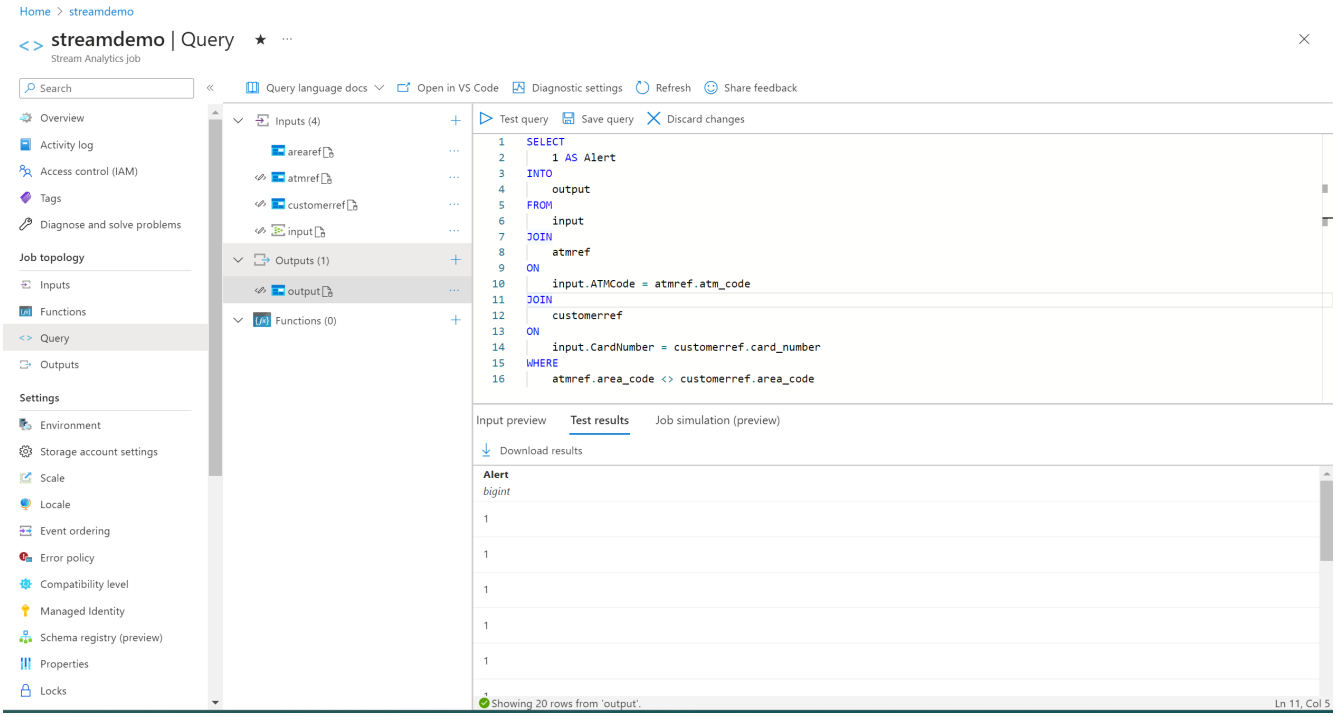

Figure 8: Query 8