

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:

```
1 SELECT
2     SUM(Amount) AS TotalAmount
3 INTO
4     output
5 FROM
6     input
7 WHERE
8     Type = 0
9     AND ATMCCode = 21
10 GROUP BY
11     SlidingWindow(minute, 10)
```

Output

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

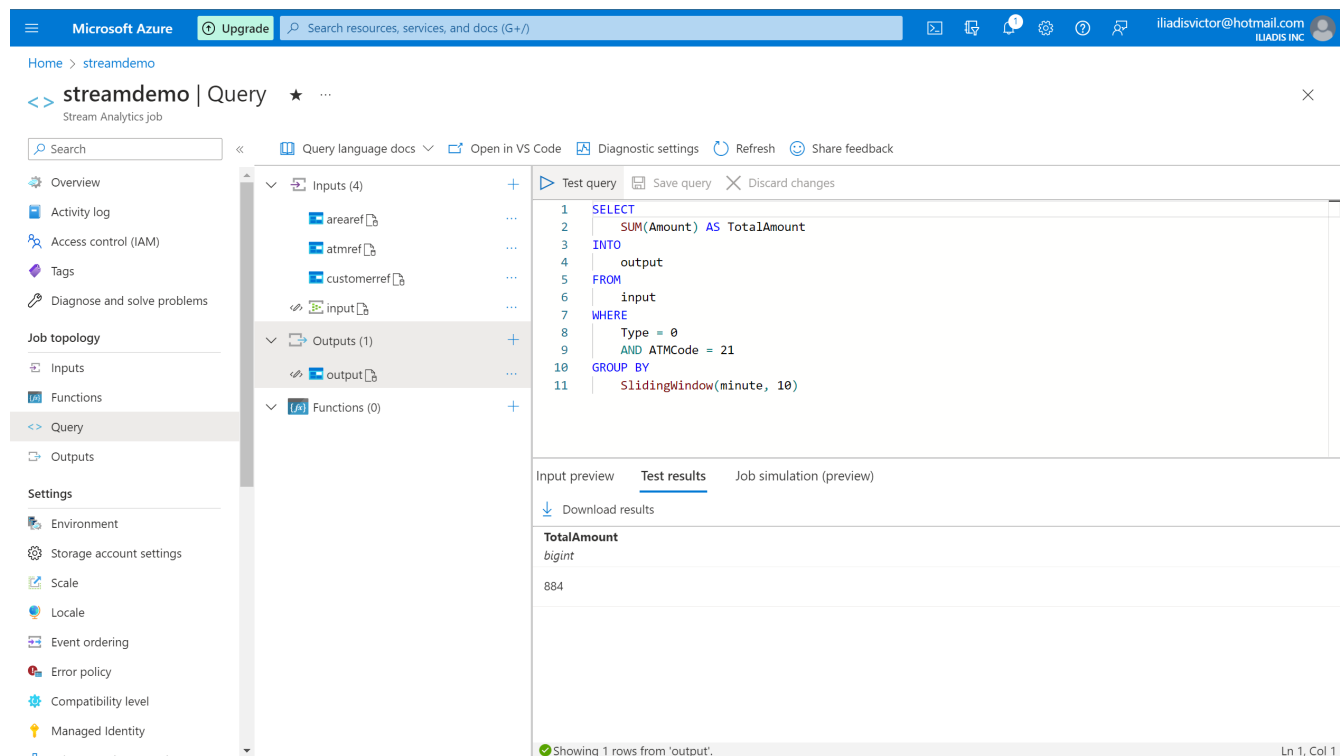


Figure 1: Query 1

Query 2

Query code:

```
1 SELECT
2     SUM(Amount) AS TotalAmount
3 INTO
4     output
5 FROM
6     input
7 WHERE
8     Type = 1
9     AND ATMCCode = 21
10 GROUP BY
11     TumblingWindow(hour, 1)
```

Output

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

The screenshot shows the Azure Stream Analytics portal interface for a job named 'streamdemo'. The left sidebar contains navigation options: Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Job topology, Inputs, Functions, Query (selected), Outputs, and Settings. The main area is divided into three sections: Inputs (4), Outputs (1), and Functions (0). The 'Test query' tab is active, displaying the following SQL query:

```
1 SELECT
2     SUM(Amount) AS TotalAmount
3 INTO
4     output
5 FROM
6     input
7 WHERE
8     Type = 1
9     AND ATMCCode = 21
10 GROUP BY
11     TumblingWindow(hour, 1)
```

Below the query editor, the 'Test results' tab is selected, showing a single output row:

TotalAmount
1025

At the bottom, a status bar indicates 'Showing 0 rows from 'output'.' and 'Ln 10, Col 9'.

Figure 2: Query 2

Query 3

Query code:

```
1 SELECT
2     SUM(Amount) AS TotalAmount
3 INTO
4     output
5 FROM
6     input
7 WHERE
8     Type = 1
9     AND ATMCCode = 21
10 GROUP BY
11     HoppingWindow(minute, 60, 30)
```

Output

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

The screenshot shows the Azure Stream Analytics portal interface for a job named 'streamdemo'. The 'Query' tab is active, displaying the following SQL query:

```
1 SELECT
2     SUM(Amount) AS TotalAmount
3 INTO
4     output
5 FROM
6     input
7 WHERE
8     Type = 1
9     AND ATMCCode = 21
10 GROUP BY
11     HoppingWindow(minute, 60, 30)
```

The 'Test results' tab shows the output of the query as a table with two rows of 'TotalAmount' values (1025).

TotalAmount
1025
1025

The interface also includes a left sidebar with navigation options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Job topology, Inputs, Functions, Query, Outputs, and Settings. The bottom status bar indicates 'Showing 0 rows from 'output'.'

Figure 3: Query 3

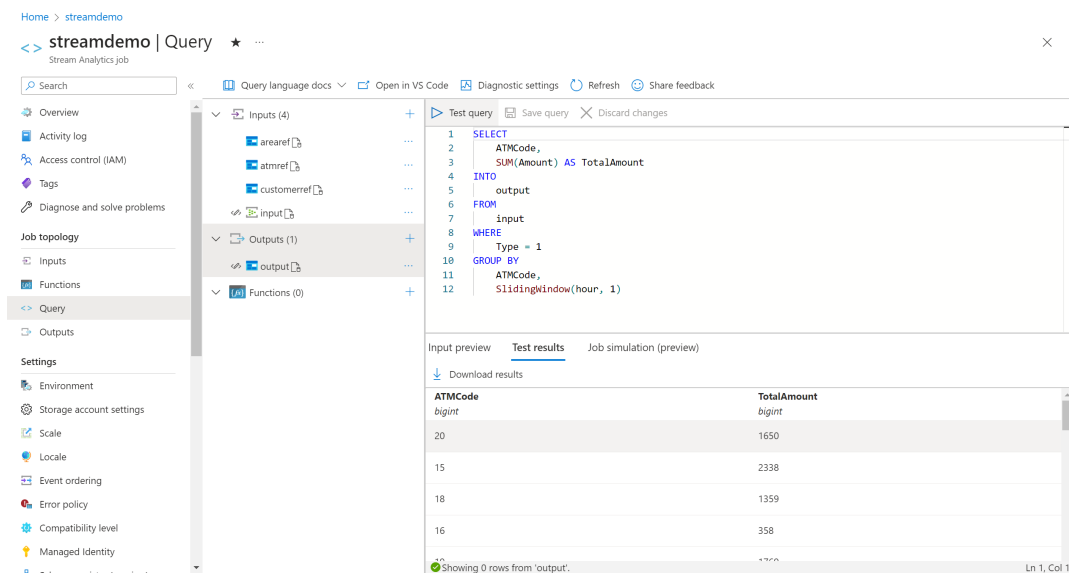
Query 4

Query code:

```
1 SELECT
2     ATMCode,
3     SUM(Amount) AS TotalAmount
4 INTO
5     output
6 FROM
7     input
8 WHERE
9     Type = 1
10 GROUP BY
11     ATMCode,
12     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}]
```



The screenshot shows the Azure Stream Analytics portal interface. On the left, there's a navigation pane with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Job topology, Inputs, Functions, <> Query, Outputs, and Settings. The main area displays the query code for 'streamdemo | Query'. The query is a SELECT statement that calculates the sum of 'Amount' for each 'ATMCode' over a sliding window of 1 hour, and stores the results in an output named 'output'. Below the query editor, there's a section for 'Test results' showing a table with the following data:

ATMCode	TotalAmount
20	1650
15	2338
18	1359
16	358

The table shows 13 rows of data in total. The status bar at the bottom indicates 'Showing 0 rows from 'output''.

Figure 4: Query 4

Query 5

Query code:

```
1 SELECT
2     atmref.area_code AS AreaCode ,
3     SUM(input.Amount) AS TotalAmount
4 INTO
5     output
6 FROM
7     input
8 JOIN
9     atmref
10 ON
11     input.ATMCode = atmref.atm_code
12 WHERE
13     input.Type = 1
14 GROUP BY
15     atmref.area_code ,
16     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}]
```

Home > streamdemo

<> streamdemo | Query ★ ...

Stream Analytics job

Search

Query language docs Open in VS Code Diagnostic settings Refresh Share feedback

Test query Save query Discard changes

```
1 SELECT
2   ATMCode,
3   SUM(Amount) AS TotalAmount
4 INTO
5   output
6 FROM
7   input
8 WHERE
9   Type = 1
10 GROUP BY
11   ATMCode,
12   SlidingWindow(hour, 1)
```

Input preview Test results Job simulation (preview)

Download results

ATMCode	TotalAmount
<i>bigint</i>	<i>bigint</i>
20	1650
15	2338
18	1359
16	358
17	1700

Showing 0 rows from 'output'. Ln 1, Col 1

Figure 5: Query 5

Query 6

Query code:

```
1 SELECT
2     arearef.area_city AS City ,
3     customerref.gender AS Gender ,
4     SUM(input.Amount) AS TotalAmount
5 INTO
6     output
7 FROM
8     input
9 JOIN
10    customerref
11 ON
12    input.CardNumber = customerref.card_number
13 JOIN
14    atmref
15 ON
16    input.ATMCode = atmref.atm_code
17 JOIN
18    arearef
19 ON
20    atmref.area_code = arearef.area_code
21 GROUP BY
22     arearef.area_city ,
23     customerref.gender ,
24     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}
]
```


Home > streamdemo

<> streamdemo | Query ★ ...

Stream Analytics job

Query language docs Open in VS Code Diagnostic settings Refresh Share feedback

Search

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Job topology

Inputs

Functions

<> Query

Outputs

Settings

Environment

Storage account settings

Scale

Locale

Event ordering

Error policy

Compatibility level

Managed Identity

Schema registry (preview)

Properties

Locks

Developer tools

Visual Studio Code

Monitoring

Logs

Job diagram (preview)

Metrics

Alert rules

Diagnostic settings

Automation

Test query Save query Discard changes

```

1 SELECT
2   arearef.area_city AS City,
3   customerref.gender AS Gender,
4   SUM(input.Amount) AS TotalAmount
5 INTO
6   output
7 FROM
8   input
9 JOIN
10  customerref
11 ON
12  input.CardNumber = customerref.card_number
13 JOIN
14  arearef
15 ON
16  input.ATMCode = arearef.atm_code
17 JOIN
18  arearef
19 ON
20  arearef.area_code = arearef.area_code
21 GROUP BY
22  arearef.area_city,
23  customerref.gender,
24  TumblingWindow(hour, 1)
25

```

Input preview Test results Job simulation (preview)

Download results

City	Gender	TotalAmount
string	string	bigint
"Schauenburg"	"Female"	4307
"Baltimore"	"Male"	627
"Omaha"	"Female"	931
"Tacoma"	"Male"	21
"Memphis"	"Male"	2913
"Tacoma"	"Female"	1207
"Greeley"	"Male"	45
"Memphis"	"Female"	1132
"Vancouver"	"Female"	17

Showing 19 rows from 'output'.

Ln 1, Col 1

Figure 6: Query 6

Query 7

Query code:

```

1 SELECT
2   1 AS Alert
3 INTO
4   output
5 FROM
6   (
7     SELECT
8       CardNumber,
9       COUNT(*) AS TransactionCount
10    FROM
11      input
12    WHERE
13      Type = 1
14    GROUP BY
15      CardNumber,
16      SlidingWindow(hour, 1)
17  ) AS Transactions
18 WHERE
19   TransactionCount >= 2

```

Output(20 Alerts)

```

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

```

Query language tools

Open in VS Code

Diagnostic settings

Refresh

Share feedback

Inputs (4)

arearef

atmref

customerref

input

Outputs (1)

output

Functions (0)

Test query

Save query

Discard changes

```

1 SELECT
2     1 AS Alert
3 INTO
4     output
5 FROM
6     (
7         SELECT
8             CardNumber,
9             COUNT(*) AS TransactionCount
10        FROM
11            input
12        WHERE
13            Type = 1
14        GROUP BY
15            CardNumber,
16            SlidingWindow(hour, 1)
17        ) AS Transactions
18 WHERE
19     TransactionCount >= 2

```

Input preview

Test results

Job simulation (preview)

Download results

Alert
bigint
1
1
1
1
1
1
1

Showing 20 rows from 'output'.

Ln 1, Col

Figure 7: Query 7

Query 8

Query code:

```
1 SELECT
2     1 AS Alert
3 INTO
4     output
5 FROM
6     input
7 JOIN
8     atmref
9 ON
10    input.ATMCode = atmref.atm_code
11 JOIN
12    customerref
13 ON
14    input.CardNumber = customerref.card_number
15 WHERE
16    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},
{"Alert":1}]
.....
```

Home > streamdemo

<> streamdemo | Query ★ ...

Stream Analytics job

Search

Query language docs Open in VS Code Diagnostic settings Refresh Share feedback

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Job topology

Inputs

Functions

Query

Outputs

Settings

Environment

Storage account settings

Scale

Locale

Event ordering

Error policy

Compatibility level

Managed Identity

Schema registry (preview)

Properties

Locks

Inputs (4)

- arearef
- atmref
- customerref
- input

Outputs (1)

- output

Functions (0)

Test query Save query Discard changes

```
1 SELECT
2   1 AS Alert
3 INTO
4   output
5 FROM
6   input
7 JOIN
8   atmref
9 ON
10  input.ATMCode = atmref.atm_code
11 JOIN
12  customerref
13 ON
14  input.CardNumber = customerref.card_number
15 WHERE
16  atmref.area_code <> customerref.area_code
```

Input preview Test results Job simulation (preview)

Download results

Alert
bigint
1
1
1
1
1
1

Showing 20 rows from 'output'. Ln 11, Col 5

Figure 8: Query 8