

Assignment 5

Panourgia Evangelia (t8190130)
Papadatos Ioannis (t8190314)
Professor: Chatziantoniou Damianos

Latest Update: May 8, 2022

School of Management Science and Technology,
Athens University of Economics and Business

Contents

0. Introduction.....	3
1. Create a trial account.....	3
2. Setup an Event Hub.....	3
3. Generate a Security Access Signature.....	4
4. Edit Generator.html.....	5
5. Feed the Event Hub.....	5
6. Setup a Storage account.....	5
7. Upload the Reference Data files to your storage account.....	7
8. Setup a Stream Analytics Job.....	8
9. Input setup.....	8
10. Output setup.....	10

II Queries

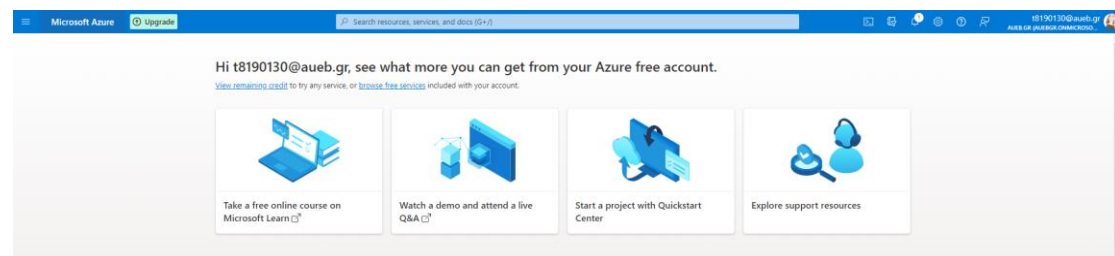
Queries.....	10
Query 1.....	11
Query 2.....	12
Query 3.....	13
Query 4	14
Query 5.....	15
Query 6.....	16
Query 7.....	17
Query 8	18
11.Conclusion.....	19
12. References.....	19

0.Introduction

Azure Stream Analytics is a fully managed, real-time analytics service designed to help you analyze and process fast moving streams of data that can be used to get insights, build reports or trigger alerts and actions [1]. For this assignment, we used Stream Analytics by setting up a processing environment and running queries using real-time randomly generated data and reference data (data exists in folder named data).

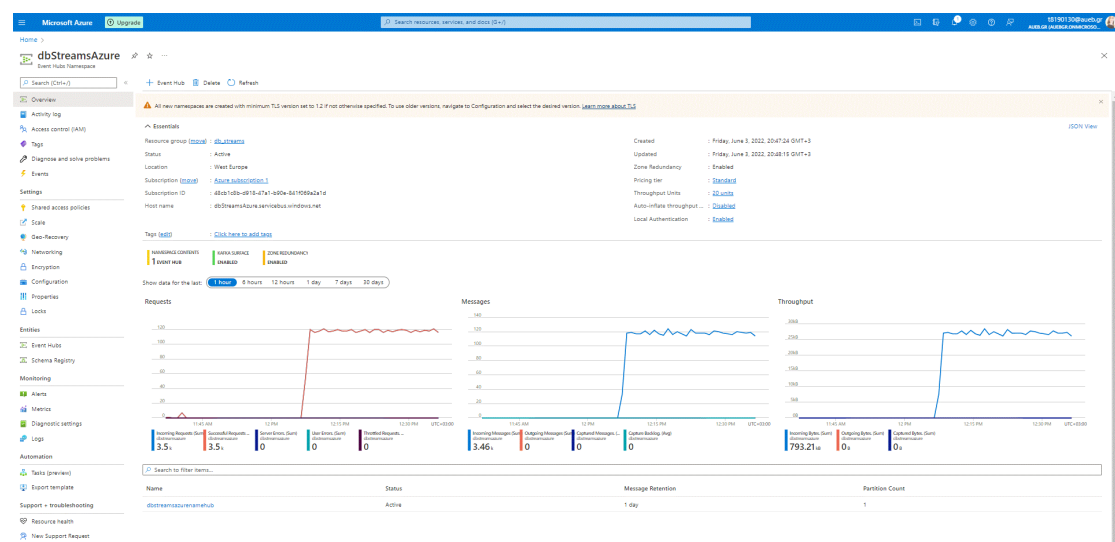
1. Create a Trial Account

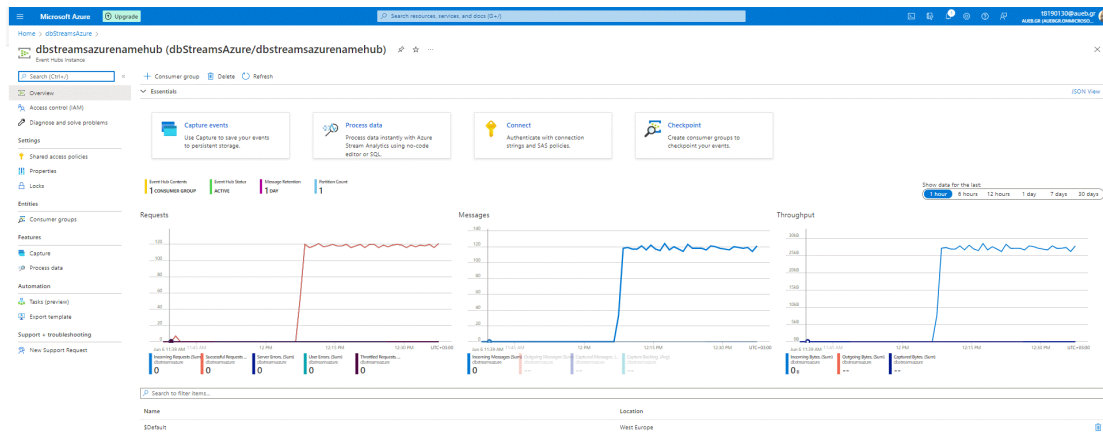
I created a Microsoft Azure account using my university e-mail address.



2. Setup an event hub

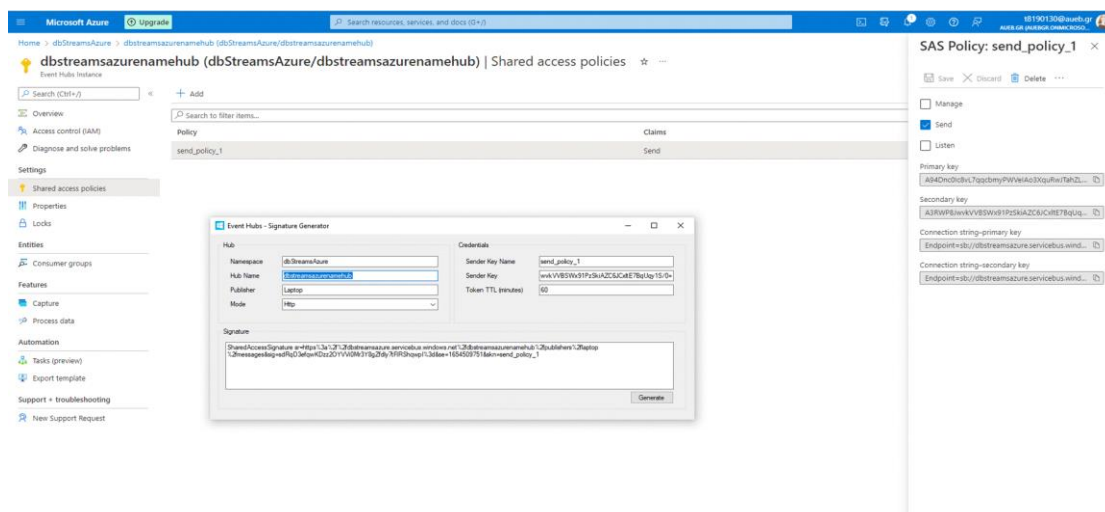
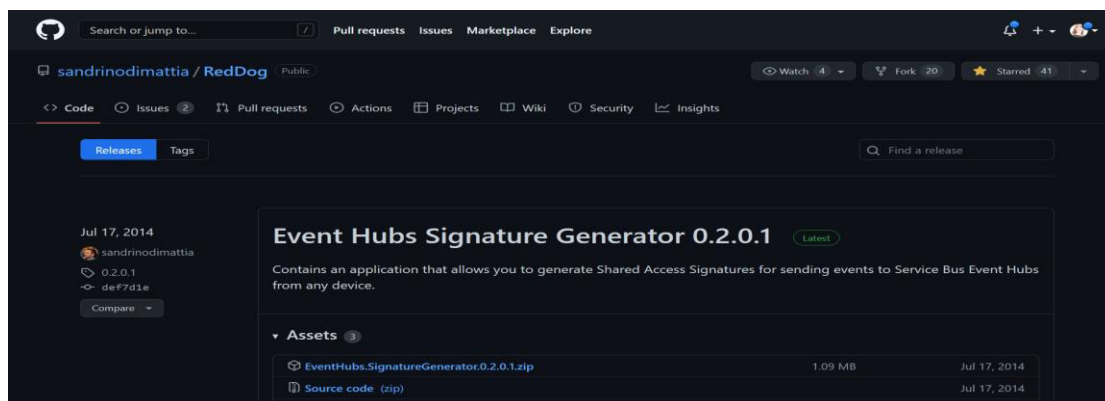
Event Hubs is a fully managed, real-time data ingestion service that's simple, trusted, and scalable [2].





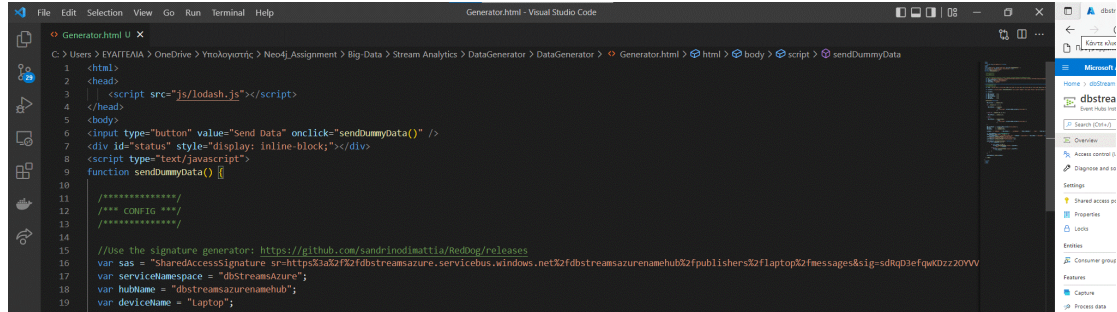
3. Generate a Security Access Signature

We generated a security access signature. Specifically, we visited GitHub repository named: <https://github.com/sandrinodimattia/RedDog/releases> which contains an application that allows you to generate Shared Access Signatures for sending events to Service Bus Event Hubs from any device. So, we download [EventHubs.SignatureGenerator.0.2.0.1.zip](#) (unzip it) and then we returned to azure interface again and created the security access signature as shows the following pictures.



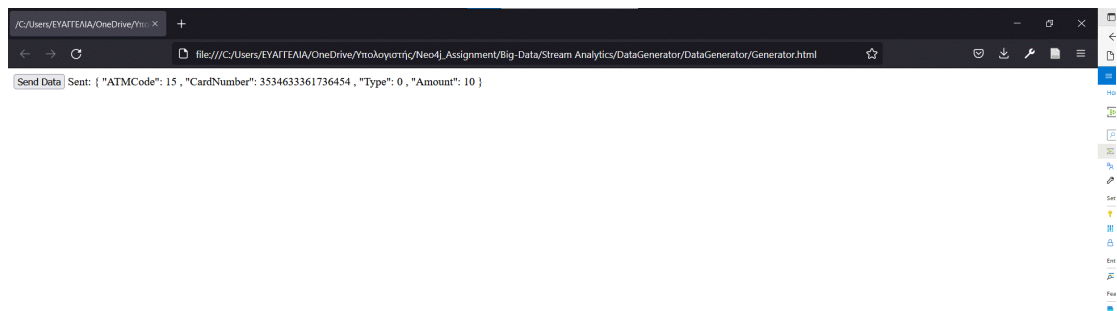
4. Edit Generator.html

We edited `Generator.html`. We opened it with visual studio and we updated the `CONFIG` variables with our security access signature bearing in mind the suitable variables values that we created in step 3 (the previous step).



5. Feed the Event Hub

We fed the Event Hub with the use of `Generator.html`. Specifically, we opened `Generator.html` in a web browser and pressed the “Send Data” button.



6. Set up a Storage account

An Azure storage account contains all of your Azure Storage data objects, including blobs, file shares, queues, tables, and disks. The storage account provides a unique namespace for your Azure Storage data that's accessible from anywhere in the world over HTTP or HTTPS. Data in your storage account is durable and highly available, secure, and massively scalable [3]. For the assignment a storage account was created in order to store the reference data to be used in the queries, as well as their output.

Microsoft Azure Upgrade Search resources, services, and docs (G+J) 18190130@auweb.gr AUWEB.GR (AUWEBGR.ONMICROSOFT...)

Home > Storage accounts >

Create a storage account

Basics Advanced Networking Data protection Encryption Tags Review + create

Create new

Instance details

If you need to create a legacy storage account type, please click [here](#).

Storage account name

Region

Performance ☒ Standard: Recommended for most scenarios (general-purpose v2 account)
☐ Premium: Recommended for scenarios that require low latency.

Redundancy
☒ Make read access to data available in the event of regional unavailability.

Review + create < Previous Next: Advanced >

Microsoft Azure Upgrade Search resources, services, and docs (G+J) 18190130@auweb.gr AUWEB.GR (AUWEBGR.ONMICROSOFT...)

Home > Storage accounts >

Create a storage account

Running final validation...

Basics Advanced Networking Data protection Encryption Tags Review + create

Basics

Subscription	Azure subscription 1
Resource Group	db_streams
Location	westeurope
Storage account name	evastorage1
Deployment model	Resource manager
Performance	Standard
Replication	Read-access geo-redundant storage (RA-GRS)

Advanced

Secure transfer	Enabled
Allow storage account key access	Enabled
Allow cross-tenant replication	Enabled
Default to Azure Active Directory	Disabled

Microsoft Azure Upgrade Search resources, services, and docs (G+J) 18190130@auweb.gr AUWEB.GR (AUWEBGR.ONMICROSOFT...)

Home >

evastorage1_1654533299219 | Overview

Deployment

Search (Ctrl+J) Delete Cancel Redeploy Refresh

We'd love your feedback! →

Your deployment is complete

Deployment name: evastorage1_1654533299219
Subscription: Azure subscription 1
Resource group: db_streams

Start time: 6/6/2022, 7:35:10 PM
Correlation ID: 7b3e05b2-5a72-476c-bb18-706604115c4d

Deployment details (Download)
Next steps
Go to resource

Cost Management
Get notified to stay within your budget and prevent unexpected charges on your bill.
Set up cost alerts >

Microsoft Defender for Cloud
Secure your apps and infrastructure
Go to Microsoft Defender for Cloud >

7. Upload the Reference Data files to your storage account

A blob container was created and the reference JSON files were uploaded.

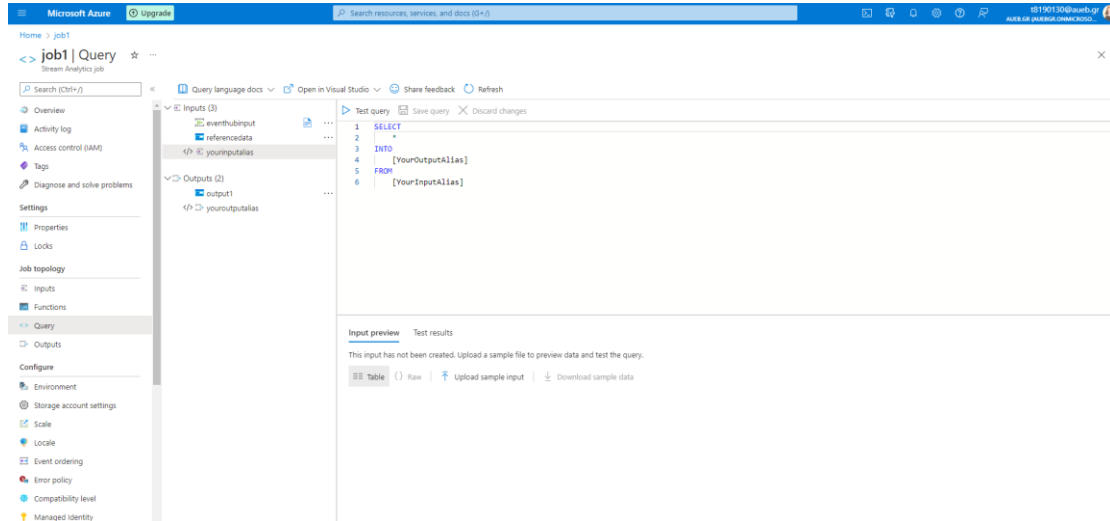
The screenshot shows the Microsoft Azure portal interface. On the left, the navigation pane is open, showing the 'evastorage1' storage account. The 'Overview' tab is selected. The main pane displays the 'Essentials' section with various properties like Resource group, Location, and Subscription. On the right, the 'Upload blob' dialog is open, showing a list of current uploads: 'Customer.json' (3 KB / 3 KB), 'Atm.json' (1 KB / 1 KB), and 'Area.json' (990 B / 990 B). The 'Upload' button is visible at the bottom of the dialog.

The screenshot shows the Microsoft Azure portal interface for the 'container1' blob container. The 'Overview' tab is selected, displaying a table of blobs. The table has columns for Name, Modified, Access tier, Archive status, Blob type, Size, and Lease state. Three blobs are listed: 'Area.json', 'Atm.json', and 'Customer.json', all with a size of 990 B and a lease state of 'Available'.

Name	Modified	Access tier	Archive status	Blob type	Size	Lease state
Area.json	6/7/2022, 12:38:10 PM	Hot (inferred)		Block blob	990 B	Available
Atm.json	6/7/2022, 12:38:10 PM	Hot (inferred)		Block blob	1,04 KB	Available
Customer.json	6/7/2022, 12:38:10 PM	Hot (inferred)		Block blob	3,23 KB	Available

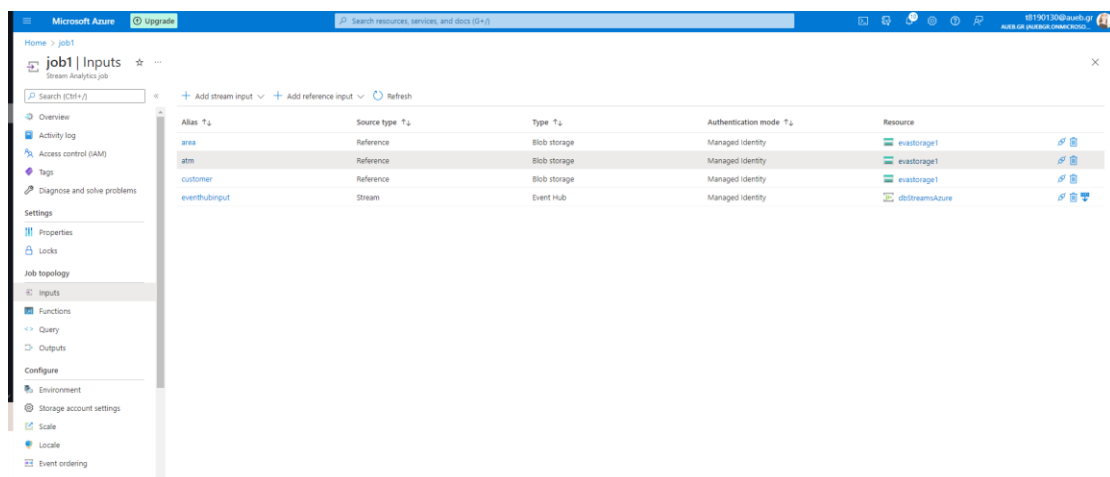
8. Setup a Stream Analytics Job

We created a stream analytics job bearing in mind the instructions from link[4].



9. Input setup

The reference data, as well as the Event Hub stream were set as inputs for the job. Successful setup of the stream input was confirmed by sampling data from it.



Microsoft Azure | Upgrade | Search resources, services, and docs (Ctrl+J)

Home > job1 | Query

Stream Analytics job

Query language docs | Open in Visual Studio | Share feedback | Refresh

Test query | Save query | Discard changes

```

1 SELECT
2 *
3 FROM
4 [YourOutputAlias]
5 FROM
6 eventhubinput

```

Input preview | Test results

Showing 22 rows from 'youroutputalias':

ATMCode	CardNumber	Type	Amount	EventProcessedUtcTime	PartitionId	EventEnqueuedUtcTime
12	5893112367133403000	1	22	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:18.49900...
10	5446210381593272	1	40	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:18.54300...
17	3549670931666297	1	13	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:17.84300...
19	520025312538103	0	32	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:17.97700...
18	5602238863017460	0	30	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:15.54600...
15	201634801435467	1	26	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:14.26300...
16	3583237214000023	1	35	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:12.84300...
17	18476700316281017	1	43	*2022-06-07T10:25:01.42693...	0	*2022-06-07T09:25:12.80000...

For the sample we changed the range as following :

Microsoft Azure | Upgrade | Search resources, services, and docs (Ctrl+J)

Home > job1 | Query

Stream Analytics job

Query language docs | Open in Visual Studio | Share feedback | Refresh

Test query | Save query | Discard changes

```

1

```

Input preview | Test results

Showing sample events from 'eventhubinput':

ATMCode	CardNumber	Type	Amount	EventProcessedUtcTime
15	3542024987623740	1	48	*2022-06-07T14:30:22.561
18	604513436397576	0	33	*2022-06-07T14:30:22.561
16	3583237214000023	0	18	*2022-06-07T14:30:22.561
13	5610627137784218	1	19	*2022-06-07T14:30:22.561
15	201634801435467	0	10	*2022-06-07T14:30:22.561
10	5446210381593272	1	21	*2022-06-07T14:30:22.561
10	560222217915588000	1	44	*2022-06-07T14:30:22.561

Sample data

eventhubinput

Sampling events from the input source in this time range will retrieve up to 1000 events or 1 MB (whichever comes first).

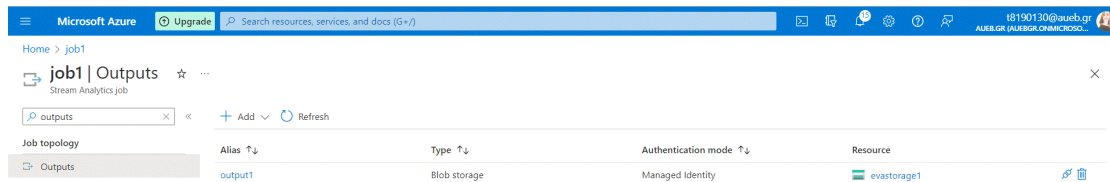
Start time: 06/07/2022 5:31:52 PM Local time (UTC+03:00)

Duration: 9 Days 0 Hours 3 Minutes 9 Seconds

Sample

10. Output setup

We created a stream analytics job bearing in mind the instructions from link[4].

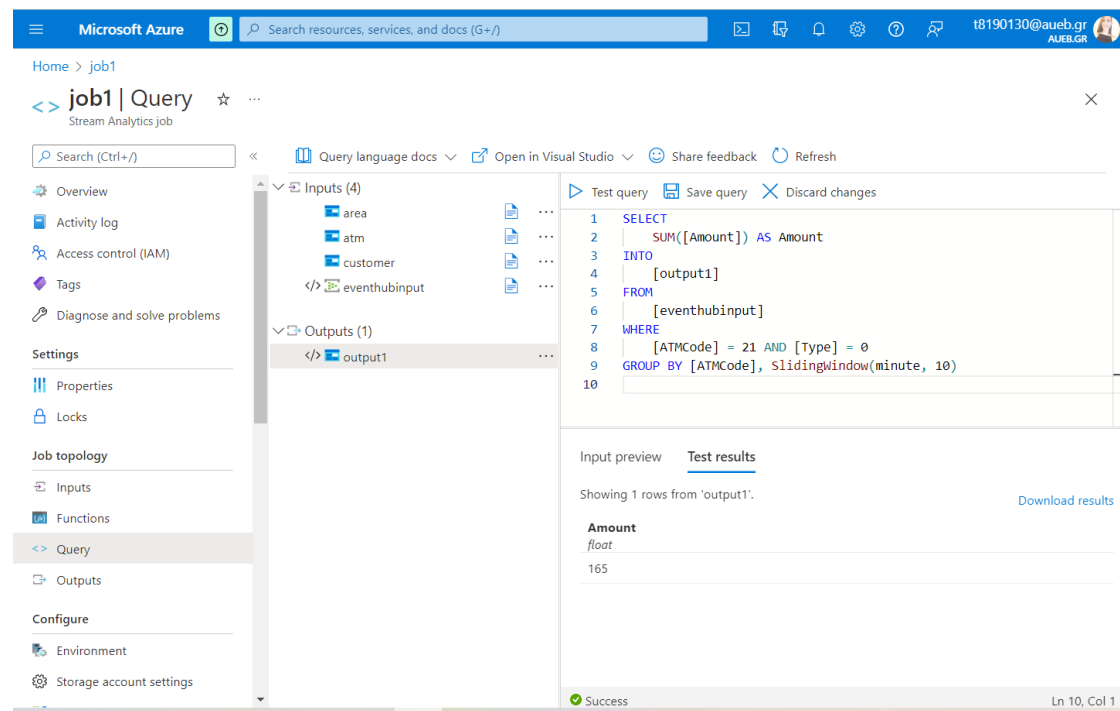


II Queries

In the interest of reproducibility the jobs were tested using the data sampled in Section 9 of this report, however jobs did also produce blobs in the output container, successfully. Both the sampled data and the results of the queries have been provided alongside this report in the folder **output_json**.

Query 1: Show the total “Amount” of “Type = 0” transactions at “ATM Code = 21” of the last 10 minutes. Repeat as new events keep flowing in (use a sliding window).

```
SELECT
    SUM([Amount]) AS Amount
INTO
    [output1]
FROM
    [eventhubinput]
WHERE
    [ATMCode] = 21 AND [Type] = 0
GROUP BY [ATMCode], SlidingWindow(minute, 10)
```



The screenshot shows the Microsoft Azure portal interface for a Stream Analytics job named 'job1'. The 'Query' tab is active, displaying the following SQL query:

```
1 SELECT
2     SUM([Amount]) AS Amount
3 INTO
4     [output1]
5 FROM
6     [eventhubinput]
7 WHERE
8     [ATMCode] = 21 AND [Type] = 0
9 GROUP BY [ATMCode], SlidingWindow(minute, 10)
10
```

The 'Test results' tab shows the output of the query, displaying a single row with the value 165 for the 'Amount' column.

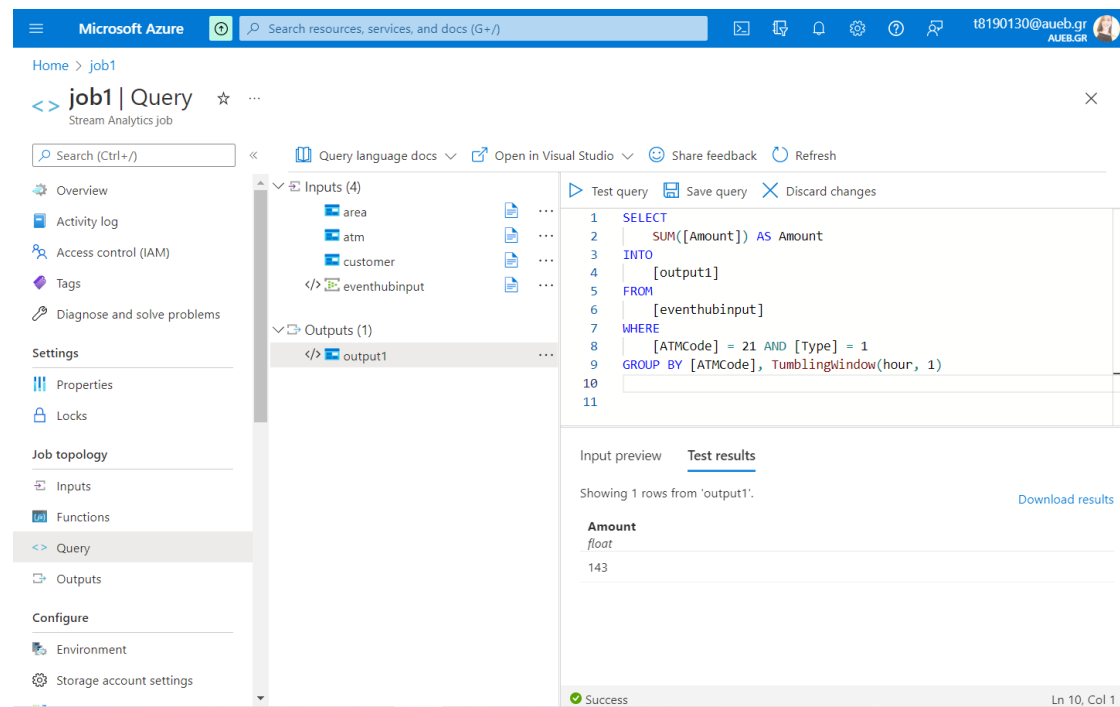
Amount
165

Reference [5]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 2: Show the total “Amount” of “Type = 1” transactions at “ATM Code = 21” of the last hour. Repeat once every hour (use a tumbling window).

```
SELECT
    SUM([Amount]) AS Amount
INTO
    [output1]
FROM
    [eventhubinput]
WHERE
    [ATMCode] = 21 AND [Type] = 1
GROUP BY [ATMCode], TumblingWindow(hour, 1)
```



The screenshot displays the Microsoft Azure portal interface for a Stream Analytics job named 'job1'. The 'Query' tab is active, showing the following SQL query:

```
1 SELECT
2     SUM([Amount]) AS Amount
3 INTO
4     [output1]
5 FROM
6     [eventhubinput]
7 WHERE
8     [ATMCode] = 21 AND [Type] = 1
9 GROUP BY [ATMCode], TumblingWindow(hour, 1)
```

The 'Test results' tab shows the output of the query, displaying a single row with the value 143 for the 'Amount' column.

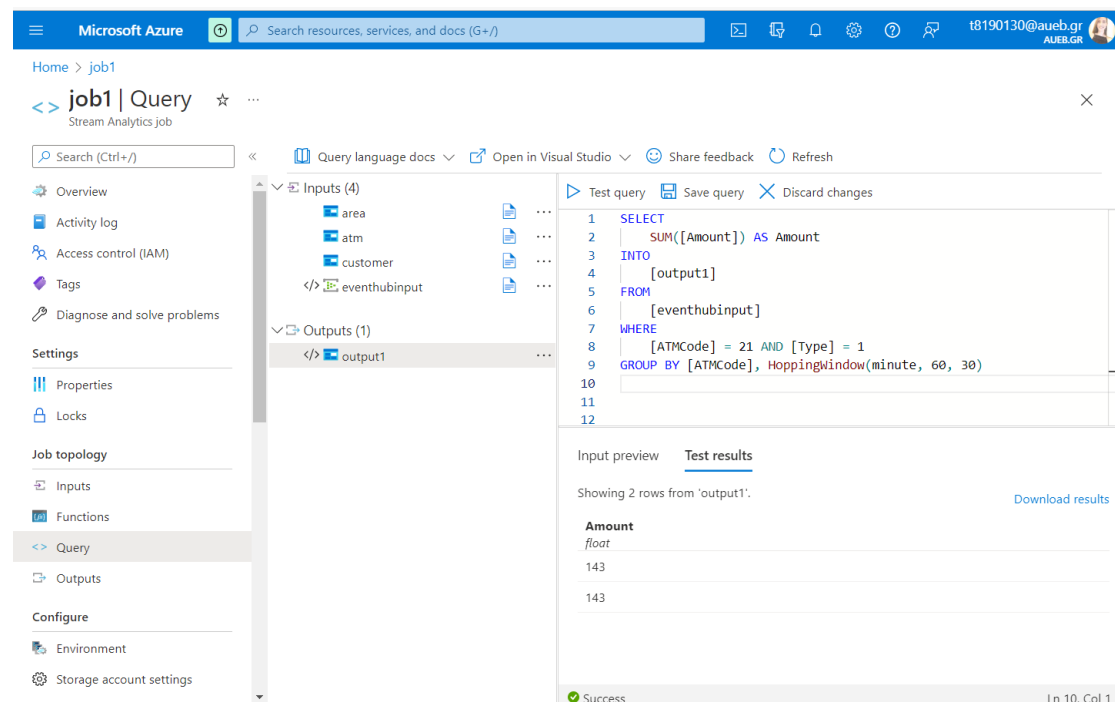
Amount
143

Reference [6]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 3: Show the total “Amount” of “Type = 1” transactions at “ATM Code = 21” of the last hour. Repeat once every 30 minutes (use a hopping window).

```
SELECT
    SUM([Amount]) AS Amount
INTO
    [output1]
FROM
    [eventhubinput]
WHERE
    [ATMCode] = 21 AND [Type] = 1
GROUP BY [ATMCode], HoppingWindow(minute, 60, 30)
```



The screenshot displays the Microsoft Azure Stream Analytics job 'job1' interface. The query editor shows the SQL query for Query 3. The test results section shows two rows of output with the column 'Amount' and values 143 and 143.

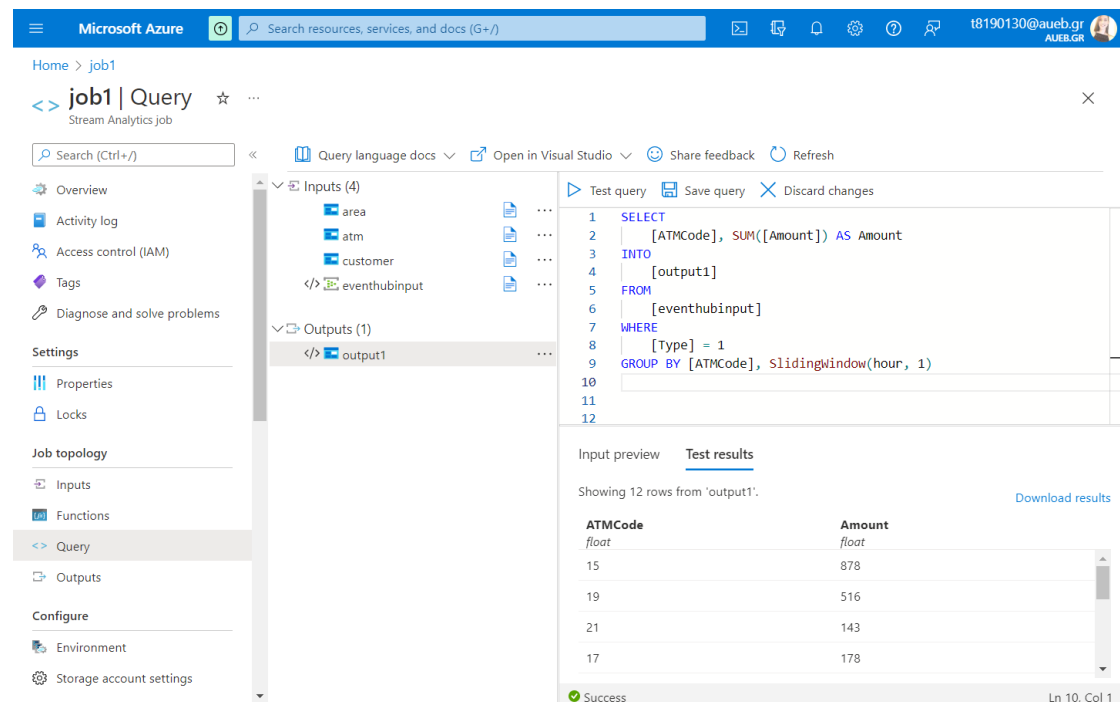
Amount
143
143

Reference [7]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 4: Query 4: Show the total “Amount” of “Type = 1” transactions per “ATM Code” of the last one hour (use a sliding window).

```
SELECT
    [ATMCode], SUM([Amount]) AS Amount
INTO
    [output1]
FROM
    [eventhubinput]
WHERE
    [Type] = 1
GROUP BY [ATMCode], SlidingWindow(hour, 1)
```



The screenshot shows the Microsoft Azure portal interface for a Stream Analytics job named 'job1'. The 'Query' tab is active, displaying the SQL query for Query 4. The 'Test results' tab is also visible, showing a table with 12 rows of data for 'output1'.

Query:

```
SELECT
    [ATMCode], SUM([Amount]) AS Amount
INTO
    [output1]
FROM
    [eventhubinput]
WHERE
    [Type] = 1
GROUP BY [ATMCode], SlidingWindow(hour, 1)
```

Test results:

ATMCode	Amount
15	878
19	516
21	143
17	178

Reference[5]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 5: Show the total “Amount” of “Type = 1” transactions per “Area Code” of the last hour. Repeat once every hour (use a tumbling window).

SELECT

[atm].[area_code], SUM([eventhubinput].[Amount]) AS Amount

INTO

[output1]

FROM

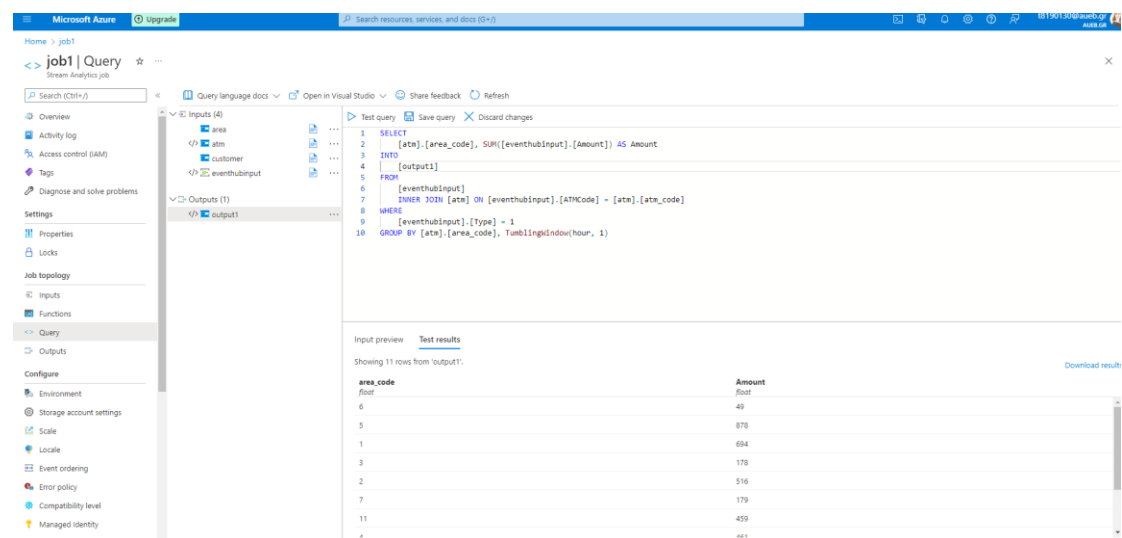
[eventhubinput]

INNER JOIN [atm] ON [eventhubinput].[ATMCode] = [atm].[atm_code]

WHERE

[eventhubinput].[Type] = 1

GROUP BY [atm].[area_code], TumblingWindow(hour, 1)



The screenshot shows the Microsoft Azure portal interface for a Stream Analytics job named 'job1'. The query editor is open, displaying the following SQL query:

```

1 SELECT
2   [atm].[area_code], SUM([eventhubinput].[Amount]) AS Amount
3 INTO
4   [output1]
5 FROM
6   [eventhubinput]
7 INNER JOIN [atm] ON [eventhubinput].[ATMCode] = [atm].[atm_code]
8 WHERE
9   [eventhubinput].[Type] = 1
10 GROUP BY [atm].[area_code], TumblingWindow(hour, 1)

```

The left sidebar shows the job configuration, including inputs (area, atm, customer, eventhubinput) and outputs (output1). The right sidebar shows the 'Test results' tab, displaying a table with 11 rows of data:

area_code	Amount
6	49
5	878
1	694
3	178
2	516
7	179
11	439
4	401

Reference[6]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 6: Show the total “Amount” per ATM’s “City” and Customer’s “Gender” of the last hour. Repeat once every hour (use a tumbling window).

SELECT

```
[area].[area_city], [customer].[gender], SUM([eventhubinput].[Amount]) AS Amount
```

INTO

```
[output1]
```

FROM

```
[eventhubinput]
```

```
INNER JOIN [atm] ON [eventhubinput].[ATMCode] = [atm].[atm_code]
```

```
INNER JOIN [area] ON [atm].[area_code] = [area].[area_code]
```

```
INNER JOIN [customer] ON [eventhubinput].[CardNumber] = [customer].
```

```
[card_number]
```

```
GROUP BY [area].[area_city], [customer].[gender], TumblingWindow(hour, 1)
```

Microsoft Azure | Upgrade | Search resources, services, and docs (G+)

Home > job1

job1 | Query | Stream Analytics job

Search (Ctrl+I) | Query language docs | Open in Visual Studio | Share feedback | Refresh

Overview | Activity log | Access control (IAM) | Tags | Diagnose and solve problems | Settings | Properties | Locks | Job topology | Inputs | Functions | Query | Outputs | Configure | Environment | Storage account settings | Scale | Locale | Event ordering | Error policy | Compatibility level | Managed Identity | General

Inputs (4): area, atm, customer, eventhubinput

Outputs (1): output1

Test query | Save query | Discard changes

```
1 SELECT
2   [area].[area_city], [customer].[gender], SUM([eventhubinput].[Amount]) AS Amount
3 INTO
4   [output1]
5 FROM
6   [eventhubinput]
7   INNER JOIN [atm] ON [eventhubinput].[ATMCode] = [atm].[atm_code]
8   INNER JOIN [area] ON [atm].[area_code] = [area].[area_code]
9   INNER JOIN [customer] ON [eventhubinput].[CardNumber] = [customer].[card_number]
10  GROUP BY [area].[area_city], [customer].[gender], TumblingWindow(hour, 1)
```

Input preview | Test results

Showing 16 rows from 'output1'.

area_city	gender	Amount
string	string	float
"Tacoma"	"Female"	347
"Schaumburg"	"Female"	1172
"Vancouver"	"Male"	290
"Baltimore"	"Female"	112
"Springfield"	"Male"	792
"Canton"	"Male"	210
"Vancouver"	"Female"	39
"Vancouver"	"Female"	...

Download results

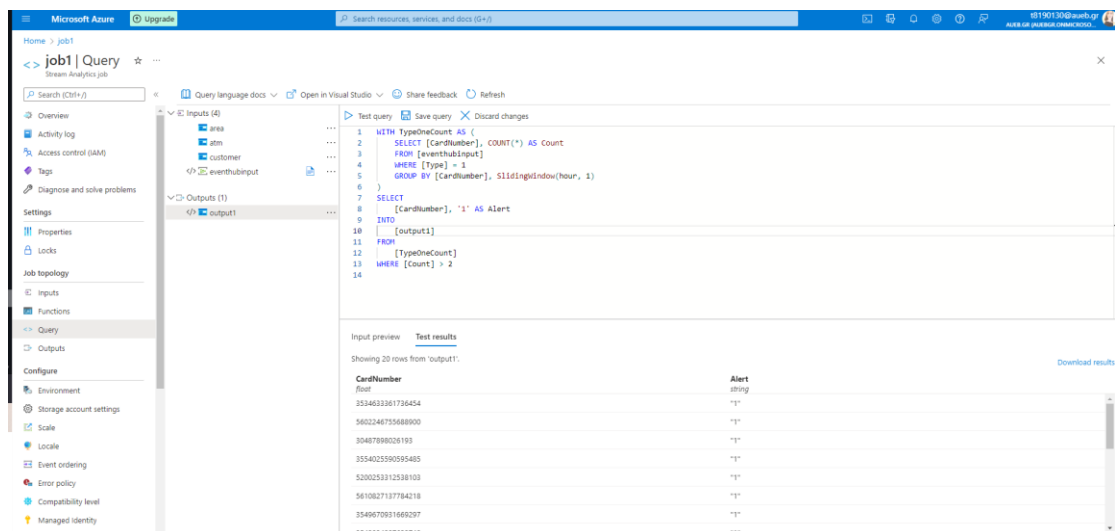
Success | Ln 2, Col 34 | 25°C

Reference[6]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 7: Alert (SELECT “1”) if a Customer has performed two transactions of “Type = 1” in a window of an hour (use a sliding window).

```
WITH TypeOneCount AS (  
    SELECT [CardNumber], COUNT(*) AS Count  
    FROM [eventhubinput]  
    WHERE [Type] = 1  
    GROUP BY [CardNumber], SlidingWindow(hour, 1)  
)  
SELECT  
    [CardNumber], '1' AS Alert  
INTO  
    [output1]  
FROM  
    [TypeOneCount]  
WHERE [Count] > 2
```



The screenshot shows the Microsoft Azure portal interface for a Stream Analytics job named 'job1'. The query editor is open, displaying the SQL query for Query 7. The 'Test results' tab is selected, showing a table with two columns: 'CardNumber' and 'Alert'. The 'Alert' column contains the value '1' for several card numbers, indicating that a customer has performed two transactions of 'Type = 1' in a window of an hour.

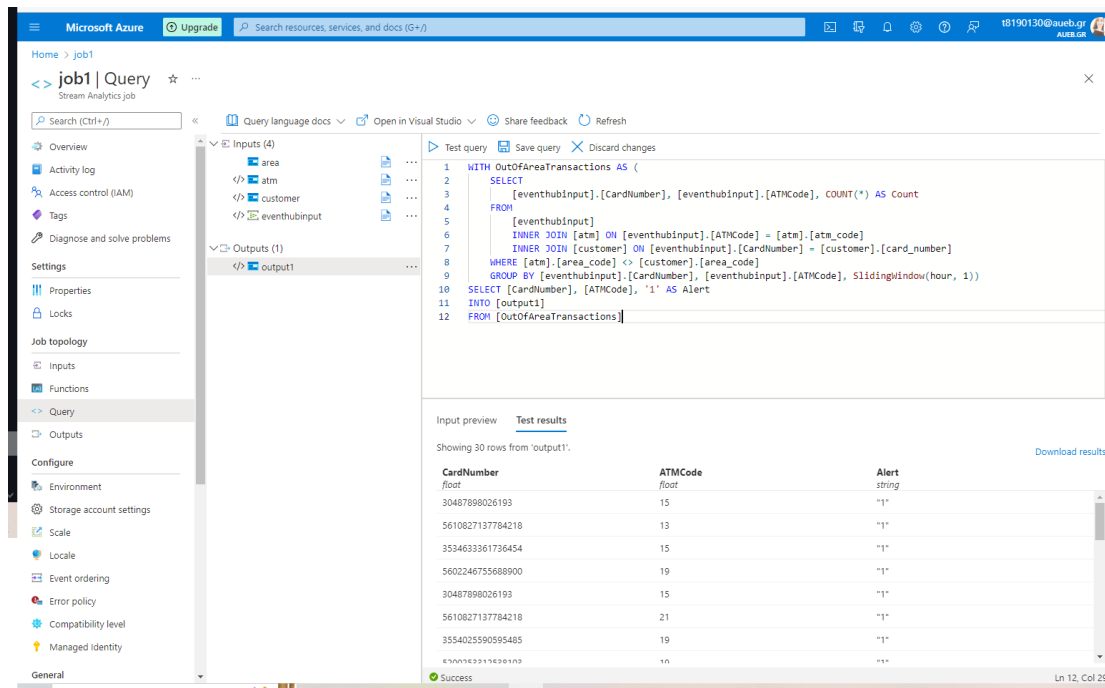
CardNumber	Alert
2534633361726454	"1"
5602246755688900	"1"
30487898026193	"1"
3554025590595485	"1"
520025311258103	"1"
5610827137784218	"1"
3549670031666207	"1"
1647074607612740	"1"

Reference[8]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

Query 8: Alert (SELECT “1”) if the “Area Code” of the ATM of the transaction is not the same as the “Area Code” of the “Card Number” (Customer’s Area Code) - (use a sliding window)

```
WITH OutOfAreaTransactions AS (
    SELECT
        [eventhubinput].[CardNumber], [eventhubinput].[ATMCode], COUNT(
*) AS Count
    FROM
        [eventhubinput]
        INNER JOIN [atm] ON [eventhubinput].[ATMCode] = [atm].[atm_code]
]
    INNER JOIN [customer] ON [eventhubinput].[CardNumber] = [customer].[card_number]
    WHERE [atm].[area_code] <> [customer].[area_code]
    GROUP BY [eventhubinput].[CardNumber], [eventhubinput].[ATMCode], SlidingWindow(hour, 1))
SELECT [CardNumber], [ATMCode], '1' AS Alert
INTO [output1]
FROM [OutOfAreaTransactions]
```



The screenshot displays the Microsoft Azure portal interface for a Stream Analytics job named 'job1'. The query editor shows the following SQL query:

```
WITH OutOfAreaTransactions AS (
    SELECT
        [eventhubinput].[CardNumber], [eventhubinput].[ATMCode], COUNT(*) AS Count
    FROM
        [eventhubinput]
        INNER JOIN [atm] ON [eventhubinput].[ATMCode] = [atm].[atm_code]
]
    INNER JOIN [customer] ON [eventhubinput].[CardNumber] = [customer].[card_number]
    WHERE [atm].[area_code] <> [customer].[area_code]
    GROUP BY [eventhubinput].[CardNumber], [eventhubinput].[ATMCode], SlidingWindow(hour, 1))
SELECT [CardNumber], [ATMCode], '1' AS Alert
INTO [output1]
FROM [OutOfAreaTransactions]
```

The 'Test results' tab shows the output of the query, displaying 30 rows from 'output1'. The output table has three columns: CardNumber (float), ATMCode (float), and Alert string (string). The data shows transactions where the ATM area code does not match the customer's area code, resulting in an Alert string of '1'.

CardNumber	ATMCode	Alert string
30487898026193	15	"1"
5610827137784218	13	"1"
3534633361736454	15	"1"
5602246755688900	19	"1"
30487898026193	15	"1"
5610827137784218	21	"1"
3554025590595485	19	"1"
610001623116201193	10	"1"

Reference[8]

Note : in folder output_json exists the exported output from azure interfaces for all queries.

11. Conclusion

In conclusion, as this assignment is the last one, we would like to mention that this course helped as to learn useful tools – systems for big data. Specifically, we worked with Hadoop, Redis, neo4j and azure in practice. Last but not least, we would like to thank our professor Damianos Chatziantoniou for the course. Good summer!

12. References

- [Azure Streams Analytics][1] : [Azure Stream Analytics documentation | Microsoft Docs](#)
- [Event Hub][2] : <https://azure.microsoft.com/en-us/services/event-hubs/#overview>
- [Storage Account][3] : <https://docs.microsoft.com/en-us/azure/storage/common/storage-account-overview>
- [Create Stream Analytics Job][4] : <https://docs.microsoft.com/en-us/azure/stream-analytics/stream-analytics-quick-create-portal>
- [Sliding Window][5] : [Sliding Window \(Azure Stream Analytics\) - Stream Analytics Query | Microsoft Docs](#)
- [Tumbling Window][6] : [Tumbling Window \(Azure Stream Analytics\) - Stream Analytics Query | Microsoft Docs](#)
- [Hopping Window][7] : [Hopping Window \(Azure Stream Analytics\) - Stream Analytics Query | Microsoft Docs](#)
- [Sliding Window][8] : [Sliding Window \(Azure Stream Analytics\) - Stream Analytics Query | Microsoft Docs](#)