

**ΟΙΚΟΝΟΜΙΚΟ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΑΘΗΝΩΝ**



**ATHENS UNIVERSITY  
OF ECONOMICS  
AND BUSINESS**

**MSc in Business Analytics – DMBI  
Assignment #3**

Azure Stream Analytics

Vogiatzis George: p2821827

Zourou Mirsini: p2821828

Fall Quarter Pt. 2018



## Contents

Steps of the procedure .....	3
Namespace creation .....	3
Setting policies .....	4
Setting up signature .....	5
Editing html configuration .....	5
Feeding event hub with Generator.html .....	5
Blob Storage Account Creation .....	6
Storage account container creation .....	7
Reference data files .....	7
Stream demo creation .....	8
Stream Demo Inputs Outputs and Settings .....	8
Input sample data for stream demo .....	9
Upload input file for stream demo .....	10
Test Query .....	10
Reference Data Settings .....	11
Stream Output .....	11
Generated Blob .....	12
Queries .....	13
Query 1 .....	13
Query 2 .....	14
Query 3 .....	16
Query 4 .....	17
Query 5 .....	19
Query 6 .....	20
Query 7 .....	21
Query 8 .....	22

## Steps of the procedure

### Namespace creation

The screenshot shows the 'Create Namespace' page in the Azure portal. The breadcrumb navigation is 'Home > New > Event Hubs > Create Namespace'. The page title is 'Create Namespace' with a sub-header 'Event Hubs'. The form includes the following fields and options:

- Name:** 'asgnmnt3hub' (with a checkmark icon and '.servicebus.windows.net' suffix).
- Pricing tier:** 'Standard (20 Consumer groups, 1000 Brokered connections)' (with a dropdown arrow).
- Enable Kafka:** ☐ (with a help icon).
- Make this namespace zone redundant:** ☐ (with a help icon).
- Subscription:** 'Azure for Students' (with a dropdown arrow).
- Resource group:** '(New) msba\_rg' (with a dropdown arrow and a 'Create new' link).
- Location:** 'West Europe' (with a dropdown arrow).
- Throughput Units:** A slider set to '1'.
- Enable Auto-Inflate:** ☐ (with a help icon).

A blue 'Create' button is located at the bottom left of the form.

### Event hub creation

The screenshot shows the 'Create Event Hub' page in the Azure portal. The breadcrumb navigation is 'Home > asgnmnt3hub - Event Hubs > Create Event Hub'. The page title is 'Create Event Hub' with a sub-header 'Event Hubs'. The form includes the following fields and options:

- Name:** 'myEventHub' (with a checkmark icon).
- Partition Count:** A slider set to '2'.
- Message Retention:** A slider set to '1'.
- Capture:** A toggle switch set to 'On'.

A blue 'Create' button is located at the bottom left of the form.

## Setting policies

Microsoft Azure

Dashboard > myeventhub - Shared access policies

myeventhub - Shared access policies

Search (Ctrl+/)

Overview

Access control (IAM)

Diagnose and solve problems

Settings

Shared access policies

Properties

Locks

Automation script

Entities

Consumer groups

Features

Capture

Support + troubleshooting

New support request

Search to filter items...

POLICY	CLAIMS
mySendPolicy	Send
myRecPolicy	Listen

SAS Policy: mySendPolicy

Save Discard More

Manage

☒ Send

☐ Listen

Primary key

ng1Q6aY+6Qz9CUsikwN8pY469Pp...

Secondary key

nRPa1qkHfH2M64P8jKGSqd054yAs1mb...

Connection string-primary key

Endpoints=sb://asgmt3hub.servicebu...

Connection string-secondary key

Endpoints=sb://asgmt3hub.servicebu...

Microsoft Azure

Dashboard > myeventhub - Shared access policies

myeventhub - Shared access policies

Search (Ctrl+/)

Overview

Access control (IAM)

Diagnose and solve problems

Settings

Shared access policies

Properties

Locks

Automation script

Entities

Consumer groups

Features

Capture

Support + troubleshooting

New support request

Search to filter items...

POLICY	CLAIMS
mySendPolicy	Send
myRecPolicy	Listen

SAS Policy: myRecPolicy

Save Discard More

Manage

☐ Send

☒ Listen

Primary key

GBKW2hYjVE2w2Cvahg6Q7VCR0fh9euf...

Secondary key

720+1zPOHQoDPmN2f5l6hteIX8BS/L...

Connection string-primary key

Endpoints=sb://asgmt3hub.servicebu...

Connection string-secondary key

Endpoints=sb://asgmt3hub.servicebu...

## Setting up signature

Event Hubs - Signature Generator

**Hub**

Namespace: asgmnmt3hub

Hub Name: myeventhub

Publisher: Laptop

Mode: Http

**Credentials**

Sender Key Name: mySendPolicy

Sender Key: r+b0Qr9C/UsiklwNBpY4k9PYpoZ9sIS+BsRcg=

Token TTL (minutes): 7200

**Signature**

SharedAccessSignature sr=https%3a%2f%2fasgmnmt3hub.servicebus.windows.net%2fmyeventhub%2fpublishers%2flaptop%2fmessages&sig=N2sF%2biU3M61Ok1mtVd%2fHcacZSIVwUr%2fZPwXNU9xn18%3d&se=1548019643&skn=mySendPolicy

Generate

## Editing html configuration

```
/*  
*** CONFIG ***  
*/  
...  
//Use the signature generator: https://github.com/sandrinodimattia/RedDog/releases  
var sas = "SharedAccessSignature sr=https%3a%2f%2fmscba-aueb.servicebus.windows.net%2feventhubdemo%2fpublishers%2flaptop%  
var serviceNamespace = "mscba-aueb";  
var hubName = "eventhubdemo";  
var deviceName = "Laptop";
```

## Feeding event hub with Generator.html

File | file:///C:/Users/George/Desktop/ergAzzureStreams/Assignment3/StreamDataGenerator/DataGenerator/Generator.html

Send Data Sent: { "vehicleTypeID": 931 , "licensePlate": "RJC-4310" , "speed": "0" , "colorID": 9 , "checkpointID": 1 , "spotType": "Toll\_Station" }

## Blob Storage Account Creation

### Create storage account

[Basics](#) [Advanced](#) [Tags](#) [Review + create](#)

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more](#)

#### PROJECT DETAILS

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

\* Subscription

\* Resource group  [Create new](#)

#### INSTANCE DETAILS

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

\* Storage account name ⓘ  ✓

\* Location

Performance ⓘ ☒ Standard ☐ Premium

Account kind ⓘ

Replication ⓘ

ⓘ Accounts with the selected kind, replication and performance type only support block and append blobs. Page blobs, file shares, tables, and queues will not be available.

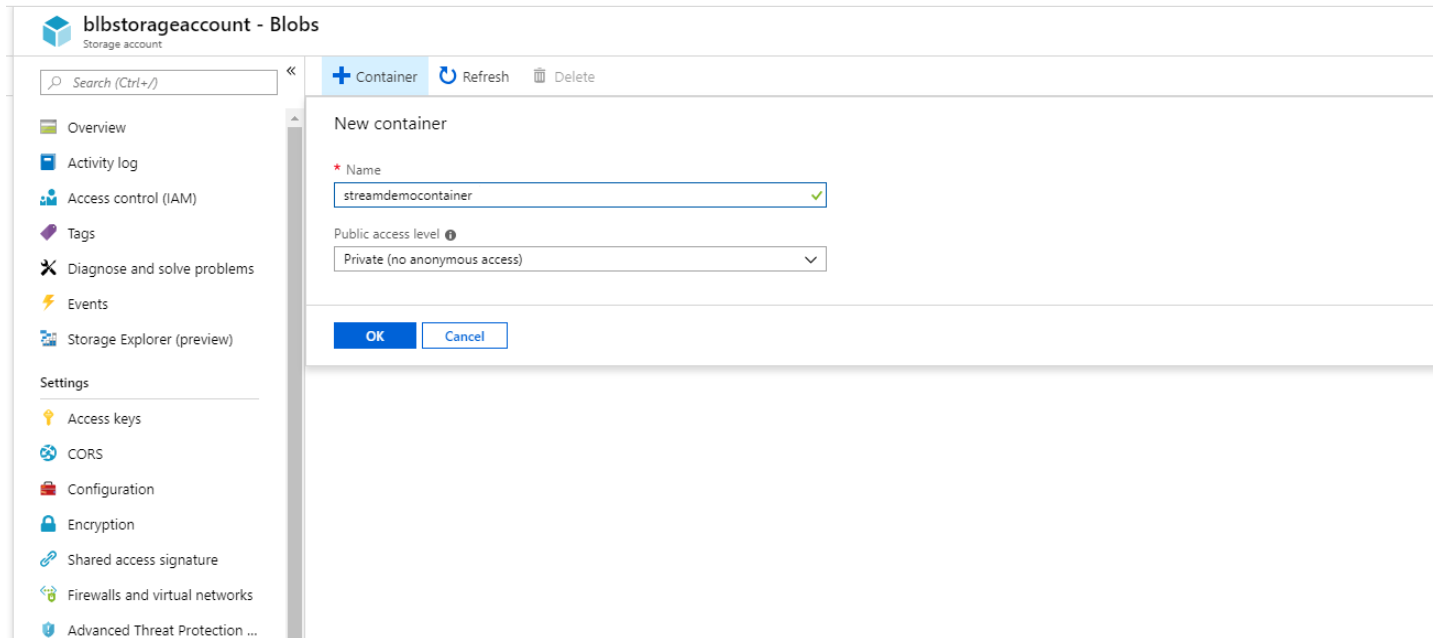
Access tier (default) ⓘ ☐ Cool ☒ Hot

[Review + create](#)

[Previous](#)

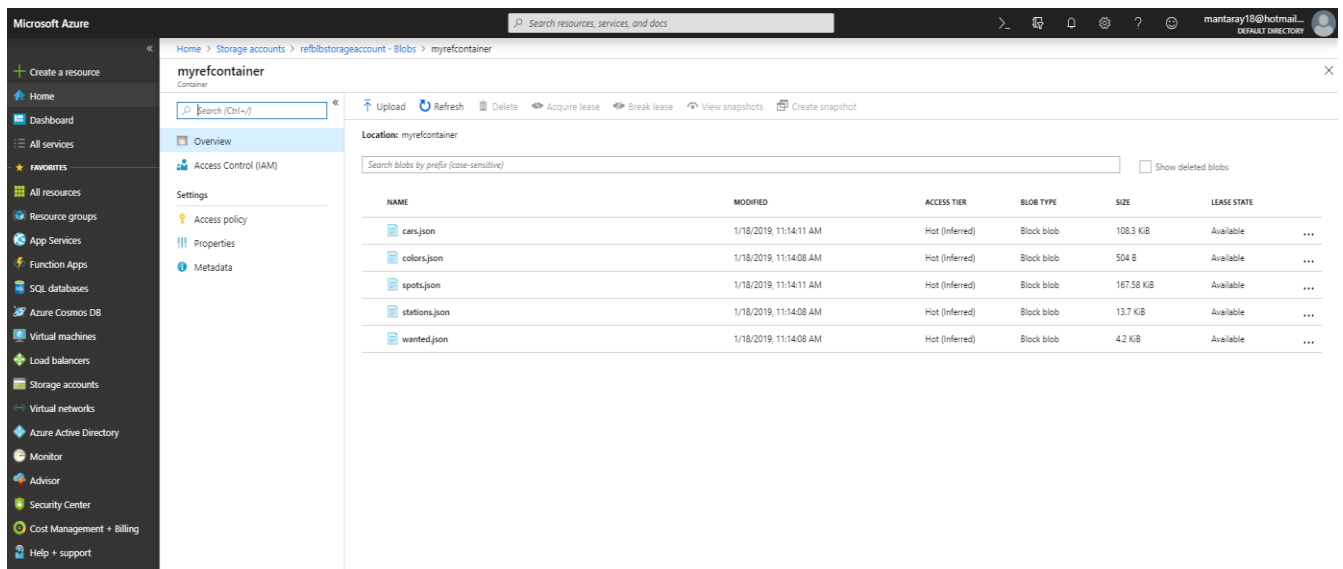
[Next : Advanced >](#)

## Storage account container creation



## Reference data files

To make better use of our reference data files we first had to convert the txt files to csv via Excel and then to json with the help of an online converter. Particularly for the txt with the wanted license plates we had to add a “column” name on the beginning of the file to convert to a valid json file. After the transformations we uploaded our json reference file to another blob storage account



Create a resource

Home

Dashboard

All services

FAVORITES

All resources

Resource groups

App Services

Function Apps

SQL databases

Azure Cosmos DB

Virtual machines

Load balancers

Storage accounts

Virtual networks

Azure Active Directory

Monitor

Advisor

Security Center

Cost Management + Billing

Help + support

Search resources, services, and docs

Home > New > New Stream Analytics job

New Stream Analytics job

\* Job name  
streamdemo ✓

\* Subscription  
Azure for Students ▼

\* Resource group  
msba\_rg ▼  
[Create new](#)

\* Location  
West Europe ▼

Hosting environment ⓘ  
Cloud Edge

Streaming units (1 to 120) ⓘ  
 3

Create Automation options

Dashboard > streamdemo > Inputs

Inputs

+ Add stream input

+ Add reference input

NAME	SOURCE TYPE	SOURCE
cars	Reference	Blob storage ...
colors	Reference	Blob storage ...
input	Stream	Event Hub <a href="#">Sample data</a> ...
spots	Reference	Blob storage ...
stations	Reference	Blob storage ...
wanted	Reference	Blob storage ...



Input details

input

Test

Delete

\* Input alias

input

Provide Event Hub settings manually

Select Event Hub from your subscriptions

Subscription

Azure for Students

\* Event Hub namespace

asgnmnt3hub

\* Event Hub name

Create new

Use existing

myeventhub

\* Event Hub policy name

myRecPolicy

Event Hub policy key

\*\*\*\*\*

Event Hub consumer group

\* Event serialization format

JSON

Encoding

UTF-8

Event compression type

None

Output details

output

Test

Delete

\* Output alias

output

Provide Blob storage settings manually

Select Blob storage from your subscriptions

Subscription

Azure for Students

\* Storage account

blbstorageaccount

Storage account key

\*\*\*\*\*

\* Container

Create new

Use existing

streamdemocontainer

Path pattern

Date format

YYYY/MM/DD

Time format

HH

\* Event serialization format

JSON

Encoding

UTF-8

Format

Line separated

## Input sample data for stream demo

```

1 {
2   "vehicleTypeID": 396,
3   "licensePlate": "CPH-0523",
4   "speed": "76",
5   "colorID": 4,
6   "checkpointID": 21,
7   "spotType": "Speed Limit Camera",
8   "EventProcessedUtcTime": "2019-01-15T21:49:30.2294651Z",
9   "PartitionId": 1,
10  "EventEnqueuedUtcTime": "2019-01-15T21:49:25.2820000Z"
11 },
12 {
13   "vehicleTypeID": 950,
14   "licensePlate": "JGL-3532",
15   "speed": "0",
16   "colorID": 9,
17   "checkpointID": 19,
18   "spotType": "Toll Station",
19   "EventProcessedUtcTime": "2019-01-15T21:49:30.2294651Z",
20   "PartitionId": 1,
21   "EventEnqueuedUtcTime": "2019-01-15T21:49:26.2820000Z"
22 },
23 {
24   "vehicleTypeID": 270,
25   "licensePlate": "FVG-2343",
26   "speed": "109",
27   "colorID": 5,
28   "checkpointID": 372,
29   "spotType": "Speed Limit Camera",
30   "EventProcessedUtcTime": "2019-01-15T21:49:30.2294651Z",
31   "PartitionId": 1,
32   "EventEnqueuedUtcTime": "2019-01-15T21:49:27.2980000Z"
33 },
34 {
35   "vehicleTypeID": 671,
36   "licensePlate": "VPC-6951",
37   "speed": "31",
38   "colorID": 1,
39   "checkpointID": 344,
40   "spotType": "Speed Limit Camera",
41   "EventProcessedUtcTime": "2019-01-15T21:49:30.2294651Z",
42   "PartitionId": 1,
43   "EventEnqueuedUtcTime": "2019-01-15T21:49:28.2830000Z"
44 }

```

## Upload input file for stream demo

The screenshot shows the Stream Analytics Query Editor interface. On the left is a navigation pane with options like Overview, Activity log, Access control (IAM), Tags, Diagnose and solve problems, Settings, Locks, Job topology, Inputs, Functions, Query (selected), Outputs, Configure, Scale, Locale, Event ordering, Error policy, Compatibility level, General, Tools, Properties, Monitoring, Metrics, and Alert rules. The main area displays a query: `1 select * from Input`. A 'Test' button is visible in the top right. A 'streamdemo - Query' dialog box is open, showing the file 'streamdemo-stream\_input.json' selected. The dialog has radio buttons for JSON (selected), CSV, and Avro. An 'OK' button is at the bottom right.

## Test Query

The screenshot shows the Stream Analytics Query Editor interface with the 'Test' button highlighted. The query `1 select * from Input` is shown. Below the query, a message states: 'Your query could be put in logs that are in a potentially different geography. Missing some language constructs? Let us know! (Powered by UserVoice - Privacy Policy)'. The 'Results' section shows the output of the query, which is a table with 95 rows. The table has columns: VEHICLEID, LICENSEPLATE, SPEED, COLORID, CHECKPOINTID, SPOTTYPE, EVENTPROCESSEDU..., PARTITIONID, and EVENTENQUEUEDU... The first few rows of data are as follows:

VEHICLEID	LICENSEPLATE	SPEED	COLORID	CHECKPOINTID	SPOTTYPE	EVENTPROCESSEDU...	PARTITIONID	EVENTENQUEUEDU...
396	"CPM-0523"	"76"	4	21	"Speed_Limit_Ca...	"2019-01-15T21:4...	1	"2019-01-15T21:4...
950	"JGL-3532"	"0"	9	19	"Toll_Station"	"2019-01-15T21:4...	1	"2019-01-15T21:4...
270	"PVG-2343"	"109"	5	372	"Speed_Limit_Ca...	"2019-01-15T21:4...	1	"2019-01-15T21:4...
671	"VPC-6951"	"31"	1	344	"Speed_Limit_Ca...	"2019-01-15T21:4...	1	"2019-01-15T21:4...
169	"CKS-7871"	"0"	1	94	"Toll_Station"	"2019-01-15T21:4...	1	"2019-01-15T21:4...
877	"DAE-3438"	"0"	5	15	"Toll_Station"	"2019-01-15T21:4...	1	"2019-01-15T21:4...
851	"CKV-8149"	"0"	3	5	"Toll_Station"	"2019-01-15T21:4...	1	"2019-01-15T21:4...
31	"FLY-2927"	"43"	3	184	"Speed_Limit_Ca...	"2019-01-15T21:4...	1	"2019-01-15T21:4...

## Reference Data Settings

The screenshot shows the 'streamdemo' job configuration in the Microsoft Azure portal. The 'Inputs' section lists several data sources:

NAME	SOURCE TYPE	SOURCE
cars	Reference	Blob storage
colors	Reference	Blob storage
input	Stream	Event Hub
spots	Reference	Blob storage
stations	Reference	Blob storage
wanted	Reference	Blob storage

The 'Input details' panel on the right shows the configuration for the 'cars' input:

- Input alias:** cars
- Subscription:** Azure for Students
- Storage account:** blobstorageaccount
- Storage account key:** [Redacted]
- Container:** Use existing (selected), streamdemocontainer
- Path pattern:** cars.json
- Date format:** YYYY/MM/DD
- Time format:** HH
- Event serialization format:** JSON
- Encoding:** UTF-8

The 'Save' button is visible at the bottom of the 'Input details' panel.

## Stream Output

The screenshot shows the 'streamdemo' job configuration in the Microsoft Azure portal. The 'Overview' tab is selected, showing the job is 'Running'.

**Inputs:** 6 (cars, colors)

**Outputs:** 1 (output)

**Query:**

```
1 select cars.car_make
2 into output
3 from input
4 inner join cars
5 on input.vehicleid = cars.id
```

**Monitoring:**

- Input Events (Sum):** 187
- Output Events (Sum):** 184
- Runtime Errors (Sum):** 0

**Resource utilization:**

- 50% Utilization (Max):** 8%

## Generated Blob

The screenshot displays the Microsoft Azure portal interface. On the left is the navigation pane with options like 'Create a resource', 'Dashboard', 'All services', 'FAVORITES', 'Resource groups', 'App Services', 'Function Apps', 'SQL databases', 'Azure Cosmos DB', 'Virtual machines', 'Load balancers', 'Storage accounts', 'Virtual networks', 'Azure Active Directory', 'Monitor', 'Advisor', 'Security Center', 'Cost Management + Billing', and 'Help + support'.

The main area shows the 'streamdemocontainer' storage container. The 'Overview' tab is selected, displaying a list of blobs. The selected blob, '0\_65c507f1ee004e9db87d8acd92eec9a3\_1.json', is highlighted. The right pane shows the blob's content, which is a JSON array of car makes.

**Blob List:**

NAME
0_13da9e83fd044e8b3b6...
0_40bcf21eb9a041eb823d...
0_4267ecd060914a20bd2...
0_49e2a240cf394985a123e...
0_55ee84090264f99329...
0_56c2f90a1374eba998e...
0_5h18e63880d948569031...
0_5d8765e4b84e48b83f3...
0_659795f3db914304a59b...
0_65c507f1ee004e9db87d8...
0_6624e595d0e6471482f...
0_669c0b993f6745d0a34b...
0_7bb16e8e18f2429689e5...
0_8601b96a049475a815e1...
0_92300d4b7eb84721a6d9...
0_ab2cdd6efaa4e44b342...
0_b6e7c598d147489db844...
0_b7b68851458d4642b437...

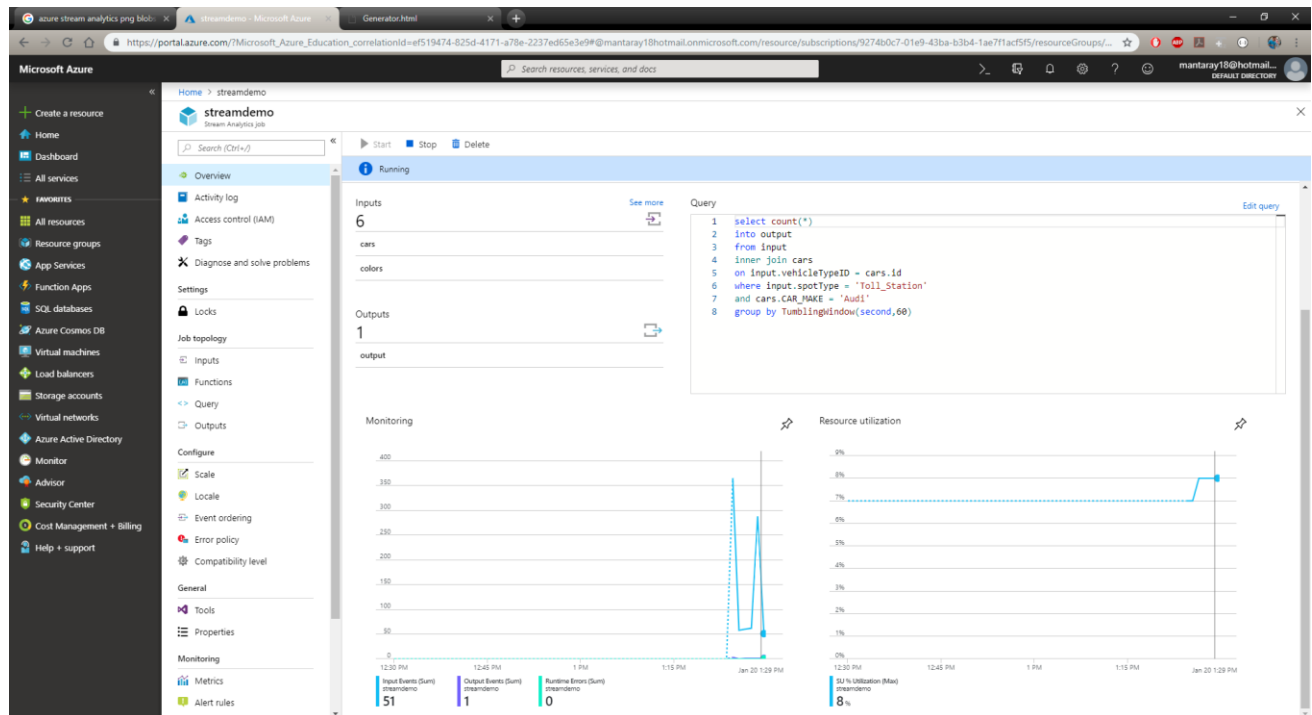
**Selected Blob Content (JSON):**

```
1 [{"car_make": "Pontiac"}]
2 [{"car_make": "Volvo"}]
3 [{"car_make": "Mazda"}]
4 [{"car_make": "Volkswagen"}]
5 [{"car_make": "Eagle"}]
6 [{"car_make": "Austiza"}]
7 [{"car_make": "Infiniti"}]
8 [{"car_make": "Ford"}]
9 [{"car_make": "Ford"}]
10 [{"car_make": "Audi"}]
11 [{"car_make": "Ford"}]
12 [{"car_make": "Toyota"}]
13 [{"car_make": "Mercury"}]
14 [{"car_make": "Mitsubishi"}]
15 [{"car_make": "Chevrolet"}]
16 [{"car_make": "Mazda"}]
17 [{"car_make": "Isuzu"}]
18 [{"car_make": "Kia"}]
19 [{"car_make": "GMC"}]
20 [{"car_make": "BMW"}]
21 [{"car_make": "Pontiac"}]
22 [{"car_make": "Lexus"}]
23 [{"car_make": "Mercedes-Benz"}]
24 [{"car_make": "Hyundai"}]
25 [{"car_make": "GMC"}]
26 [{"car_make": "Mazda"}]
27 [{"car_make": "Buick"}]
28 [{"car_make": "Hyundai"}]
29 [{"car_make": "Pontiac"}]
30 [{"car_make": "Saturn"}]
31 [{"car_make": "Audi"}]
32 [{"car_make": "Mercury"}]
33 [{"car_make": "Jeep"}]
```

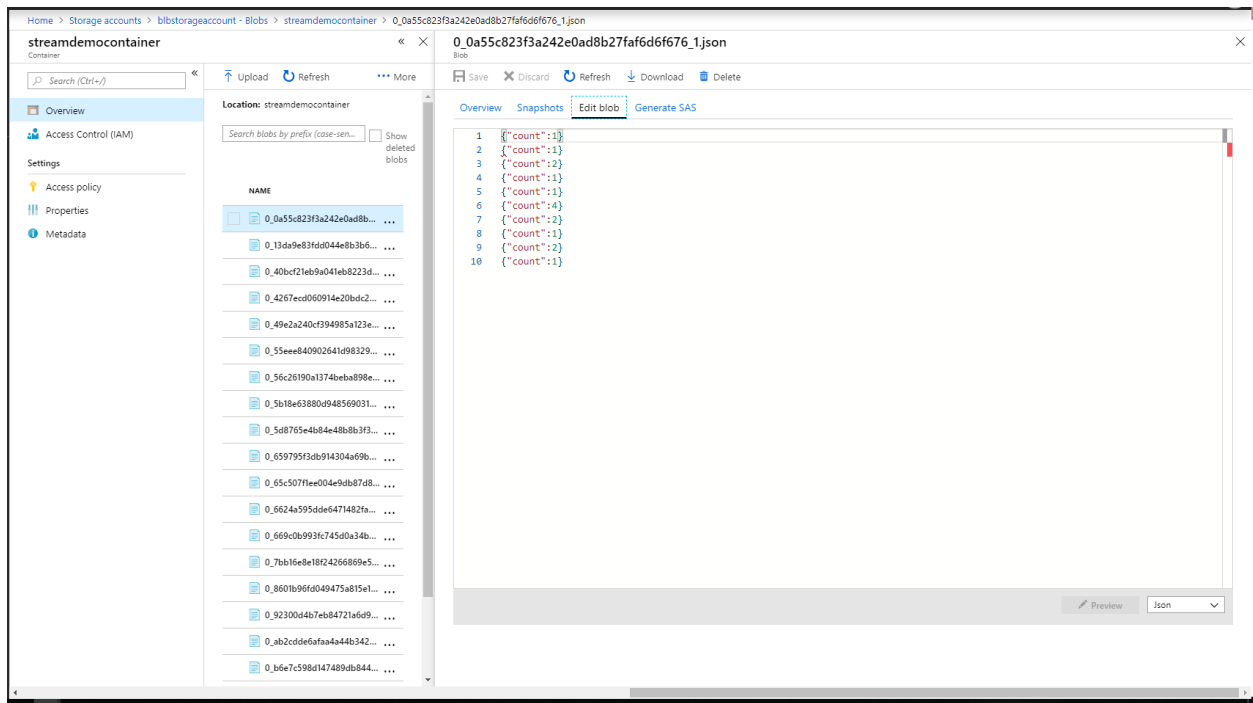
# Queries

## Query 1

```
select count(*)  
into output  
from input  
inner join cars  
on input.vehicleTypeID = cars.id  
where input.spotType = 'Toll_Station'  
and cars.CAR_MAKE = 'Audi'  
group by TumblingWindow(second,60)
```

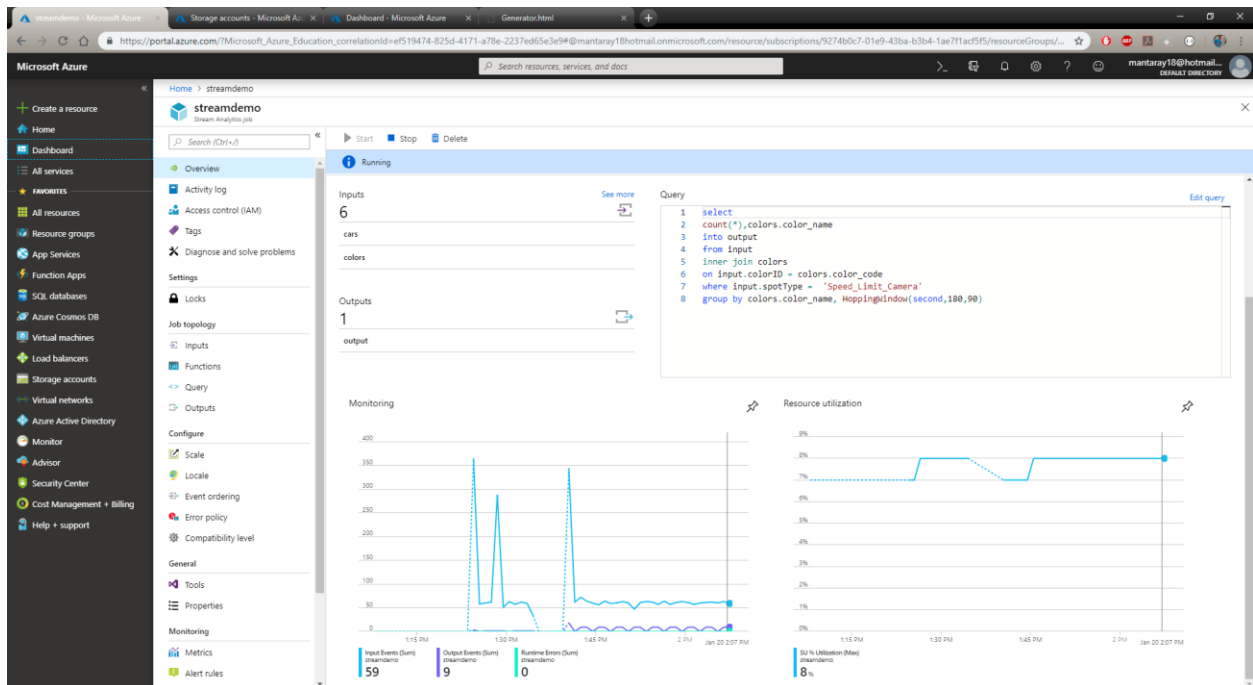


## Blob output of query 1

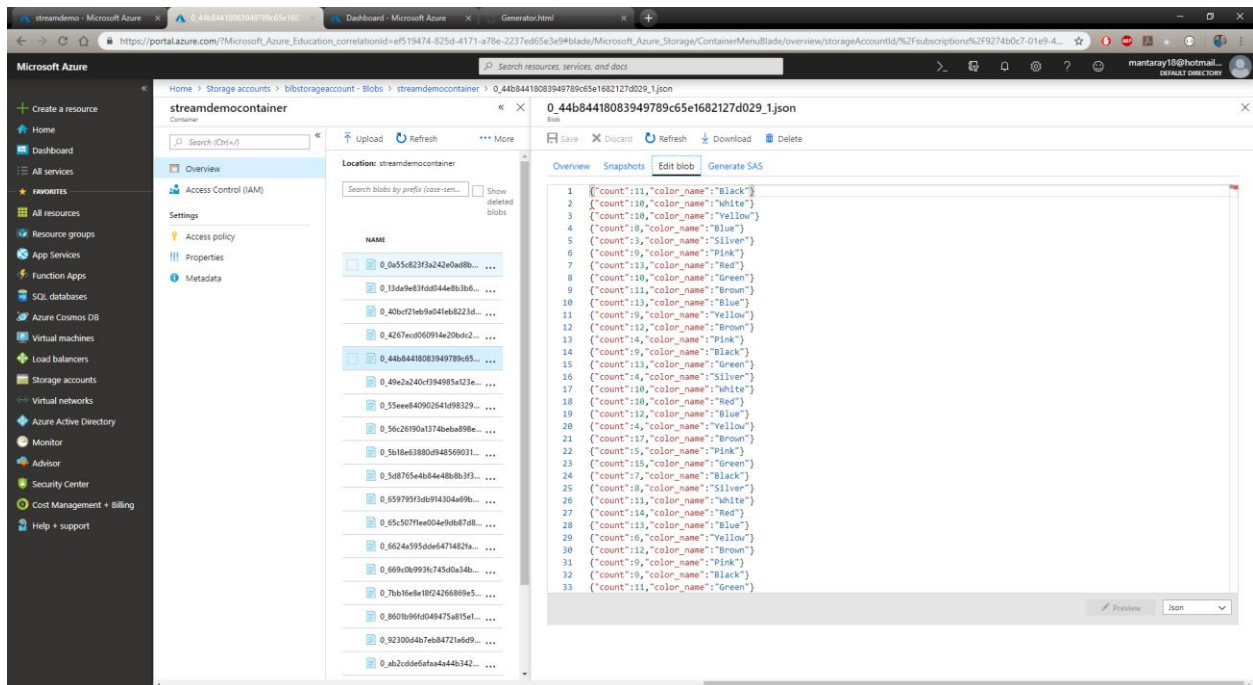


## Query 2

```
select
count(*),colors.color_name
into output
from input
inner join colors
on input.colorID = colors.color_code
where input.spotType = 'Speed_Limit_Camera'
group by colors.color_name, HoppingWindow(second,180,90)
```

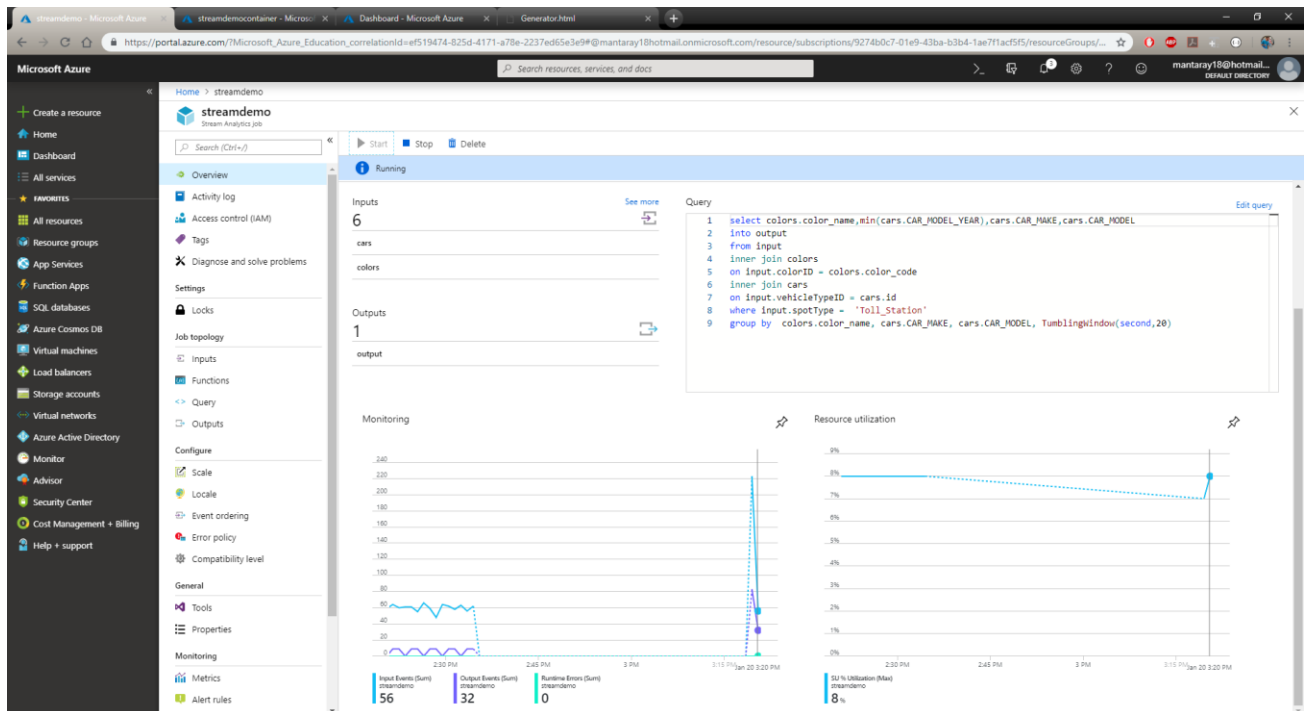


## Blob output query 2



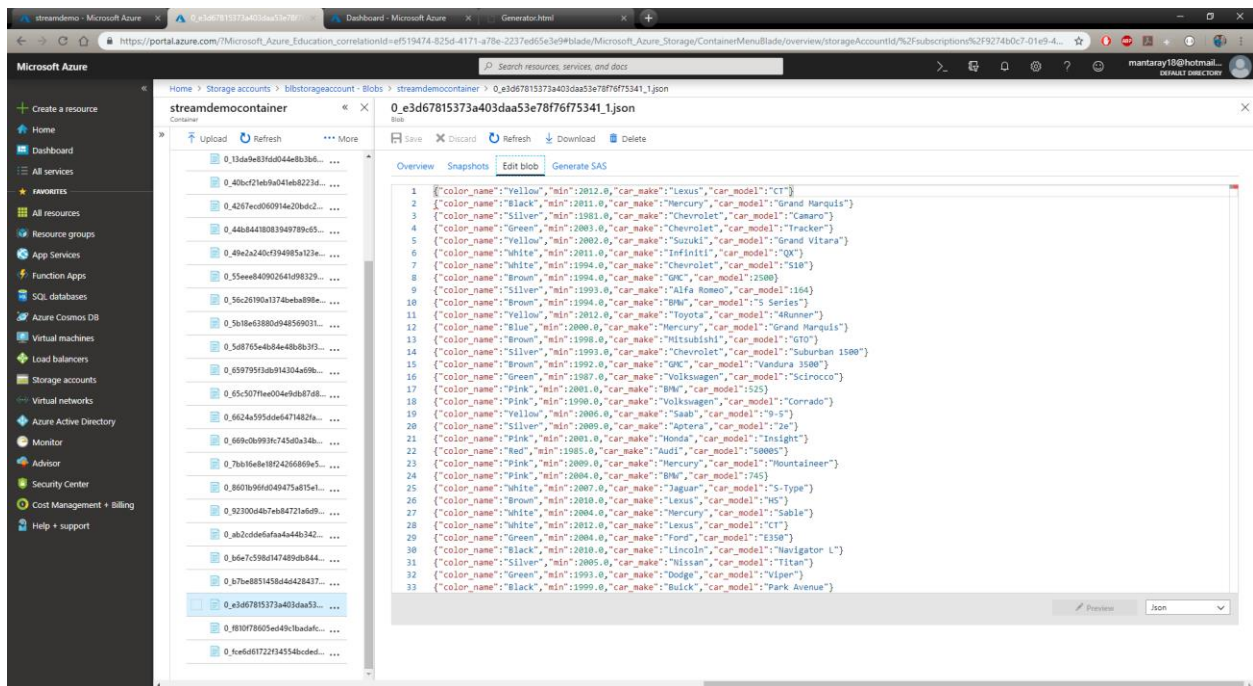
### Query 3

```
select colors.color_name,min(cars.CAR_MODEL_YEAR),cars.CAR_MAKE,cars.CAR_MODEL  
into output  
from input  
inner join colors  
on input.colorID = colors.color_code  
inner join cars  
on input.vehicleTypeID = cars.id  
where input.spotType = 'Toll_Station'  
group by colors.color_name, cars.CAR_MAKE, cars.CAR_MODEL, TumblingWindow(second,20)
```

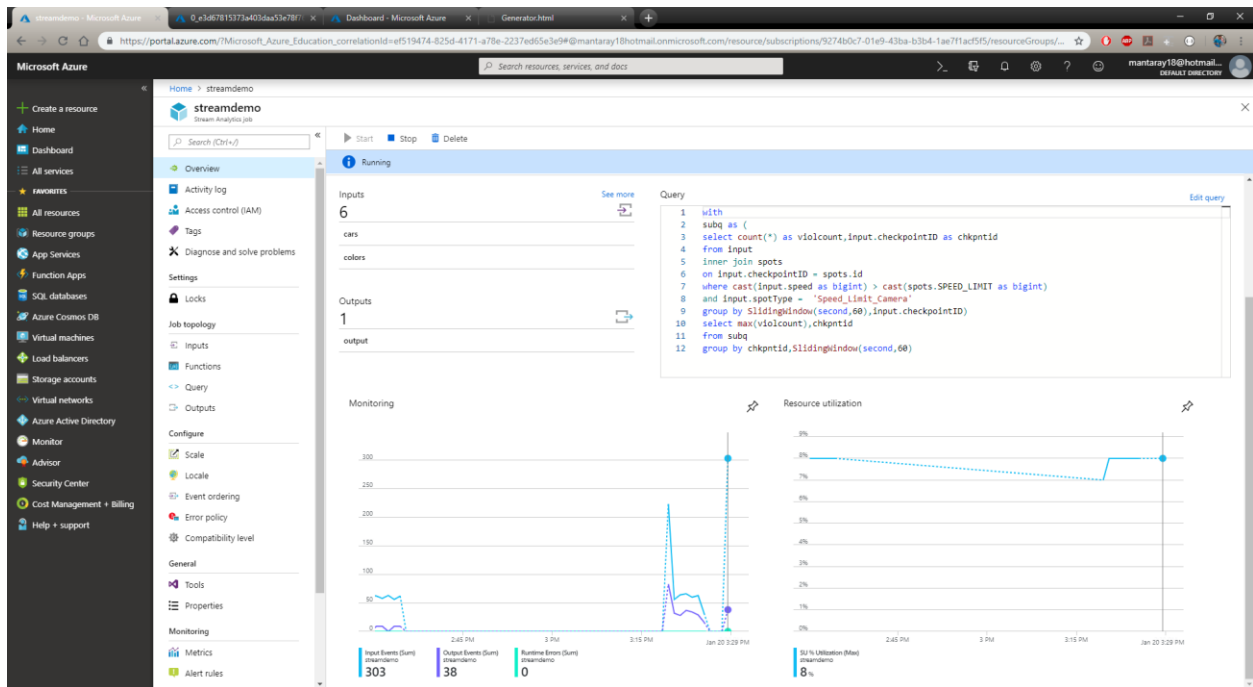




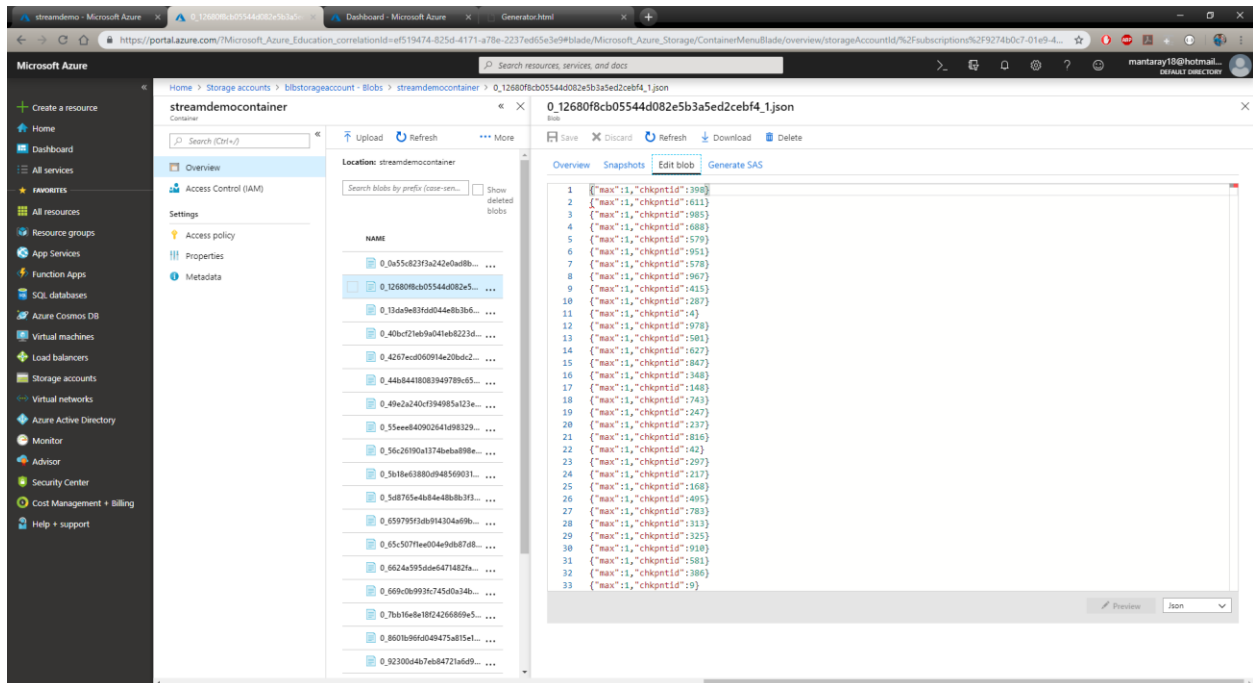
## Blob output query 3



Query 4  
with  
subq as (  
select count(\*) as violcount, input.checkpointID as chkptid  
from input  
inner join spots  
on input.checkpointID = spots.id  
where cast(input.speed as bigint) > cast(spots.SPEED\_LIMIT as bigint)  
and input.spotType = 'Speed\_Limit\_Camera'  
group by SlidingWindow(second,60), input.checkpointID)  
select max(violcount), chkptid  
from subq  
group by chkptid, SlidingWindow(second,60)

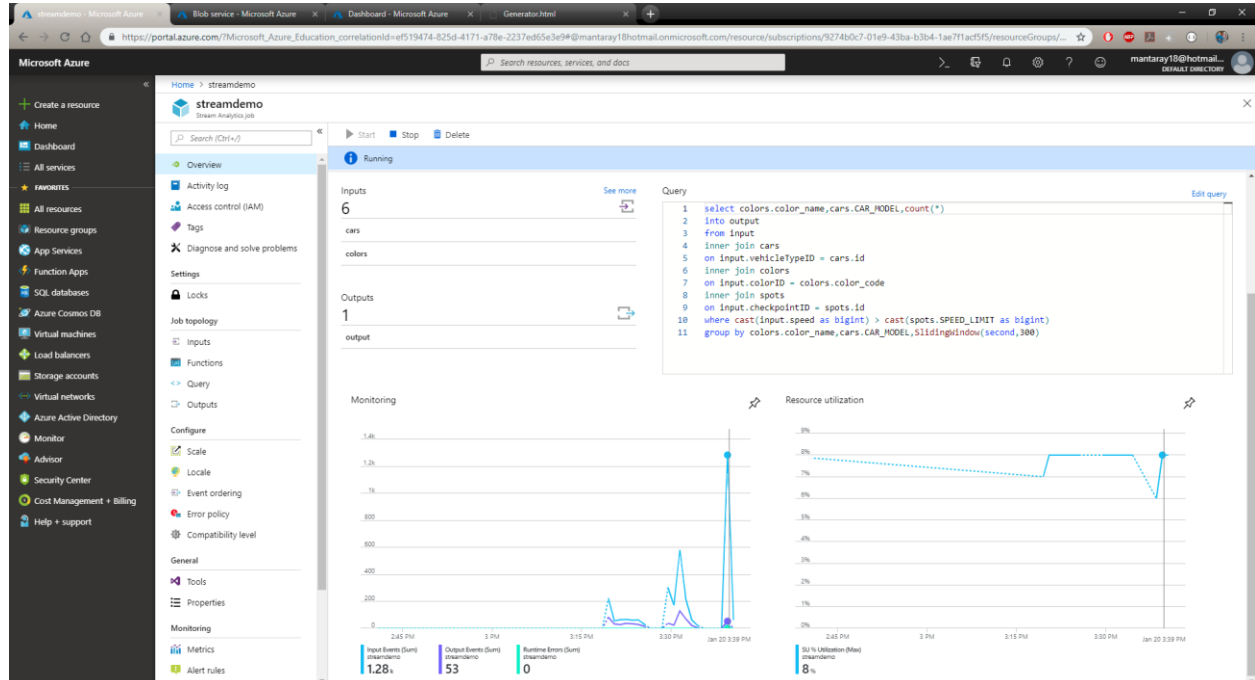


## blob output query 4

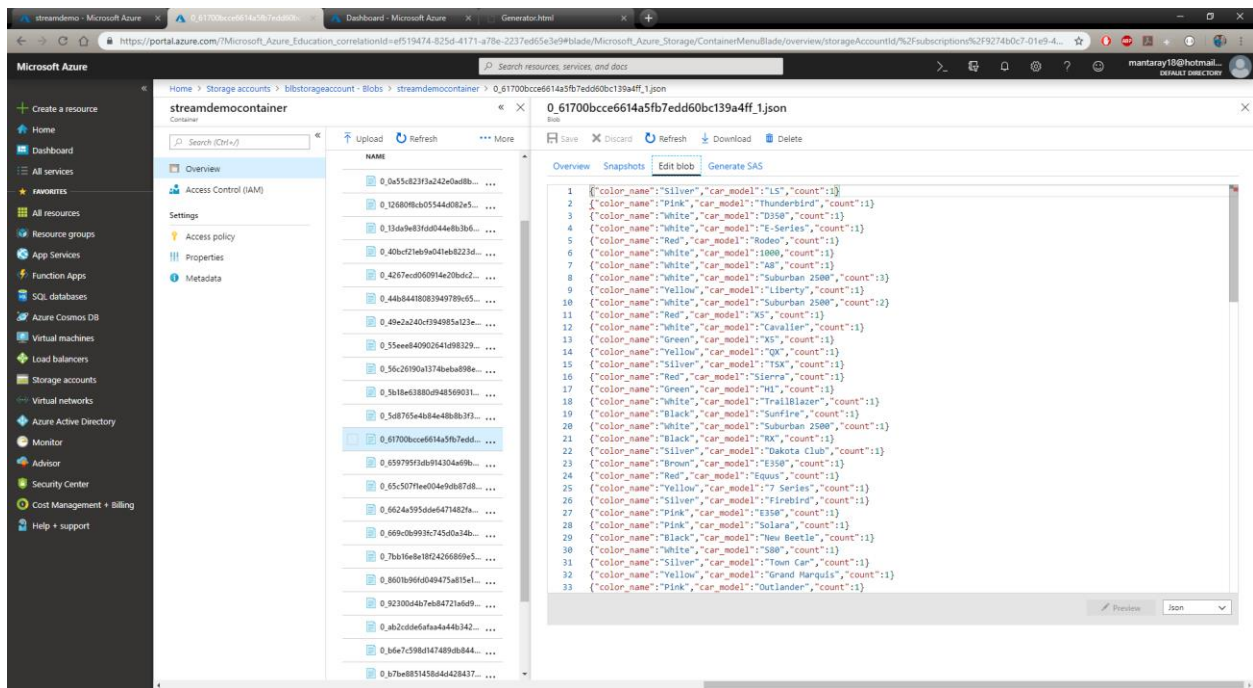


## Query 5

```
select colors.color_name,cars.CAR_MODEL,count(*)  
  
into output  
  
from input  
  
inner join cars  
  
on input.vehicleTypeID = cars.id  
  
inner join colors  
  
on input.colorID = colors.color_code  
  
inner join spots  
  
on input.checkpointID = spots.id  
  
where cast(input.speed as bigint) > cast(spots.SPEED_LIMIT as bigint)  
  
group by colors.color_name,cars.CAR_MODEL,SlidingWindow(second,300)
```



## Blob output query 5



## Query 6

select input.vehicletypeid,input.licensePlate, cars.car\_model,count(\*)

into output

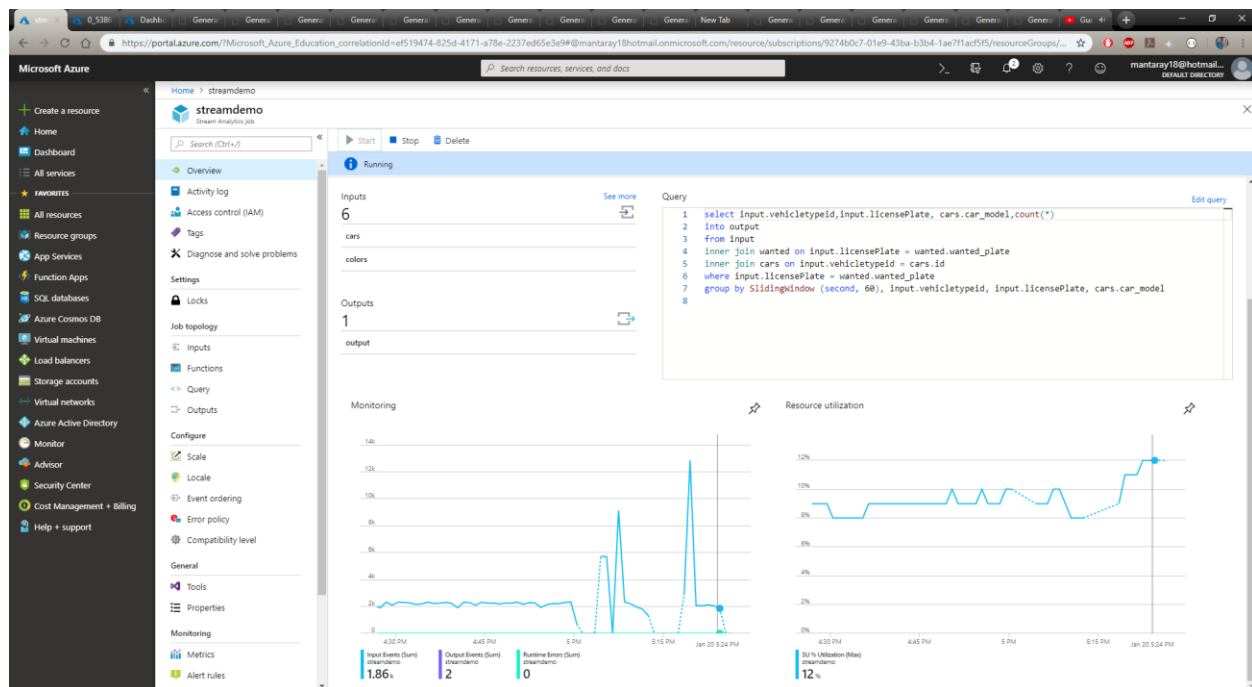
from input

inner join wanted on input.licensePlate = wanted.wanted\_plate

inner join cars on input.vehicletypeid = cars.id

where input.licensePlate = wanted.wanted\_plate

group by SlidingWindow (second, 60), input.vehicletypeid, input.licensePlate, cars.car\_model



### Query 7

```

select temp1.licensePlate ,count(*) as countTimesPassed
into output

```

```

from input as temp1

```

```

join input as temp2

```

```

on datediff(minute, temp1, temp2)

```

```

between 0 and 1

```

```

where temp1.licensePlate = temp2.licensePlate

```

```

and temp1.vehicletypeid != temp2.vehicletypeid

```

```

and temp1.checkpointID != temp2.checkpointid

```

```

and temp1.EventProcessedUtcTime = temp2.EventProcessedUtcTime

```

```

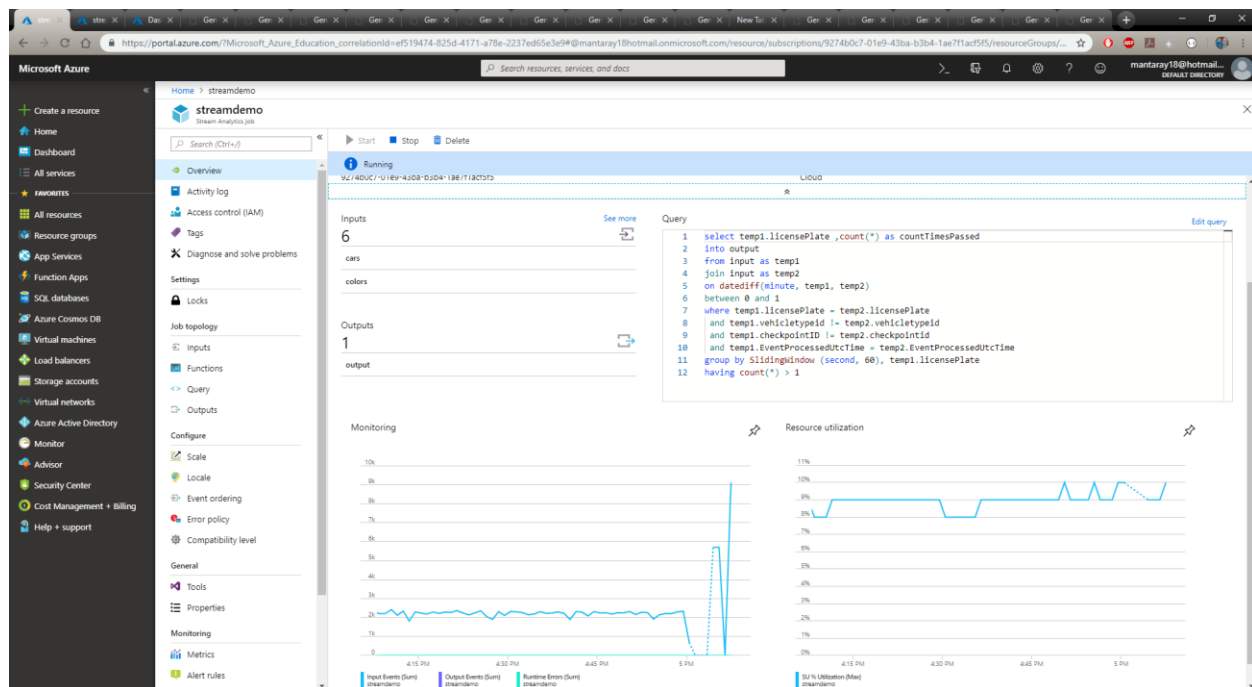
group by SlidingWindow (second, 60), temp1.licensePlate

```

```

having count(*) > 1

```



## Query 8

with

wantedcars as

(select count(\*) as bmwcount

from input

inner join cars on input.vehicleTypeID = cars.id

where cars.car\_make = 'BMW' and input.spotType = 'Speed\_Limit\_Camera'

group by TumblingWindow(second,120)),

abovelimit as

(select count(\*)as bmwviolations

from

input

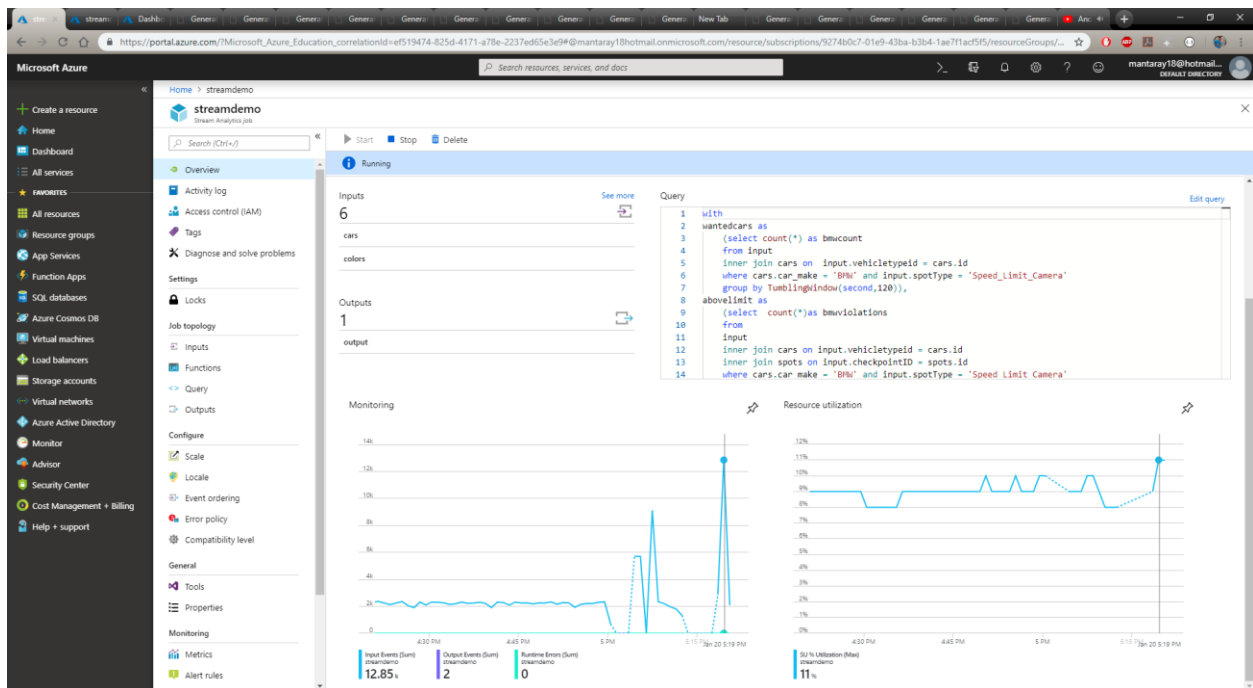
inner join cars on input.vehicleTypeID = cars.id

inner join spots on input.checkpointID = spots.id

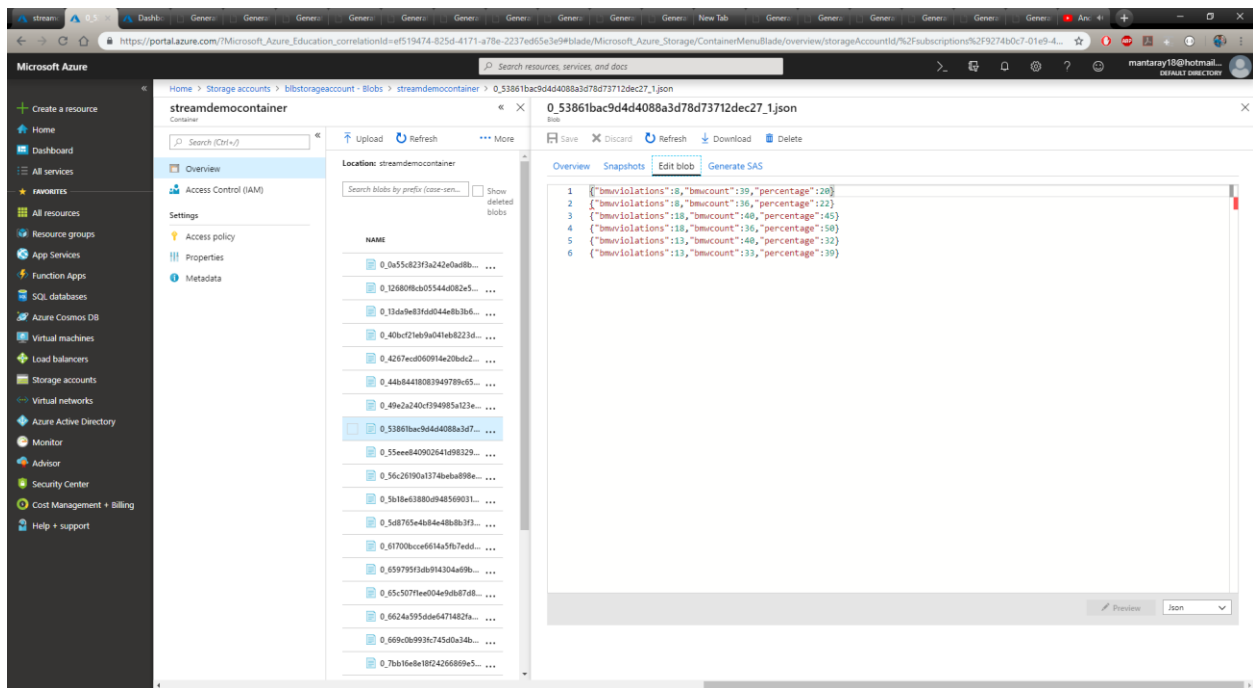
```

where cars.car_make = 'BMW' and input.spotType = 'Speed_Limit_Camera'
and cast(input.speed as bigint) > cast(spots.speed_limit as bigint)
group by TumblingWindow(second,120))
select case when abovelimit.bmwviolations is null then 0 else abovelimit.bmwviolations end as
bmwviolations,
wantedcars.bmwcount,
case when ((abovelimit.bmwviolations)*100)/((wantedcars.bmwcount) is null then 0 else
((abovelimit.bmwviolations)*100)/((wantedcars.bmwcount) end as percentage
from wantedcars
left join abovelimit
on datediff(minute, wantedcars, abovelimit)
between 0 AND 2

```



## Blob output query 8



The screenshot shows the Microsoft Azure portal interface. On the left is the navigation pane with options like 'Create a resource', 'Home', 'Dashboard', 'All services', 'Favorites', 'All resources', 'Resource groups', 'App Services', 'Function Apps', 'SQL databases', 'Azure Cosmos DB', 'Virtual machines', 'Load balancers', 'Storage accounts', 'Virtual networks', 'Azure Active Directory', 'Monitor', 'Advisor', 'Security Center', 'Cost Management + Billing', and 'Help + support'.

The main area displays the 'streamdemocontainer' storage account. The 'Overview' tab is selected, showing a list of blobs. The blob '0\_53861bac9d4d4088a3d78d73712dec27\_1.json' is selected, and its details are shown on the right. The blob's content is displayed as a JSON array of 6 objects, each containing 'bmviolations', 'bmcount', and 'percentage' fields.

```
1 [{"bmviolations":8,"bmcount":39,"percentage":28}
2 [{"bmviolations":8,"bmcount":36,"percentage":22}
3 [{"bmviolations":18,"bmcount":40,"percentage":45}
4 [{"bmviolations":18,"bmcount":36,"percentage":50}
5 [{"bmviolations":13,"bmcount":40,"percentage":32}
6 [{"bmviolations":13,"bmcount":33,"percentage":30}
```