Kafka Streams

1.Terraform apply (updated vm size to 2xStandard_D4_v3):


```
Warning: Deprecated Resource

  with kubernetes_namespace.confluent,
  on main.tf line 114, in resource "kubernetes_namespace" "confluent":
 114: resource "kubernetes_namespace" "confluent" {

 Deprecated; use kubernetes_namespace_v1.

Releasing state lock. This may take a few moments...

Apply complete! Resources: 9 added, 0 changed, 0 destroyed.

Outputs:

acr_login_server = "acrdevwesteuropenq22.azurecr.io"
aks_api_server_url = <sensitive>
aks_kubeconfig = <sensitive>
aks_name = "aks-dev-westeurope-nq22"
client_certificate = <sensitive>
resource_group_name = "rg-dev-westeurope-nq22"
storage_account_name = "stdevwesteuropenq22"
storage_primary_access_key = <sensitive>

yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_pytho
```

2.Confluent Platform pods (decreased number of replicas for zookeper to 1, also reduced amount of allocated resources for some modules):


```
yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_pytho
n_azure-master (main)
$ kubectl get pods -o wide
NAME                                  READY   STATUS    RESTARTS      AGE   IP
          NOMINATED NODE   READINESS GATES
confluent-operator-8446f997d8-8n2gk   1/1     Running   0             23m   10.244.0.154
s000000    <none>           <none>
connect-0                             1/1     Running   0             13m   10.244.0.221
s000000    <none>           <none>
controlcenter-0                       1/1     Running   0             10m   10.244.1.101
s000001    <none>           <none>
elastic-0                             1/1     Running   4 (12m ago)   13m   10.244.1.203
s000001    <none>           <none>
kafka-0                               1/1     Running   0             11m   10.244.1.131
s000001    <none>           <none>
kafka-1                               1/1     Running   0             11m   10.244.1.153
s000001    <none>           <none>
kafka-2                               1/1     Running   0             11m   10.244.0.148
s000000    <none>           <none>
ksqldb-0                              1/1     Running   0             10m   10.244.0.53
s000000    <none>           <none>
schemaregistry-0                      1/1     Running   0             10m   10.244.1.160
s000001    <none>           <none>
zookeeper-0                           1/1     Running   0             13m   10.244.0.134
s000000    <none>           <none>
zookeeper-1                           1/1     Running   0             13m   10.244.1.115
s000001    <none>           <none>
zookeeper-2                           1/1     Running   0             13m   10.244.0.171
s000000    <none>           <none>
```

3.Control center at localhost:
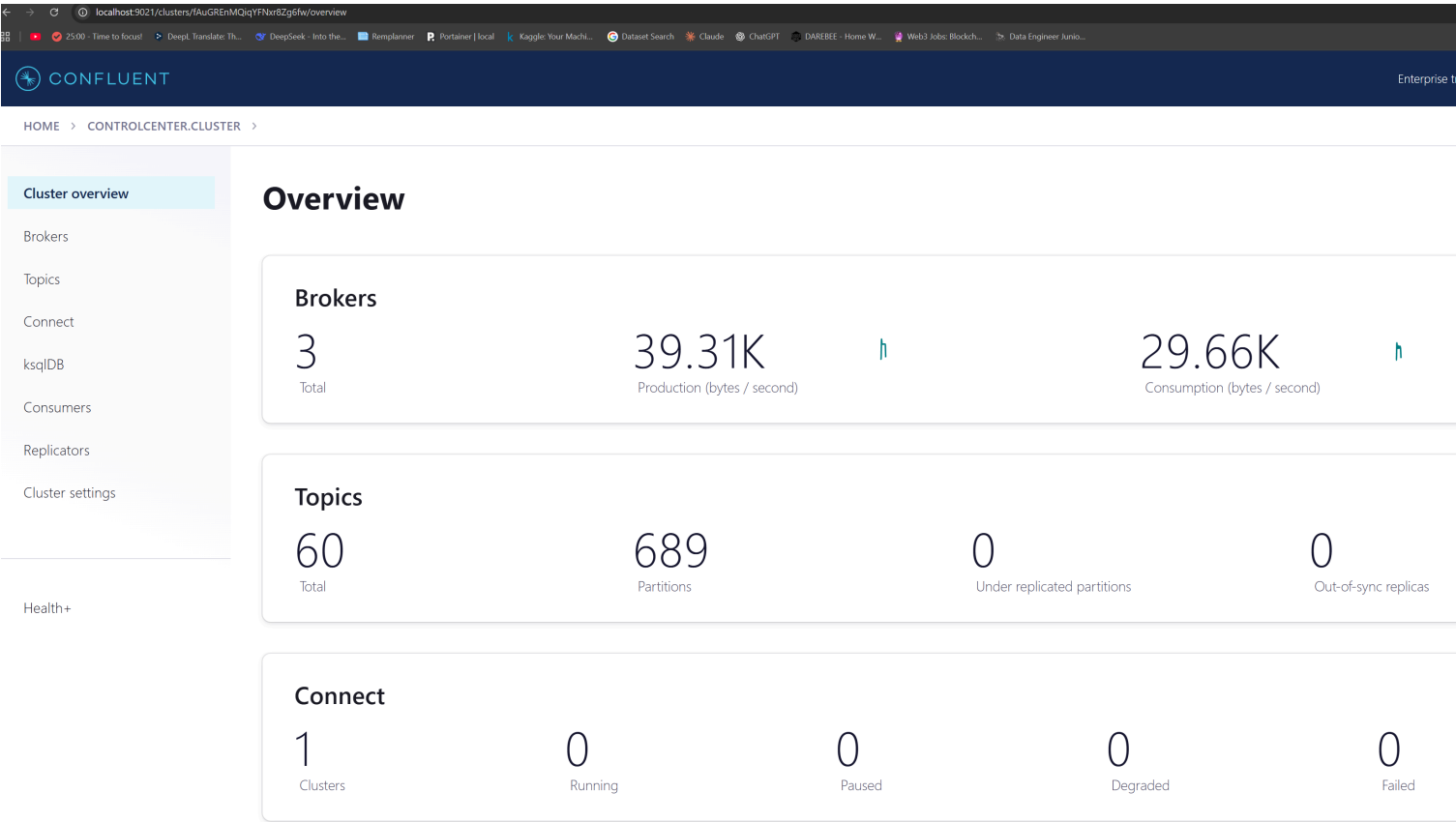
4.Kafka topic creation:

```
PS C:\Users\yura1> Start-Process powershell -WindowStyle Hidden -ArgumentList 'kubectl port-forward connect-0 8083:8083
*> $null'
PS C:\Users\yura1> kubectl exec kafka-0 -c kafka -- bash -c "/usr/bin/kafka-topics --create --topic expedia --replicatio
n-factor 3 --partitions 3 --bootstrap-server kafka:9092"
Created topic expedia.
PS C:\Users\yura1>
```

CONFLUENT

HOME > CONTROLCENTER.CLUSTER > TOPICS >

Cluster overview

Brokers

Topics

Connect

ksqlDB

Consumers

Replicators

Cluster settings

Health+

expedia

Overview   Messages   Schema   **Configuration**

General settings

name	expedia
partitions	3
min.insync.replicas	2
cleanup.policy	delete
retention.ms	604800000
max.message.bytes	1048588
retention.bytes	-1

[Edit settings]   [≡ Show full config]

5.Data loaded to Azure Container:

Home > Resource Manager | Resource groups > rg-dev-westeurope-nq22 > stdevwesteuropenq22 | Containers >

📁 data
Container

🔍 Search   × «

🖥 Overview

🔧 Diagnose and solve problems

👥 Access Control (IAM)

∨ Settings

   🔗 Shared access tokens

   👥 Manage ACL

   🔑 Access policy

+ Add Directory   ↑ Upload   ⟳ Refresh | 🗑 Delete   📋 Copy   📋 Paste   ≡ Ren

🛢 data

Authentication method: Access key (Switch to Microsoft Entra user account)

🔍 Search blobs by prefix (case-sensitive)

Showing all 1 items

	Name	Last modified	
☐	📁 topics	1/18/2026, 7:10:51 PM	

6.azure-source-cc.json prepared:

```
terraform > {} azure-source-cc.json > {} config > ᵃᵇᶜ azblob.account.name
1   {
2       "name": "expedia2",
3       "config": {
4           "connector.class": "io.confluent.connect.azure.blob.storage.AzureBlobStorageSourceConnector",
5           "tasks.max": "2",
6
7           "azblob.account.name": "stdevwesteuropenq22",
8           "azblob.account.key": "███████████████████████████████████████████████████████",
9           "azblob.container.name": "data",
10
11          "format.class": "io.confluent.connect.azure.blob.storage.format.avro.AvroFormat",
12
13          "bootstrap.servers": "kafka:9071",
14          "topics": "expedia",
15          "topics.dir": "topics",
16
17          "transforms": "maskDate",
18          "transforms.maskDate.type": "org.apache.kafka.connect.transforms.MaskField$Value",
19          "transforms.maskDate.fields": "date_time",
20          "transforms.maskDate.replacement": "0000-00-00 00:00:00"
21      }
22  }
```

7.Connector execution:

```
yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_python_azure-master/terraform (main)
$ curl -s -X POST -H "Content-Type:application/json" --data @azure-source-cc.json http://localhost:8083/connectors
```

8.Expedia Topic + Data masking demonstration:

9.Output kafka topic creation:

```
PS C:\Users\yura1> kubectl exec kafka-0 -c kafka -- bash -c "/usr/bin/kafka-topics --create --topic expedia_ext --replic
ation-factor 3 --partitions 3 --bootstrap-server kafka:9092"
WARNING: Due to limitations in metric names, topics with a period ('.') or underscore ('_') could collide. To avoid issu
es it is best to use either, but not both.
Created topic expedia_ext.
PS C:\Users\yura1>
```

10.Pushed and deployed stram-app:

```
yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_python_azure-master (main)
$ docker push "acrdevwesteuropenq22.azurecr.io"/stream-app:latest
The push refers to repository [acrdevwesteuropenq22.azurecr.io/stream-app]
138e27cba350: Pushed
7e95ebb70537: Pushed
89d573bf42b3: Pushed
91c91c91f1d2: Pushed
79d5bd8a8d26: Pushed
795dbedde24d: Pushed
1633a45502fe: Pushed
4f4fb700ef54: Pushed
081ccf923272: Pushed
c9723aa529b0: Pushed
26dfe2fac1c4: Pushed
260c7d06cc92: Pushed
4e39a27c4e8c: Pushed
latest: digest: sha256:313b1674a965df7ad86b09e0d760dfc602d82ede8c2655d96240ce99445e7bb2 size: 856

yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_python_azure-master (main)
$ kubectl apply -f stream-app.yaml
deployment.apps/kstream-app created
```

11.Messages available in expedia_ext:

expedia_ext

Overview   **Messages**   Schema   Configuration

Production	Consumption	Total messages
0 B/s	0 B/s	41,220

Filter by timestamp, offset, key or value | All partitions | Latest | Max 50 results

50 messages shown   ● Auto-refresh on

Timestamp	Offset	Partition	Key	Value
1768760571085	13866	1	""	{"id":2506661,"date_time":"0000-00-00 00:00:00","site_name":2,"posa_container":null,"user_
1768760571080	13865	1	""	{"id":2506622,"date_time":"0000-00-00 00:00:00","site_name":2,"posa_container":null,"user_
1768760571075	13663	2	""	{"id":2506595,"date_time":"0000-00-00 00:00:00","site_name":2,"posa_container":null,"user_
1768760571068	13662	2	""	{"id":1034301,"date_time":"0000-00-00 00:00:00","site_name":2,"posa_container":null,"user_
1768760571065	13661	2	""	{"id":2506552,"date_time":"0000-00-00 00:00:00","site_name":11,"posa_container":null,"user

Message details

Timestamp	Offset	Parti...
1/18/2026, 7:22:51 PM	13650	2
(1768760571000)		

Key   **Value**   Headers

```
 3      "date_time": "0000-00-00 00:00:00",
 4      "site_name": 2,
 5      "posa_container": null,
 6      "user_location_country": 66,
 7      "user_location_region": 322,
 8      "user_location_city": 47357,
 9      "orig_destination_distance":
        361.9836,
10      "user_id": 1129228,
11      "is_mobile": 0,
12      "is_package": 0,
13      "channel": 10,
14      "srch_ci": "2016-10-24",
15      "srch_co": "2016-10-26",
16      "srch_adults_cnt": 1,
17      "srch_children_cnt": 1,
18      "srch_rm_cnt": 1,
19      "srch_destination_id": 12974,
20      "srch_destination_type_id": 4,
21      "hotel_id": 142592914227.6,
22      "stay_category": "Short stay",
23      "__faust": {
24          "ns": "main.ExpediaExtRecord"
25      }
26  }
```

12.KSQL Create stream and table:

```
having trouble? Type 'help' (case-insensitive) for a rundown of how things work.

ksql> CREATE STREAM expedia_stream (
>         id BIGINT,
>         hotel_id BIGINT,
>         stay_category VARCHAR
>     ) WITH (
>         KAFKA_TOPIC='expedia_ext',
>         VALUE_FORMAT='JSON'
>     );

 Message
----------------
 Stream created
----------------
ksql>  CREATE TABLE hotels_count AS
>     SELECT
>         stay_category,
>         COUNT(hotel_id) AS hotels_amount,
>         COUNT_DISTINCT(hotel_id) AS distinct_hotels
>     FROM expedia_stream
>     GROUP BY stay_category;

 Message
-------------------------------------------
 Created query with ID CTAS_HOTELS_COUNT_1
-------------------------------------------
ksql>     SELECT * FROM hotels_count EMIT CHANGES;
+---------------------------------+----------------------------------+----------------------------------+
|STAY_CATEGORY                    |HOTELS_AMOUNT                     |DISTINCT_HOTELS                   |
+---------------------------------+----------------------------------+----------------------------------+
|Erroneous data                   |16                                |16                                |
|Standard extended stay           |20                                |20                                |
|Long stay                        |11                                |11                                |
|Standard stay                    |479                               |439                               |
|Short stay                       |4349                              |2047                              |
|Short stay                       |5125                              |2170                              |
|Long stay                        |13                                |13                                |
|Standard stay                    |564                               |509                               |
|Erroneous data                   |18                                |18                                |
|Standard extended stay           |26                                |26                                |

Press CTRL-C to interrupt
```

12.KSQL FLOW:

CONFLUENT

ksqldb

Editor | **Flow** | Streams | Tables | Persistent queries | Settings

🔍 Search

| EXPEDIA_STREAM | CREATE-TABLE | HOTELS_COUNT |
| STREAM | | TABLE |

HOTELS_COUNT ✕

Data structure TABLE

Total rows

Updates /sec

Messages | **Schema**

▶ ⏸ | 🔍 Filter by keyword | ☰ ▥

No new messages

The message browser shows messages that have arrived since this page was opened.

ksqldb

Editor   Flow   Streams   Tables   Persistent queries   Settings

```
1    select * from  HOTELS_COUNT;
```

● Add query properties

| auto.offset.reset | = | Latest ▾ | 🗑 |

＋Add another field

Unable to run query.   **Stop**   **Run query**

⚠ An unknown error has occured. Check the connection settings.   ✕

For me, ksql is not working in the UI. I verified that the pod is running and tried forwarding an additional port for ksql, but it didn't help. Since everything works correctly from the command line, I conclude that this is a Windows port-forwarding–related issue.

```
yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_python_azure-master (main)
$ kubectl get pod ksqldb-0
NAME        READY    STATUS     RESTARTS    AGE
ksqldb-0    1/1      Running    0           36m

yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_python_azure-master (main)
$ kubectl exec -it ksqldb-0 -- curl -s http://localhost:8088/info
Defaulted container "ksqldb" out of: ksqldb, config-init-container (init)
{"KsqlServerInfo":{"version":"7.8.0","kafkaClusterId":"fAuGREnMQiqYFNxr8Zg6fw","ksqlServiceId":"confluent.ksqldb_",
"serverStatus":"RUNNING"}}
yura1@pc MINGW64 ~/Desktop/New folder/hw/test/kafka/task3/m12_kafkastreams_python_azure-master (main)
$ kubectl port-forward ksqldb-0 8088:8088
Forwarding from 127.0.0.1:8088 -> 8088
Forwarding from [::1]:8088 -> 8088
```

Connectio string:

ksqldb

Editor **Flow** **Streams** **Tables** **Persistent queries** **Settings**

General

Cluster name	ksqldb
Cluster ID	ksqldb
Endpoint	http://ksqldb.confluent.svc.cluster.local:8088
Version	7.8.0 | View documentation