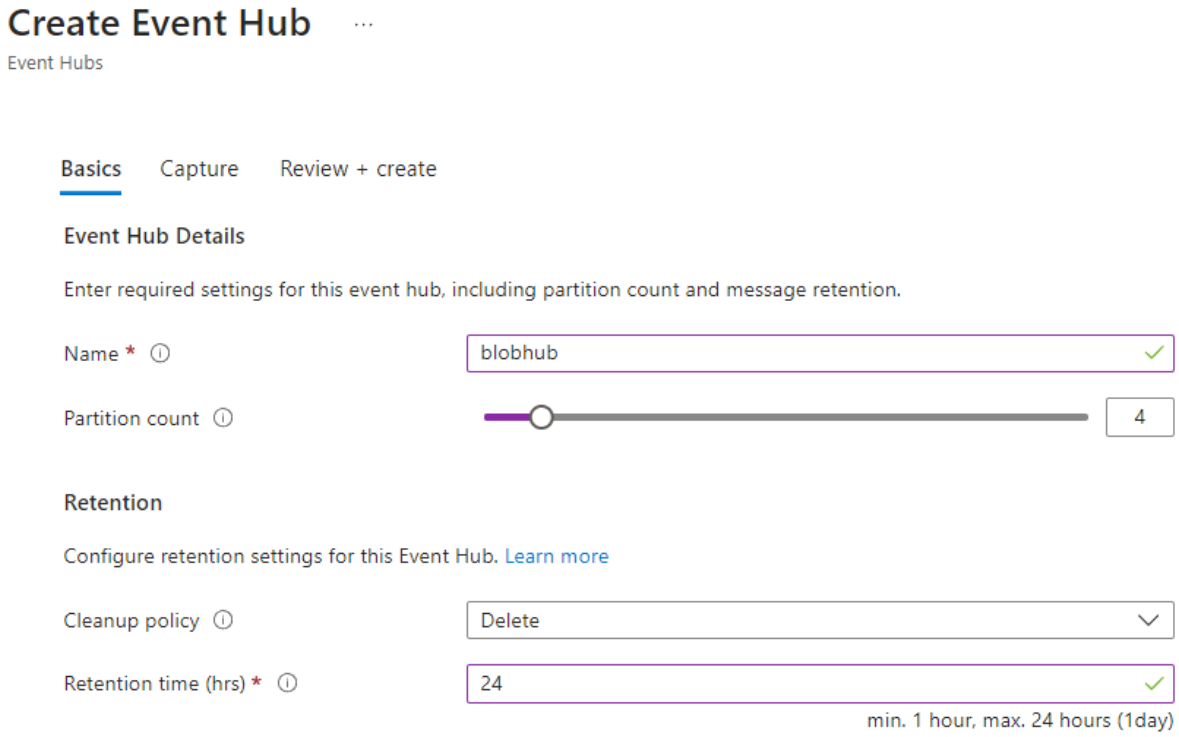
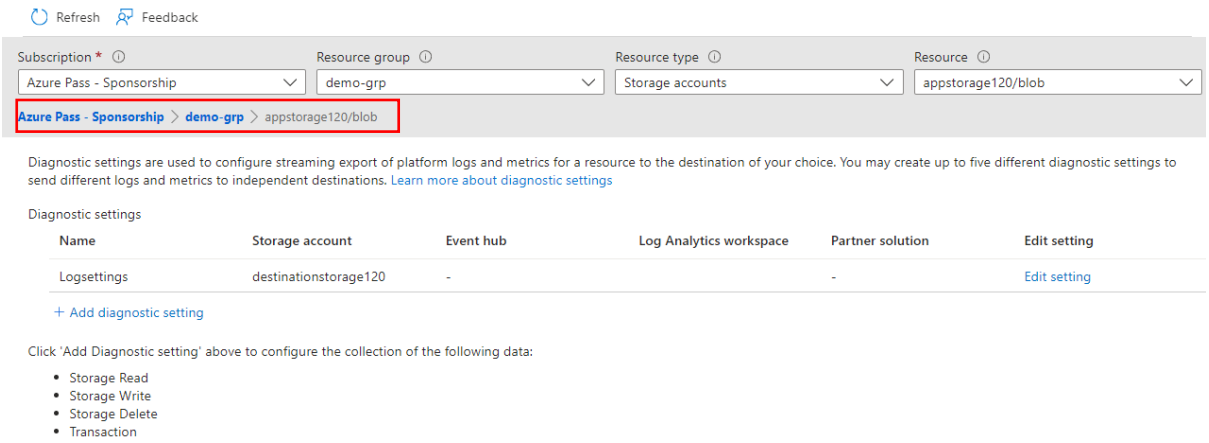
😊 Sending data to Event Hub

In this lab, the goal is to establish a data pipeline for monitoring and analyzing Blob storage activity. Initially, Blob diagnostic data is sent to Azure Event Hub for real-time processing. Then, using Stream Analytics, the data is parsed and transformed into a structured format. Finally, the processed data is stored in an SQL Database for further analysis, reporting, or integration with other systems. The end goal is to gain insights into Blob storage usage patterns, optimize performance, and ensure compliance with security and regulatory requirements.

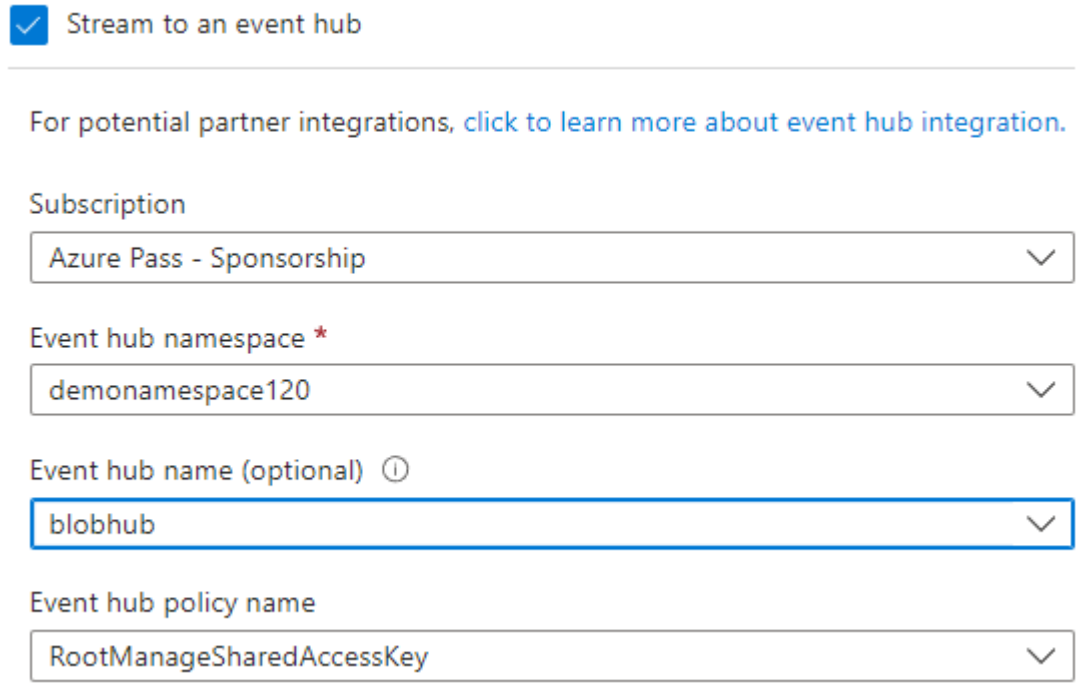
1. In this lab we want to send our Blob diagnostic data onto Azure Event Hub and then try and consume that data from our stream analytics job.
2. The first thing that we are going to do is to create an Event Hub.
3. Now you need to give your event a name then you can increase the partition count.
4. After that you need to increase the retention time to 24 hours.



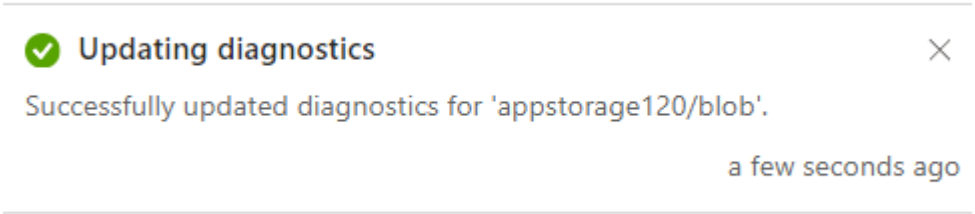
5. Then just create your Event Hub.
6. Once you have this in place, then you move to your storage account where you created your diagnostic settings. Open your blob diagnostic settings.
7. Then click on edit settings.



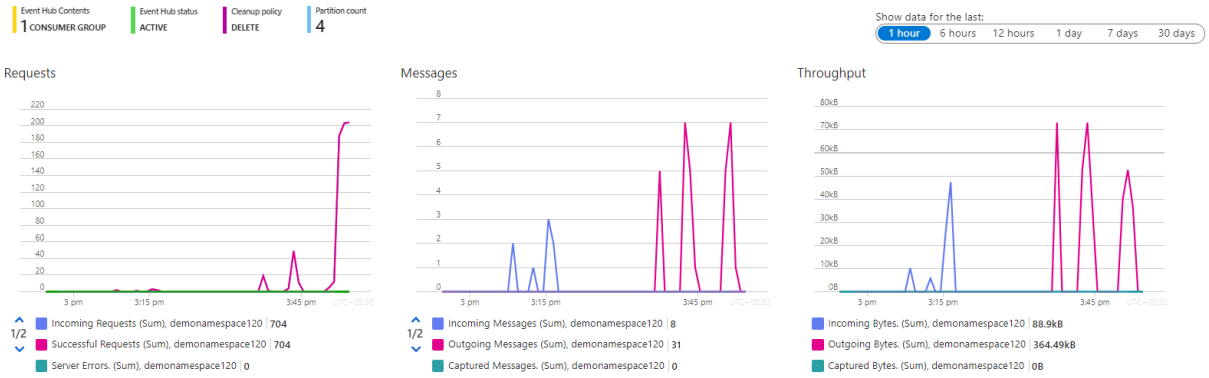
8. Now in there you have to enable Stream to an event hub. Then you need to choose your event hub name and click on save.



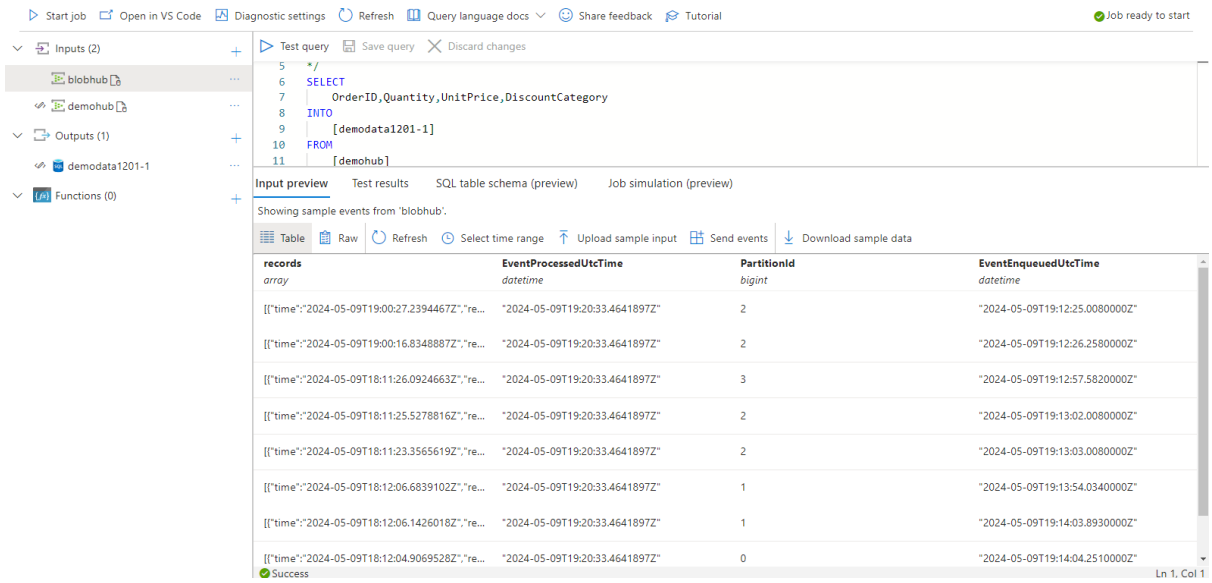
9. Once your diagnosis is complete then you have to wait for 10 mins.
10. Or what you can do is again you can create containers and upload files on them.



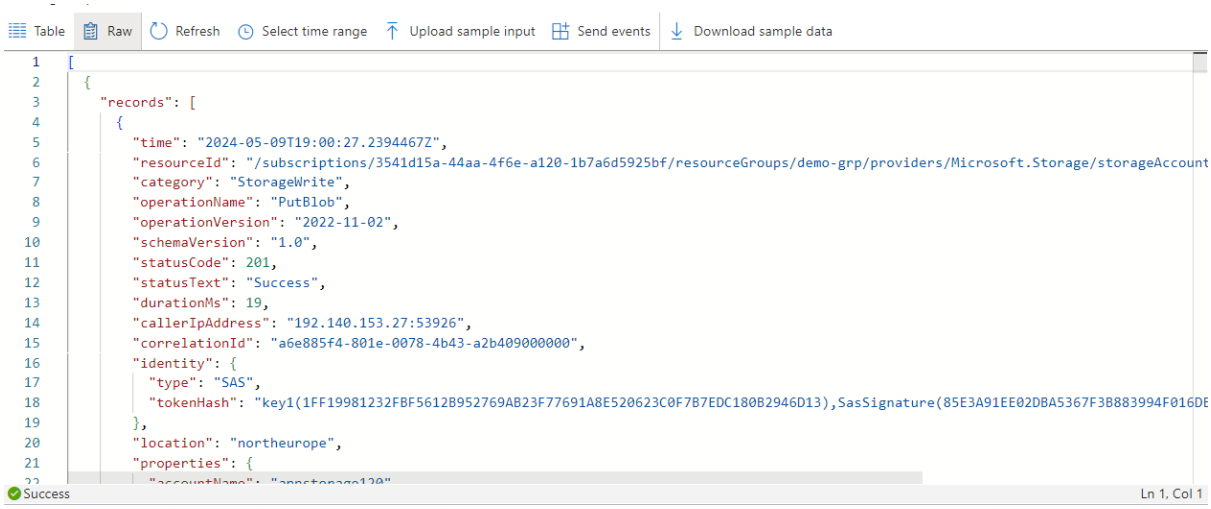
11. After some time, you need to go to your Event hub and open the blob hub event and you will see that the data was requested here.



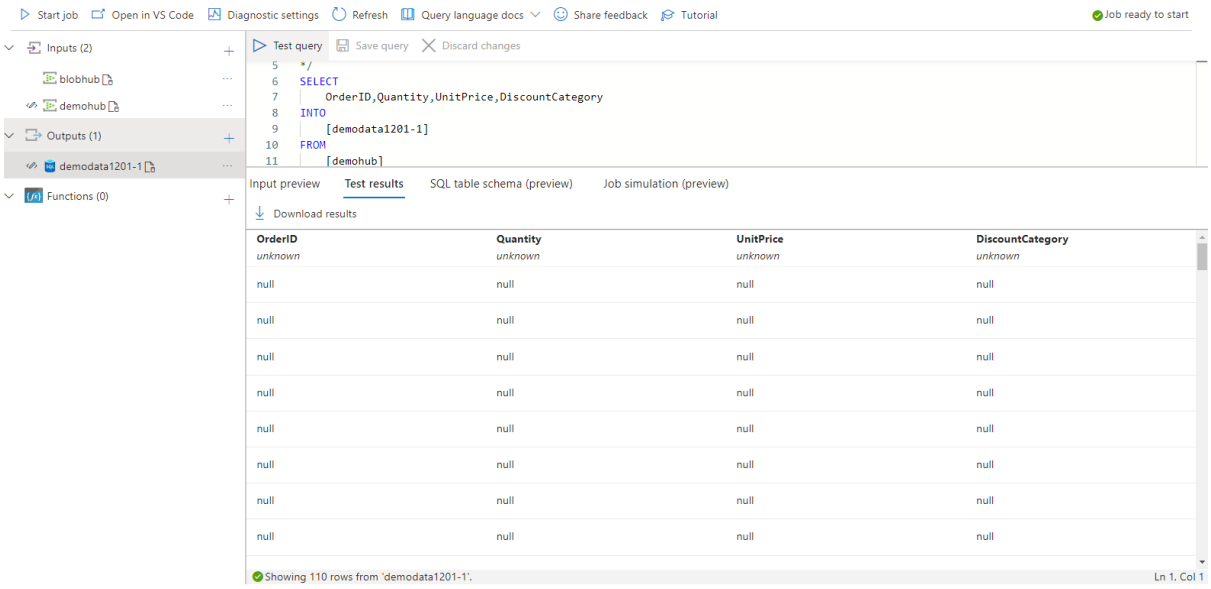
12. Now you need to go to Stream analytics then go towards inputs and add an input for the event hub which will be the blob hub.
13. Just keep things to default and click on save.
14. After that you need to go to query and when you select blob hub you will be able to see the information. But this time it is different than the last time.



15. Here you can see that if you click on Raw then you will see the data in JSON format. Now you will also notice that the data is in the form of records. Also, this is in the form of Array.



16. Then if we click on the test query you will see that we got all the Null values.



😊 Formulating Our Query:

1. Now inside of your query editor you will write this query.

```
*/
SELECT
   Records.ArrayValue.time AS TimeGenerated,
   Records.ArrayValue.resourceId AS ResourceId
INTO
   [demodata1201-1]
FROM
   [blobhub] b
   CROSS APPLY GetArrayElements(b.records) As Records
```

Test query Save query Discard changes

```
/*
Here are links to help you get started with Stream Analytics Query Language:
Common query patterns - https://go.microsoft.com/fwLink/?LinkID=619153
Query language - https://docs.microsoft.com/stream-analytics-query/query-language-elements-azure-stream-analytics
*/
SELECT
    Records.ArrayValue.time AS TimeGenerated,
    Records.ArrayValue.resourceId AS ResourceId
INTO
    [demodata1201-1]
FROM
    [blobhub] b
    CROSS APPLY GetArrayElements(b.records) As Records
```

2. Then you need to click on the test query and below are the rest for that query.

Input preview | Test results | SQL table schema (preview) | Job simulation (preview)

Download results

TimeGenerated *datetime*	ResourceId *string*
"2024-05-09T19:00:27.2394467Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T19:00:16.8348887Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T19:00:17.0202768Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T19:00:26.2784455Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T19:00:26.4486803Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T19:00:27.0564926Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T19:00:27.4450164Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T18:11:26.0924663Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...
"2024-05-09T18:11:25.5278816Z"	"/subscriptions/3541d15a-44aa-4f6e-a120-1b7a6d5925bf/resourceGroups/demo-grp/provi...

Showing 42 rows from 'demodata1201-1'. Ln 12, Col 16

3. Now from above we can understand that we selected just two records to display, and it did that. But now we will select all the records in place and display them.
4. For that we will use the query shown below.

```
*/
SELECT
   Records.ArrayValue.time AS TimeGenerated,
   Records.ArrayValue.category AS Category,
   Records.ArrayValue.operationName AS OperationName,
   Records.ArrayValue.statusCode AS StatusCode,
   Records.ArrayValue.callerIpAddress AS CallerIpAddress,
   Records.ArrayValue.[identity].type AS IdentityType
INTO
   [demodata1201-1]
FROM
   [blobhub] b
   CROSS APPLY GetArrayElements(b.records) As Records
```

Test query | Save query | Discard changes

```
/*
Here are links to help you get started with Stream Analytics Query Language:
Common query patterns - https://go.microsoft.com/fwLink/?LinkID=619153
Query language - https://docs.microsoft.com/stream-analytics-query/query-language-elements-azure-stream-analytics
*/
SELECT
    Records.ArrayValue.time AS TimeGenerated,
    Records.ArrayValue.category AS Category,
    Records.ArrayValue.operationName AS OperationName,
    Records.ArrayValue.statusCode AS StatusCode,
    Records.ArrayValue.callerIpAddress AS CallerIpAddress,
    Records.ArrayValue.[identity].type AS IdentityType
INTO
    [demodata1201-1]
FROM
    [blobhub] b
    CROSS APPLY GetArrayElements(b.records) As Records
```

5. Now for the above query run the test and you will see the data.

Input preview | Test results | SQL table schema (preview) | Job simulation (preview)

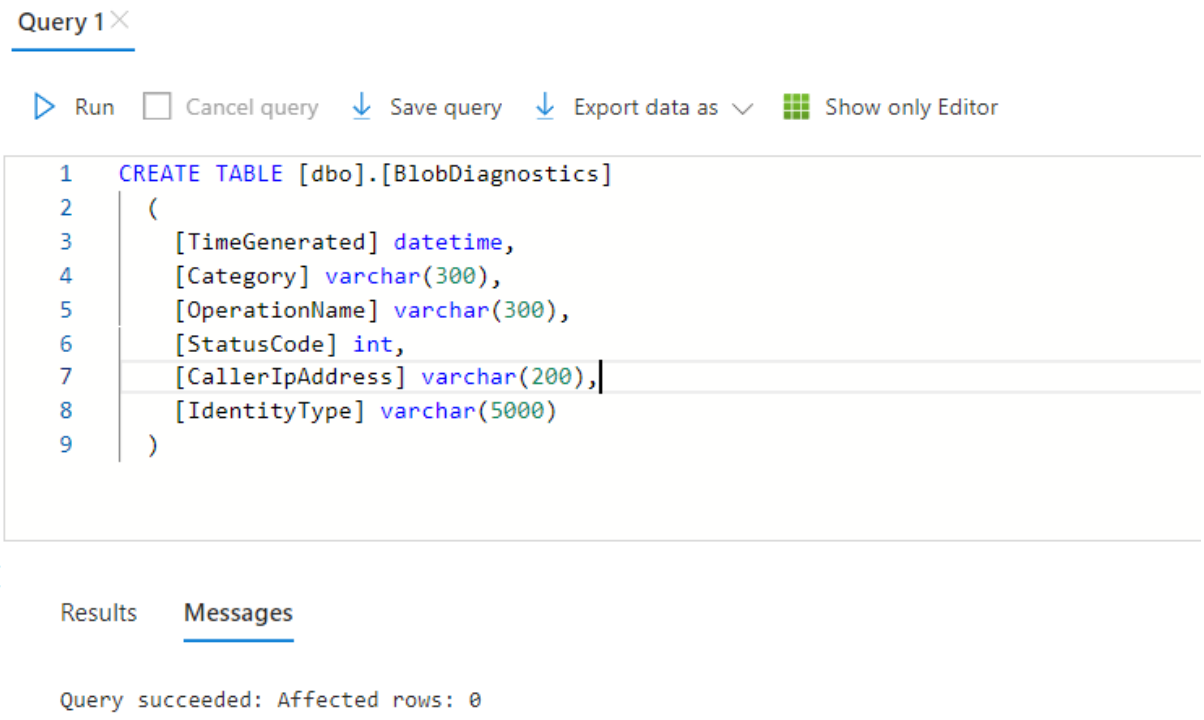
Download results

TimeGenerated *datetime*	Category *string*	OperationName *string*	StatusCode *bigint*	CallerIpAddress *string*	IdentityType *string*
"2024-05-09T19:00:27.2394...	"StorageWrite"	"PutBlob"	201	"192.140.153.27:53926"	"SAS"
"2024-05-09T19:00:16.8348...	"StorageRead"	"BlobPreflightRequest"	200	"192.140.153.27:53926"	"AnonymousPreflight"
"2024-05-09T19:00:17.0202...	"StorageRead"	"ListBlobs"	200	"192.140.153.27:53926"	"SAS"
"2024-05-09T19:00:26.2784...	"StorageRead"	"BlobPreflightRequest"	200	"192.140.153.27:53926"	"AnonymousPreflight"
"2024-05-09T19:00:26.4486...	"StorageRead"	"GetBlobProperties"	404	"192.140.153.27:53926"	"SAS"
"2024-05-09T19:00:27.0564...	"StorageRead"	"BlobPreflightRequest"	200	"192.140.153.27:53926"	"AnonymousPreflight"
"2024-05-09T19:00:27.4450...	"StorageRead"	"ListBlobs"	200	"192.140.153.27:53926"	"SAS"
"2024-05-09T18:11:26.0924...	"StorageRead"	"GetContainerProperties"	404	"10.0.11.28:41352"	"TrustedAccess"
"2024-05-09T18:11:25.5278...	"StorageWrite"	"CreateContainer"	201	"10.0.11.28:40472"	"TrustedAccess"

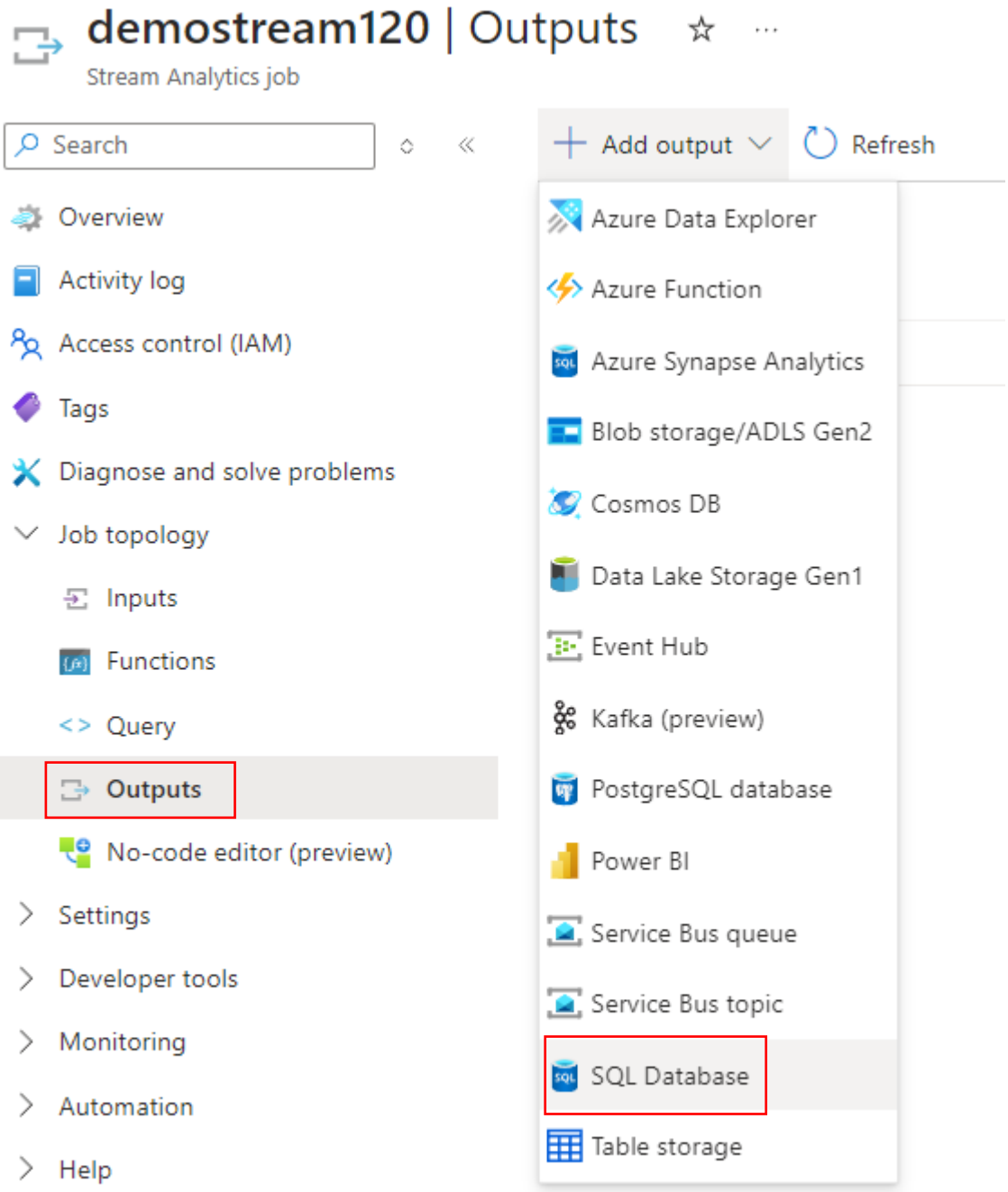
Showing 42 rows from 'demodata1201-1'. Ln 17, Col 55

😄 Reading Blob Diagnostic Data

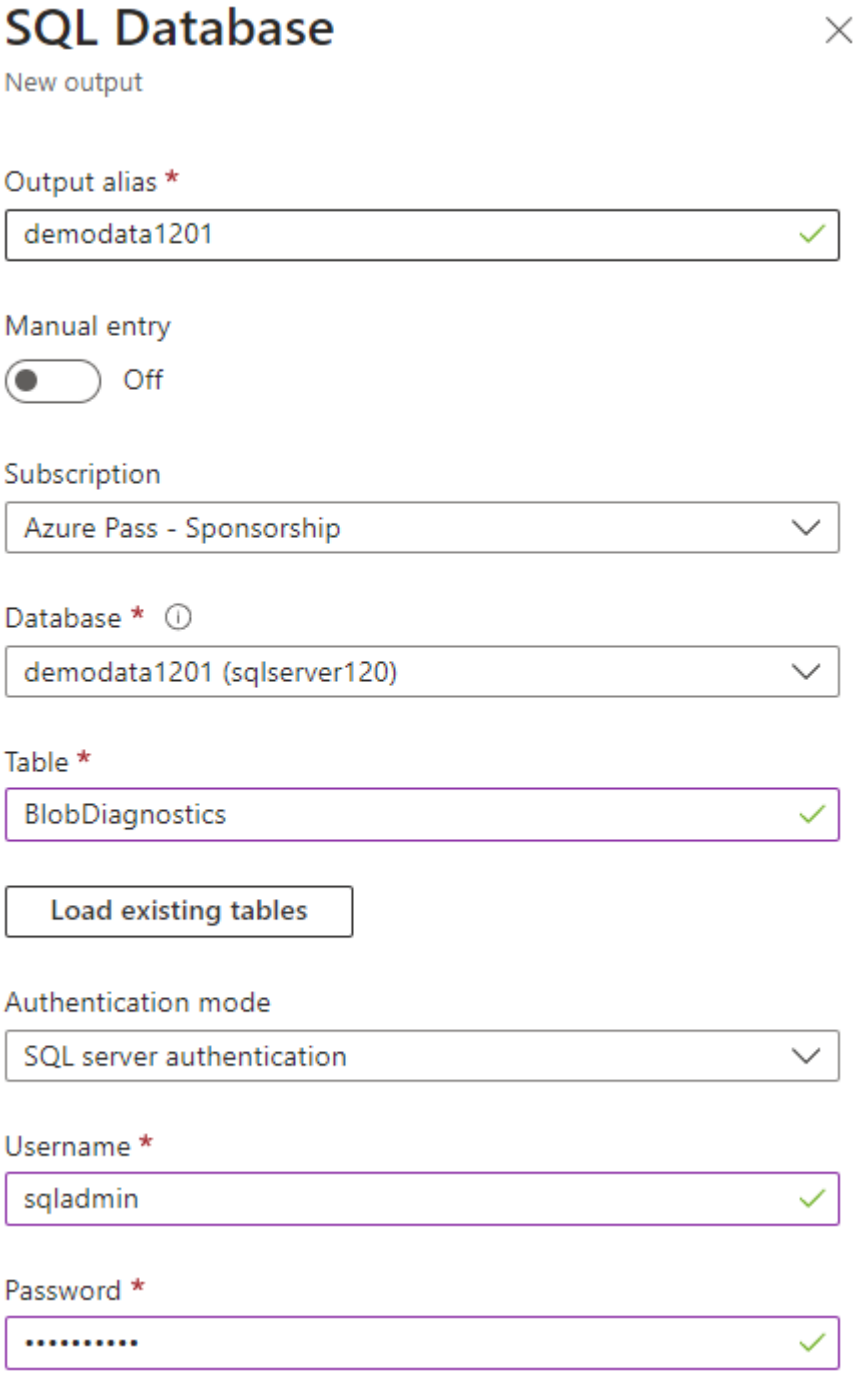
1. Now to read the blob diagnostic data first we need to create a table.
2. We will create a table based on the information that we provided in the query.
3. Now you can either create a table using SSMS or just open the query editor for in your SQL Database.



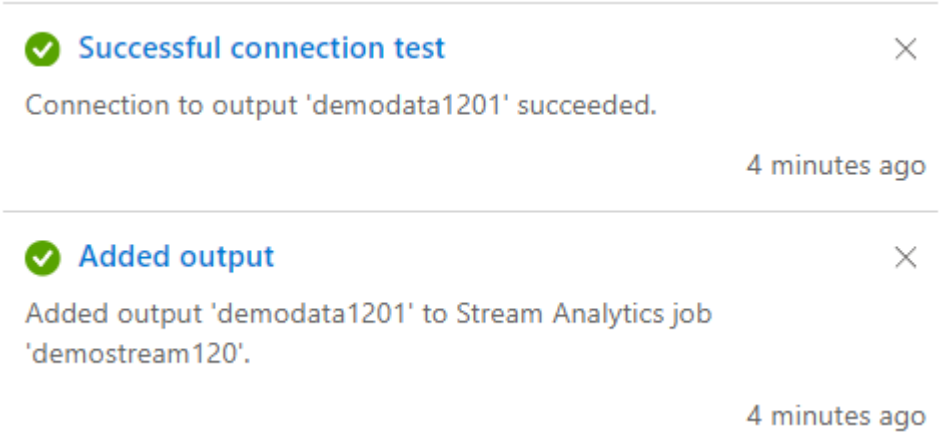
4. Once your table is created, then you need to navigate to Stream Analytics and create an output.
5. Click on Add Output and choose SQL Database.



6. Now just give the name of the table then write your username and password then click on save.

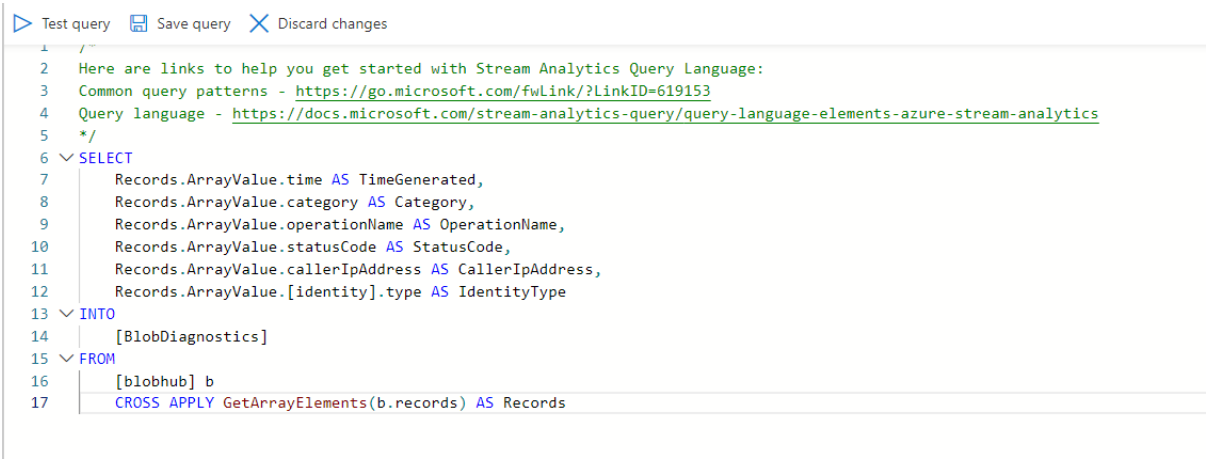


7. Below you can see that output was added and the connection was successful.

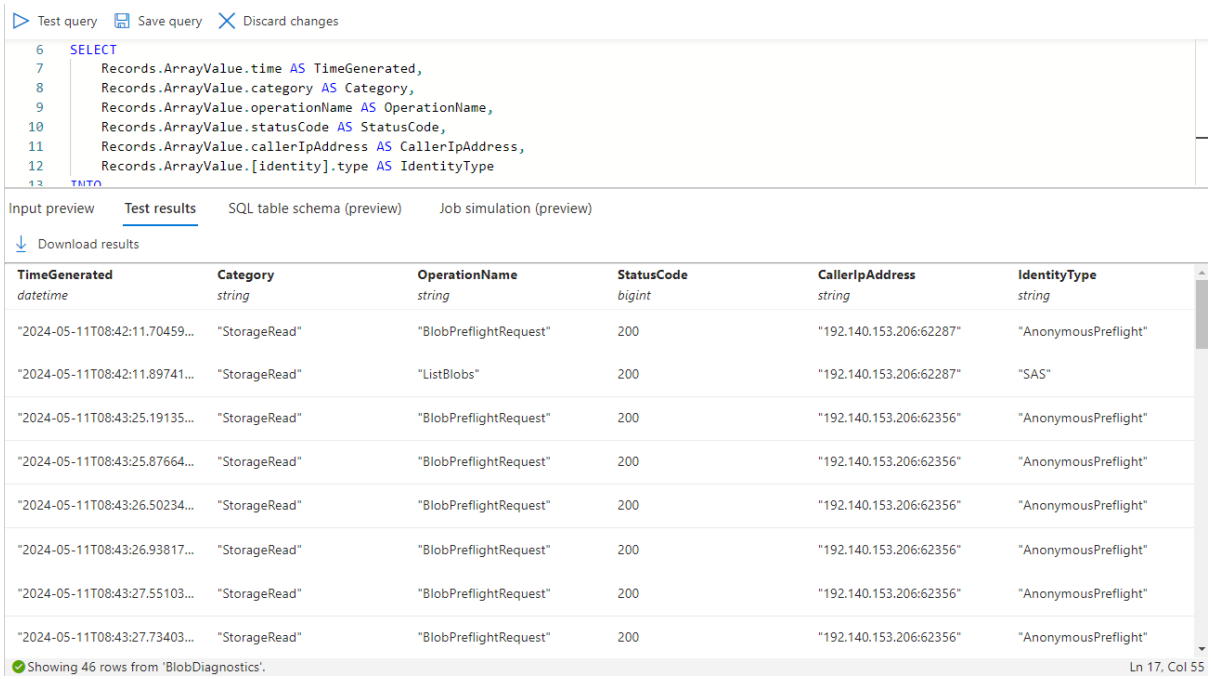


8. Then go to query and use the query mentioned below.

```
SELECT
    Records.ArrayValue.time AS TimeGenerated,
    Records.ArrayValue.category AS Category,
    Records.ArrayValue.operationName AS OperationName,
    Records.ArrayValue.statusCode AS StatusCode,
    Records.ArrayValue.callerIpAddress AS CallerIpAddress,
    Records.ArrayValue.[identity].type AS IdentityType
INTO
    [BlobDiagnostics]
FROM
    [blobhub] b
    CROSS APPLY GetArrayElements(b.records) AS Records
```



9. Then first you need to test the query, save this query, and go back to the overview of stream analytics.



10. From there you are going to start this, Job.
11. Then wait for 10-15 minutes.
12. Go back to the Query editor in the SQL database OR use SSMS and run the select statement for your table.
13. Below you can see that data in place.