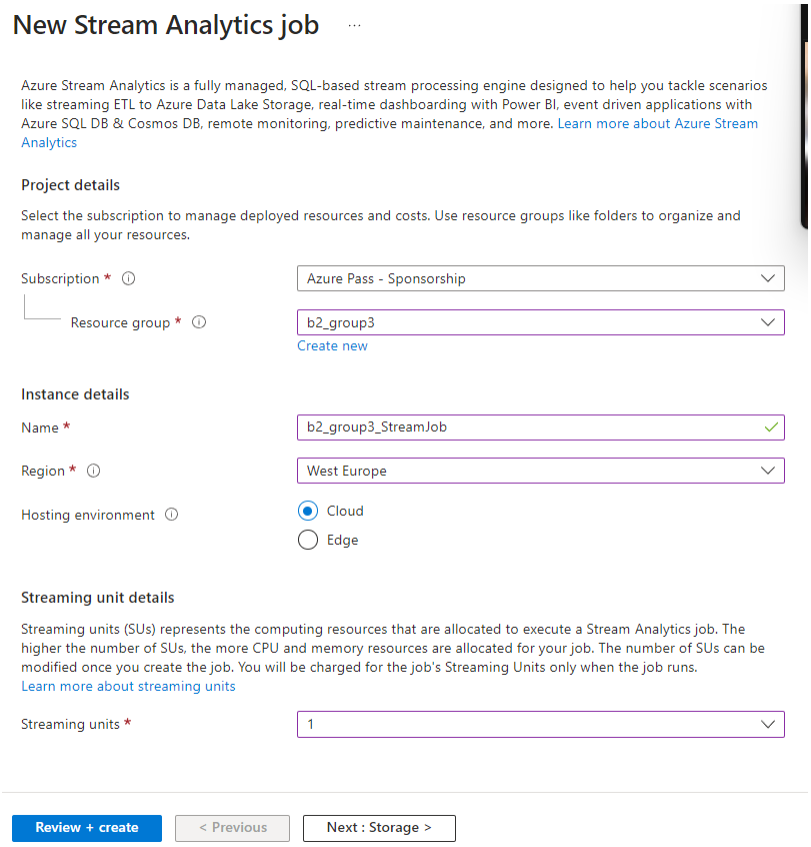
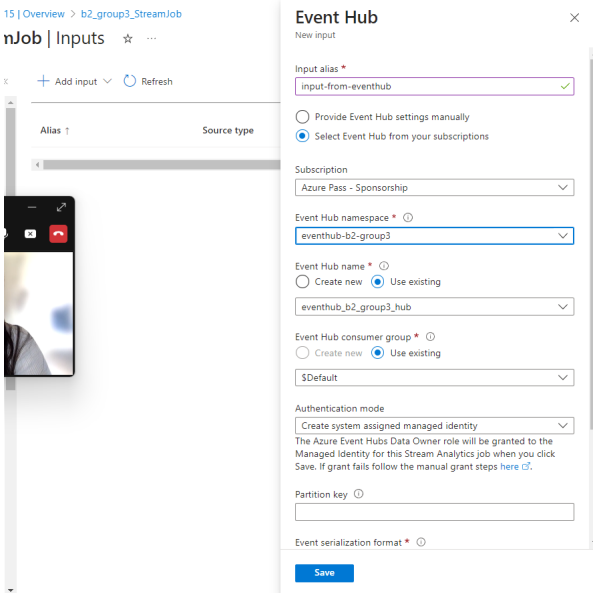
1. Created a new Stream Analytics job and chose to use 1 streaming Unit



Then configured the Input for the stream:



Remark: will we need a partition key? Not specified now.

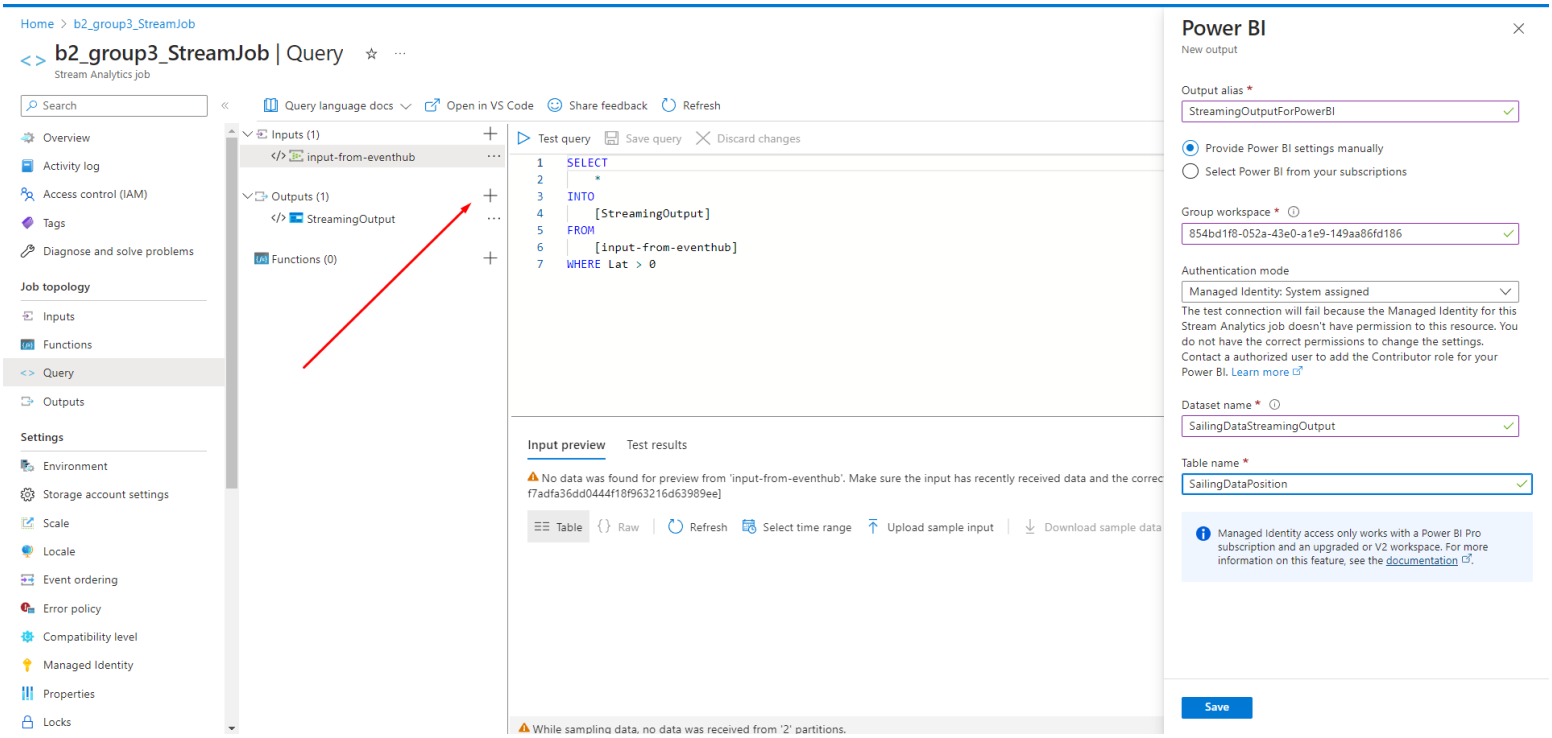
3. Configured an output sink:

Created a storage account (storageb2group3) with a Blob Storage Container(blobstorageb2group3).

4. Configured the output in the Stream Analytics Job: StreamingOutputBlobStorage

5. Selected ‘Run’on the Stream Analytics Job overview and took a first look at the results. Surprisingly, 2 datasets (JSON) were created in our BlobStorage for data with PartitionID 0 or 1…

6. Connection to PowerBI:



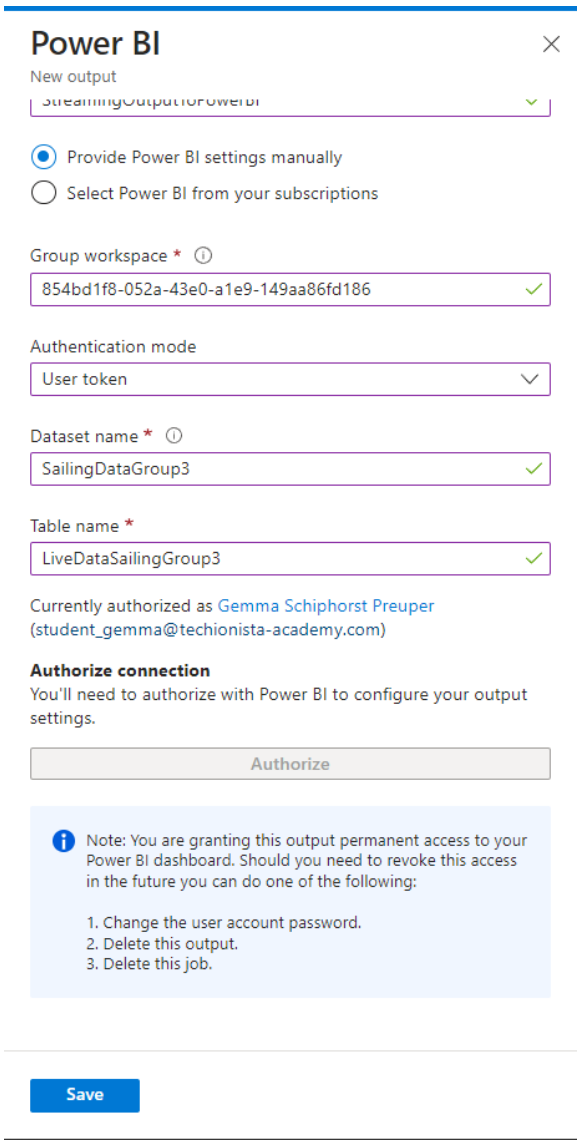
Edited the Query to only select Latitude and Longitude columns:

Test query Save query Discard changes

```
SELECT
    *
INTO
    [StreamingOutput]
FROM
    [input-from-eventhub]
WHERE
    latitude > 0

SELECT
    latitude
    longitude
INTO
    [StreamingOutputForPowerBI]
FROM
    [input-from-eventhub]
WHERE
    latitude > 0
```

PowerBI Streaming Settings:



Idea: make a Gauge Chart to see how far they've come compared to the total distance
Idea: add speed of a few boats to see their speed over the last minutes

How to decide rank: by distance from starting point with a formula.

- add a column called 'distance'
- Add the formula

```
import math

latitude = 38.586562661789316
longitude = -9.429378108391333

latCascais = 38.69225437789037
lonCascais = -9.419236159278585
R = 6378.137 km

a = (math.sin() ** 2) * ((latitude - latCascais)/2) + (math.cos(latCascais)) *
(math.cos(latitude)) * (math.sin() ** 2) * ((longitude - lonCascais)/2)
```

d = 2 * R * ((sin ** -1) * (sqrt(a))

print(d)

7. Setup a Synapse Analytics Workspace

Deployment name: Microsoft.Azure.SynapseAnalytics-20230331144...
Subscription: Azure Pass - Sponsorship
Resource group: b2_group3

Start time: 31-3-2023 14:50:17
Correlation ID: 7f32d5dd-8aa5-414b-8f22-9352022ecf29

Deployment details

Resource	Type	Status	Operation details
sailingdatabatchprocessing	Microsoft.Synapse/workspaces	Created	Operation details
datalakegen2sailingdata/default/sailingdatab	Microsoft.Storage/storageAccounts/blobSe...	Created	Operation details
datalakegen2sailingdata	Microsoft.Storage/storageAccounts	OK	Operation details

8. Tweaked the streaming output to appear in Synapse
9. Created an Apache Spark Pool (SparkPoolB2Gr3), this is costly but only when running.

New Apache Spark pool ...

Create a Synapse Analytics Apache Spark pool with your preferred configurations. Complete the Basics tab then go to Review + create to provision with smart defaults, or visit each tab to customize.

Apache Spark pool details

Name your Apache Spark pool and choose its initial settings.

Apache Spark pool name * SparkPoolB2Gr3

Node size family MemoryOptimized

Node size * Small (4 vCores / 32 GB)

Autoscale * Enabled Disabled

Number of nodes * 3

Dynamically allocate executors Enabled Disabled

Estimated price

Est. cost per hour
1.74 EUR
View pricing details

Contact an **Owner** of the storage account, and verify that the following role assignments have been made:

- Assign the workspace MSI to the **Storage Blob Data Contributor** role on the storage account
- Assign you and other users to the **Storage Blob Data Contributor** role on the storage account

Once those assignments are made, the following Spark features can be used: (1) Spark Library Management, (2) Read and Write data to SQL pool databases via the Spark SQL connector, and (3) Create Spark databases and tables

We decided to do the calculation and adding of the ‘Distance’ column in Power BI instead of in a notebook because we couldn’t get it to work.

Created a Notebook that fetches only the latest data for each boat

- The Notebook creates a table
- The Notebook runs only upon a trigger
- A trigger is set to run ‘something’ with a time interval of 5 minutes
- The trigger is connected to the Notebook using a Pipeline