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

A screenshot of a computer

Description automatically generated

Then configured the Input for the stream:

A screenshot of a web page

Description automatically generated

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:

A screenshot of a computer

Description automatically generated

Edited the Query to only select Latitude and Longitude columns:

A screenshot of a computer

Description automatically generated

PowerBI Streaming Settings:

A screenshot of a computer

Description automatically generated

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

A screenshot of a computer

Description automatically generated

8. Tweaked the streaming output to appear in Synapse

9. Created an Apache Spark Pool (SparkPoolB2Gr3), this is costly but only when running.

A screenshot of a computer

Description automatically generated

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