

04 | Activación Servicio de Azure Data Lake

<https://portal.azure.com/>



Creación del servicio

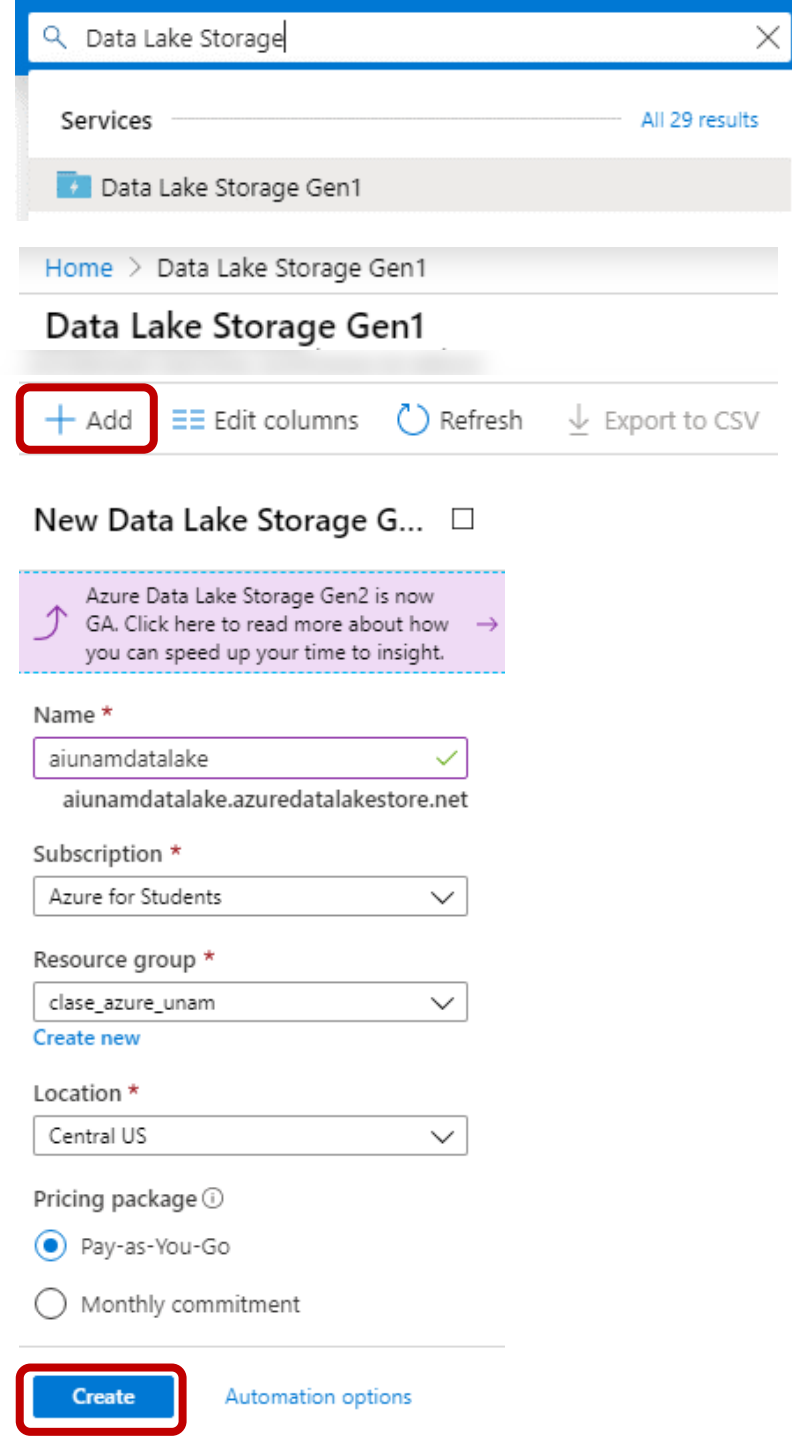
1. Buscar “Data Lake Storage” en el portal de Azure.

2. Seleccionar “add”

3. Llenar los campos con la información solicitada:

- Name
- Subscription
- Resource group
- Location

y seleccionar “create”.



Search: Data Lake Storage

Services [All 29 results](#)

Data Lake Storage Gen1

Home > Data Lake Storage Gen1

Data Lake Storage Gen1

[+ Add](#) [Edit columns](#) [Refresh](#) [Export to CSV](#)

New Data Lake Storage G... ☐

Azure Data Lake Storage Gen2 is now GA. Click here to read more about how you can speed up your time to insight. →

Name *
aiunamdatalake ✓
aiunamdatalake.azuredatastore.net

Subscription *
Azure for Students

Resource group *
clase_azure_unam
[Create new](#)

Location *
Central US

Pricing package ⓘ
☒ Pay-as-You-Go
☐ Monthly commitment

[Create](#) [Automation options](#)

4. Esperar a que finalice el despliegue del servicio y seleccionar el servicio recién creado.

[Home](#) > Data Lake Storage Gen1

Data Lake Storage Gen1

[+ Add](#) [≡ Edit columns](#) [↻ Refresh](#) [↓ Export to CSV](#) [🏷 Assign tags](#) [❤ Feedback](#) [↺ Leave preview](#)

Subscription == all Resource group == all Location == all + Add filter

Showing 1 to 1 of 1 records.

☐ Name ↑↓

Type ↑↓

Resource group ↑↓

Location ↑↓

Subscription ↑↓

☐

aiunamdatalake

Data Lake Storage Gen1

clase_azure_unam

Central US

Azure for Students

5. Seleccionar “Data Explorer”

[Home](#) > [Data Lake Storage Gen1](#) > aiunamdatalake

aiunamdatalake

Data Lake Storage Gen1

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Events

Data explorer

Delete

Azure Data Lake Storage Gen2 is now GA. Click here to read more

Essentials

aiunamdatalake

Pay-as-You-Go ⓘ

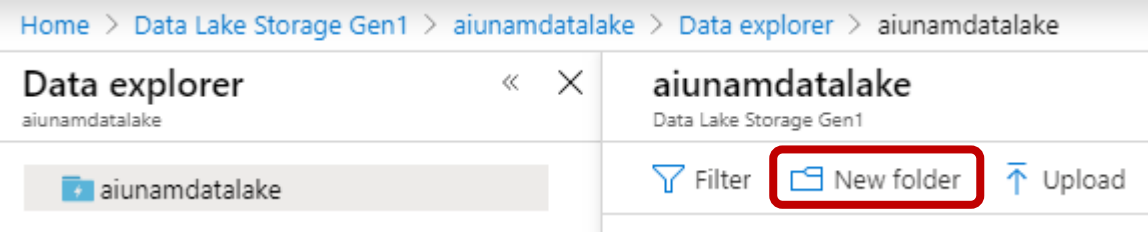
October 2019

[Pricing tier info](#)

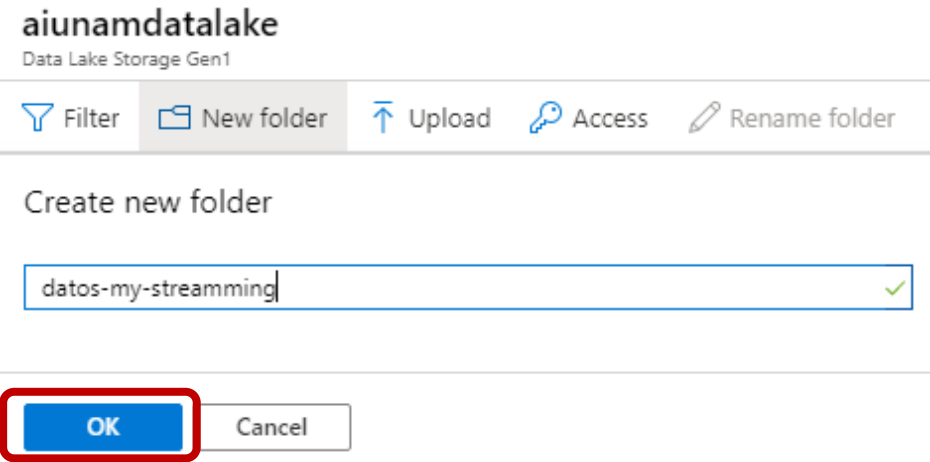
Estimated cost to date ⓘ

USD 0.00

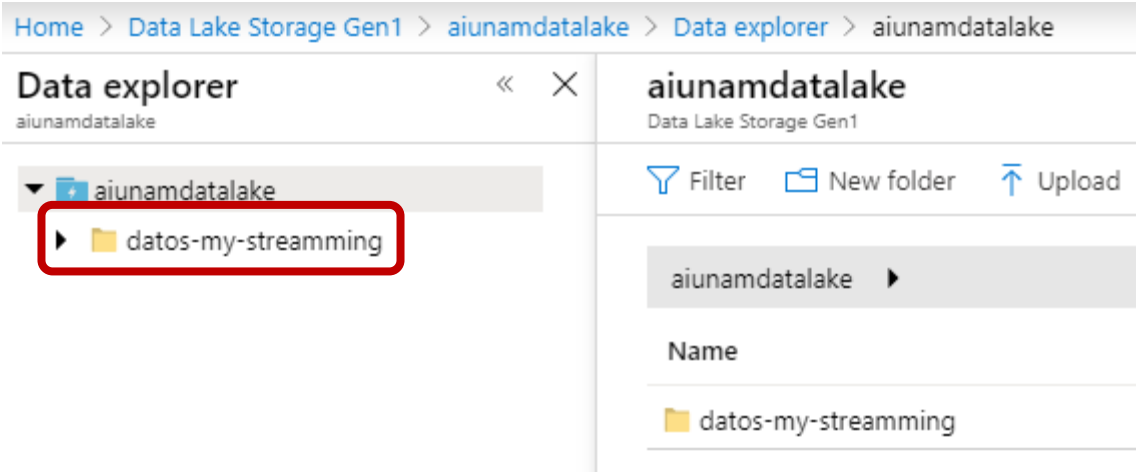
6. Seleccionar “New folder”.



7. Ingresar el nombre y seleccionar “Ok”



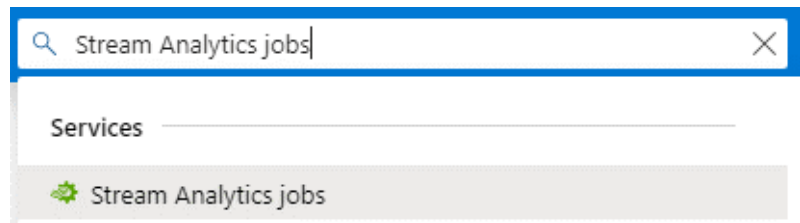
8. Acceder al nuevo folder.



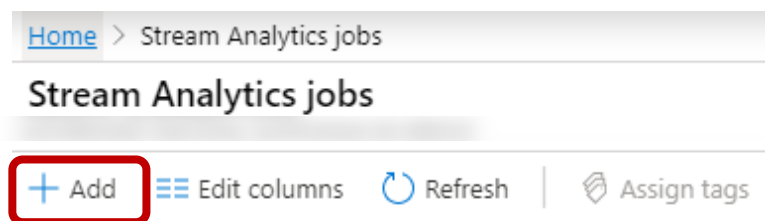


Creación de Analytics Job

1. Buscar “Stream analytics jobs” en el portal de Azure.



2. Seleccionar “add”



3. Llenar los campos con la información solicitada:

- Job name
- Subscription
- Resource group
- Location
- Hosting environment
- Streaming units

y seleccionar “create”.

A screenshot of the 'New Stream Analytics job' form in the Azure portal. The breadcrumb navigation shows 'Home > Stream Analytics jobs > New Stream Analytics job'. The form title is 'New Stream Analytics job'. The form contains several fields: 'Job name' (text input with 'IoT-job' and a green checkmark), 'Subscription' (dropdown menu with 'Azure for Students'), 'Resource group' (dropdown menu with 'clase_azure_unam'), 'Location' (dropdown menu with 'West US 2'), 'Hosting environment' (radio buttons for 'Cloud' and 'Edge', with 'Cloud' selected), and 'Streaming units (1 to 192)' (range selector with a value of 3). At the bottom of the form, there is a 'Create' button (highlighted with a red box) and a link for 'Automation options'.

4. Esperar a que finalice el despliegue del servicio y seleccionar el servicio recién creado.

[Home](#) > Stream Analytics jobs

Stream Analytics jobs

[+ Add](#)

[Edit columns](#)

[Refresh](#)

[Assign tags](#)

Subscriptions: Azure for Students


All resource groups

All locations

All tags

No grouping

1 items

<input type="checkbox"/>	Name ↑↓	Status	Created	Subscription ↑↓	Type ↑↓
<input type="checkbox"/>	 IoT-job	Created	2019-10-22T04:44:59.31...	Azure for Students	Stream Analytics job

5. Seleccionar “Inputs”

Home > Stream Analytics jobs > IoT-job

IoT-job
Stream Analytics job

Search (Ctrl+/)

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Locks

Job topology

Inputs

Functions

Query

Outputs

Start Stop

To start your job, you must first create an input.

Resource group (change)

Status

Location

Subscription (change)

Subscription ID

Inputs
0

Outputs
0

6. Seleccionar “Add stream input” y buscar el servicio “IoT hub”

Home > Stream Analytics jobs > IoT-job - Inputs

IoT-job - Inputs
Stream Analytics job

Search (Ctrl+/)

+ Add stream input + Add reference input

Overview

Activity log

Access control (IAM)

Event Hub

IoT Hub

Blob storage

7. Llenar los campos con la información solicitada:

- Input alias
- Subscription
- IoT Hub
- Endpoint
- Shared Access policy name
- Consumer group
- Event serialization format
- Encoding
- Event compression type

Y seleccionar “save”.

IoT Hub

New input

Input alias *

IoT-raspberry

☐ Provide IoT Hub settings manually

☒ Select IoT Hub from your subscriptions

Subscription

Azure for Students

IoT Hub ⓘ

my-streamming

Endpoint ⓘ

Messaging

Shared access policy name ⓘ

iothubowner

Shared access policy key ⓘ

.....

Consumer group ⓘ

\$Default

Event serialization format * ⓘ

JSON

You can implement a deserializer in C# that can read events in any format. You can try this out by [signing up for the preview program](#).

Encoding ⓘ

UTF-8

Event compression type ⓘ

None

Save

8. Deben poder visualizar el nuevo input generado. Ahora seleccionamos “Outputs”.

Home > IoT-job - Inputs

IoT-job - Inputs

Stream Analytics job

Search (Ctrl+ /)

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Locks

Job topology

Inputs

Functions

Query

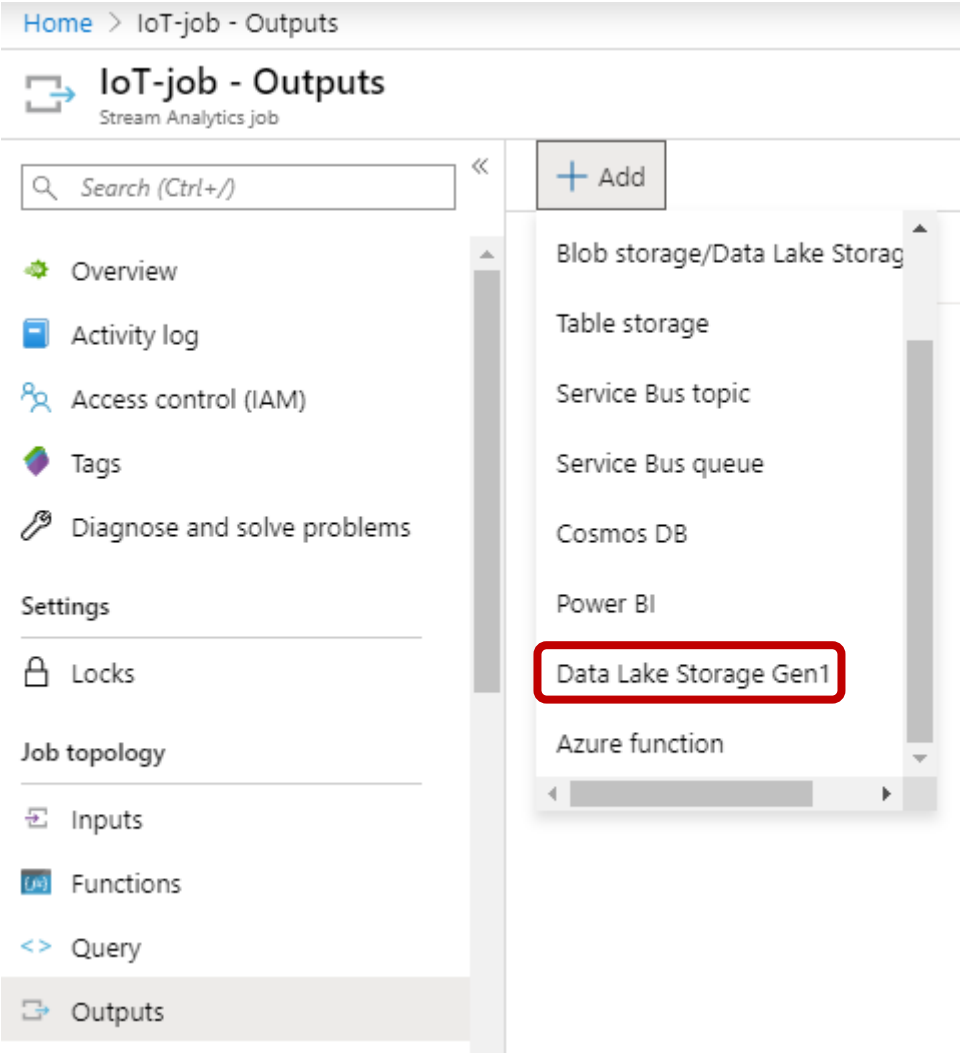
Outputs

+ Add stream input

+ Add reference input

Name	Source type
IoT-raspberry	Stream

9. Seleccionamos “Add” y buscamos el servicio “Data Lake Storage Gen 1”.



Data Lake Storage Gen1
New output

Output alias *
raspberry-data-lake ✓

☐ Provide Data Lake Storage Gen1 settings manually
☒ Select Data Lake Storage Gen1 from your subscriptions

Subscription
Azure for Students ▼

Account name
aiunamdatalake ▼

Path prefix pattern * ⓘ
datos-my-streamming/{date}/{time} ✓
Example: cluster1/logs/{date}/{time}

Date format
YYYY/MM/DD ▼

Time format
HH ▼

Event serialization format * ⓘ
JSON ▼

Encoding ⓘ
UTF-8 ▼

Format ⓘ
Line separated ▼

Authentication mode
Managed Identity ▼
✗ This resource now supports Managed Identity based authentication. Please enable it [here](#) for this job in order to use it.

Save

10. En la sección de “Authentication mode”, desplegar la lista y seleccionar “User token”

Authentication mode
Managed Identity ^
Managed Identity
User token

11. Aparecerá el siguiente recuadro, seleccionamos “Authorize”

Authentication mode
User token

Authorize connection
You'll need to authorize with Data Lake Storage Gen1 to configure your output settings.

Authorize

Don't have a Microsoft Azure Data Lake Storage Gen1 account yet?
[Sign up](#)

Note: You are granting this output permanent access to your Data Lake Storage Gen1 account. Should you need to revoke this access in the future you can do one of the following:

1. Change the user account password.
2. Delete this output.
3. Delete this job.

12. Aparecerá una pestaña de login y después deben generarse la siguiente información:

Authentication mode
User token

Currently authorized as

Authorize connection
You'll need to authorize with Data Lake Storage Gen1 to configure your output settings.

Authorize

13. Llenar los campos con la información solicitada:

Data Lake Storage Gen1
New output

Output alias *
raspberry-data-lake

☐ Provide Data Lake Storage Gen1 settings manually
☒ Select Data Lake Storage Gen1 from your subscriptions

Subscription
Azure for Students

Account name
aiunamdatalake

Path prefix pattern * ⓘ
datos-my-streamming/{date}/{time} ✓
Example: cluster1/logs/{date}/{time}

Date format
YYYY/MM/DD

Time format
HH

Event serialization format * ⓘ
JSON

Encoding ⓘ
UTF-8

Format ⓘ
Line separated

Authentication mode
User token


Save

Ya no deben observarse el letrero previo.

- Output alias
- Subscription
- Account name
- Path prefix pattern
- Date format
- Time format
- Event serialization format
- Encoding
- Format
- Authentication mode

14. Confirmamos la creación del servicio “Output” y seleccionamos “Overview”.

Home > IoT-job - Outputs

IoT-job - Outputs
Stream Analytics job

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Locks

Job topology

Inputs

Functions

Query


Outputs

+ Add

Name	Sink
raspberry-data-lake	Data Lake Storage Gen1

15. Seleccionamos “Edit query” para definir la consulta a que realizaremos al streaming.

Home > IoT-job



IoT-job

Stream Analytics job

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Settings

Locks

Job topology

Inputs

Functions

Query

Outputs

▶ Start

□ Stop

🗑 Delete

Resource group (change)

clase_azure_unam

Status

Created

Location

West US 2

Subscription (change)

Azure for Students

Subscription ID

Send feedback

UserVoice

Created

Monday, October 21, 2019, 11:44:59 PM

Started

-

Output watermark

-

Hosting environment

Cloud

Inputs

1

IoT-raspberry

IoT Hub

Outputs

Query

1 SELECT

2 *

3 INTO

4 [YourOutputAlias]

5 FROM

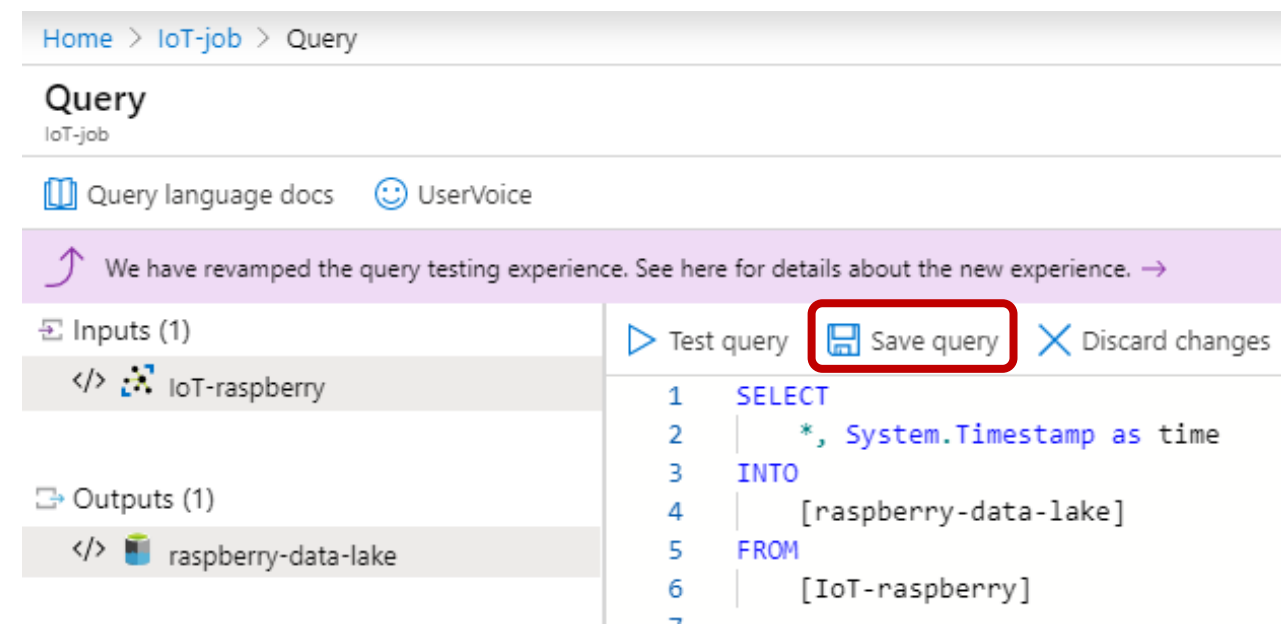
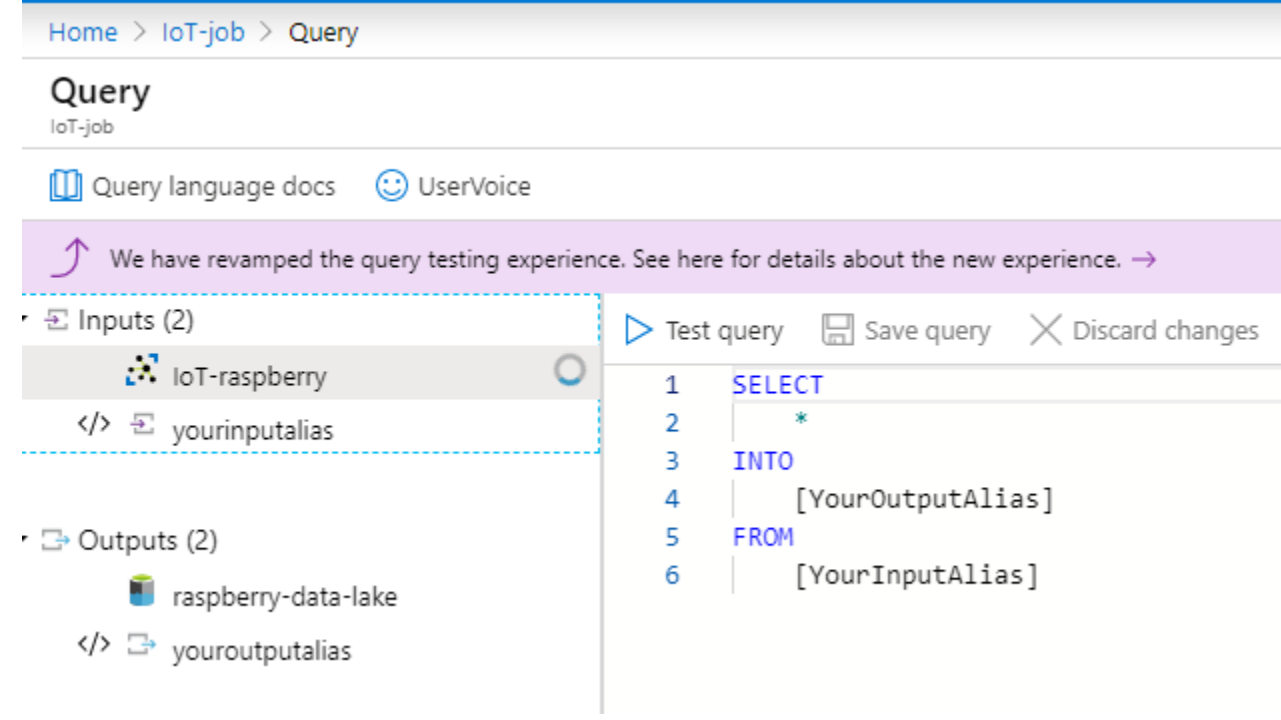
6 [YourInputAlias]

Edit query

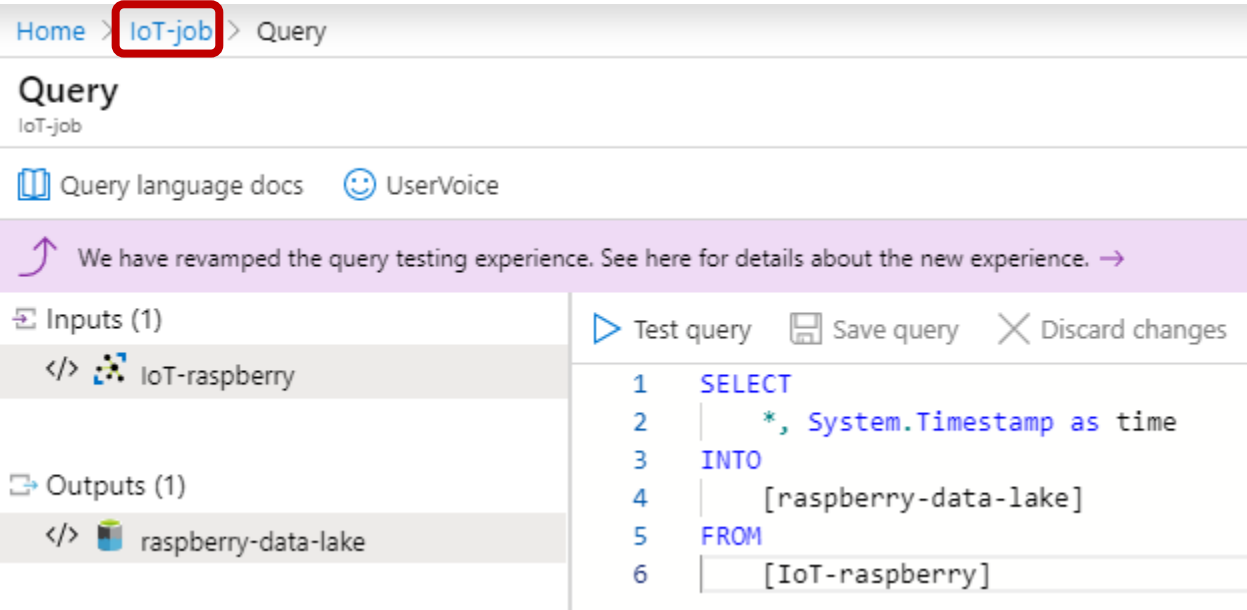
16. En esta sección se crea la consulta (query) que utilizaremos para extraer información del IoT Hub y que se guardará en el data lake.

- **SELECT:** Hace referencia a las columnas y/o transformaciones que vamos a realizar. En este caso particular vamos a traer todas las columnas de IoT Hub y además vamos a incorporar un timestamp el cual se llamará “time”.
- **INTO:** Hace referencia al “output”. En nuestro caso particular es el nombre definido previamente como “raspberry-data-lake”.
- **FROM:** Hace referencia al “input”. En nuestro caso particular es el nombre definido previamente como “IoT-raspberry”.

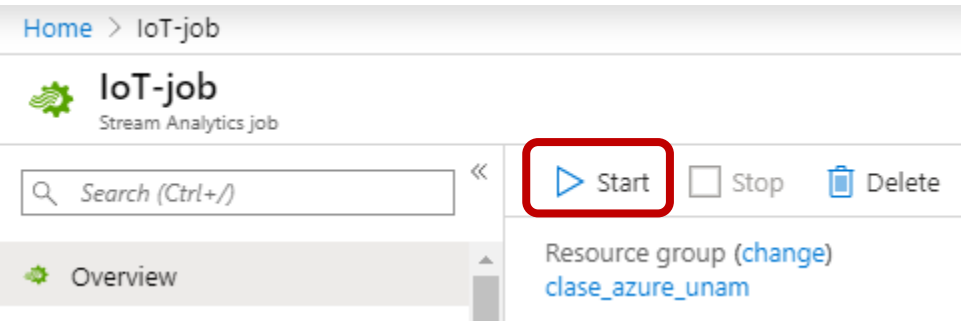
Una vez tenemos la consulta, le damos “save query”.



17. Una vez está guardada nuestra consulta, regresamos a la sección anterior seleccionando el nombre del servicio. En este caso “IoT-job”.

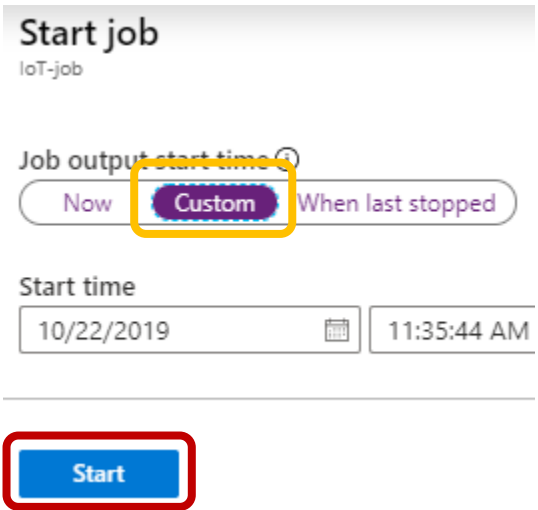
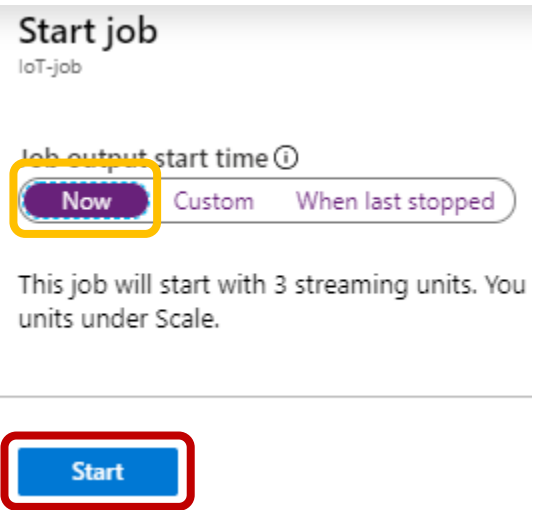


18. Seleccionamos “Start”.

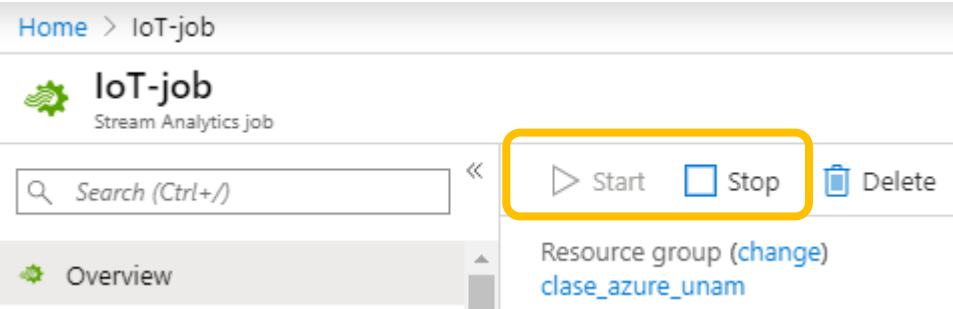


18. Seleccionamos si queremos que el “job” comience con datos desde:

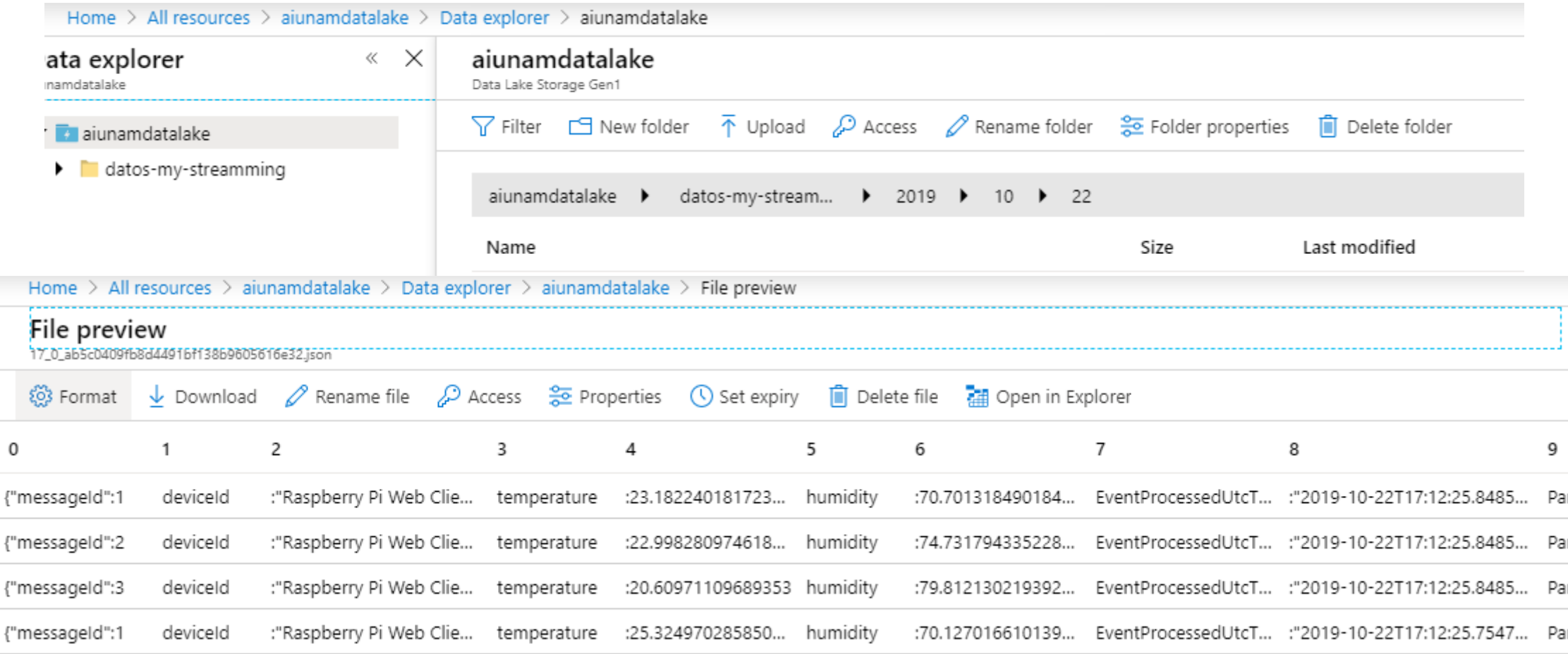
- **Now:** Hace referencia a “este momento”. Solamente guardará los datos en el data lake que vayan ingresando al IoT Hub desde el momento que den “start”.
- **Custom:** Es posible seleccionar una fecha y hora previa de tal forma que se recuperé información del IoT Hub desde que inició el dispositivo (siempre y cuando este dentro del tiempo que almacena el IoT Hub los datos)



19. Observamos que el proceso está ejecutándose:



20. Nos vamos a la carpeta del Data Lake y observamos que se crearon carpetas que contienen archivos JSON.



The image features a dark gray background with three overlapping circles in two shades of blue. A horizontal white band runs across the center, containing the word "Actividades" in a dark blue, sans-serif font.

Actividades

- Modificar la consulta para que realicé alguna transformación o aplique algún filtro a los datos
- Repetir el proceso para otros dispositivos IoT hub.
- Detener el Streaming Job, esperar 10 minutos y reiniciarlo de forma que recuperé la información generada en esos 10 minutos previos.



Microsoft

©2014 Microsoft Corporation. All rights reserved. Microsoft, Windows, Office, Azure, System Center, Dynamics and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries. The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.