04 | Activación Servicio de Azure Data Lake

Microsoft

https://portal.azure.com/

Creación del servicio

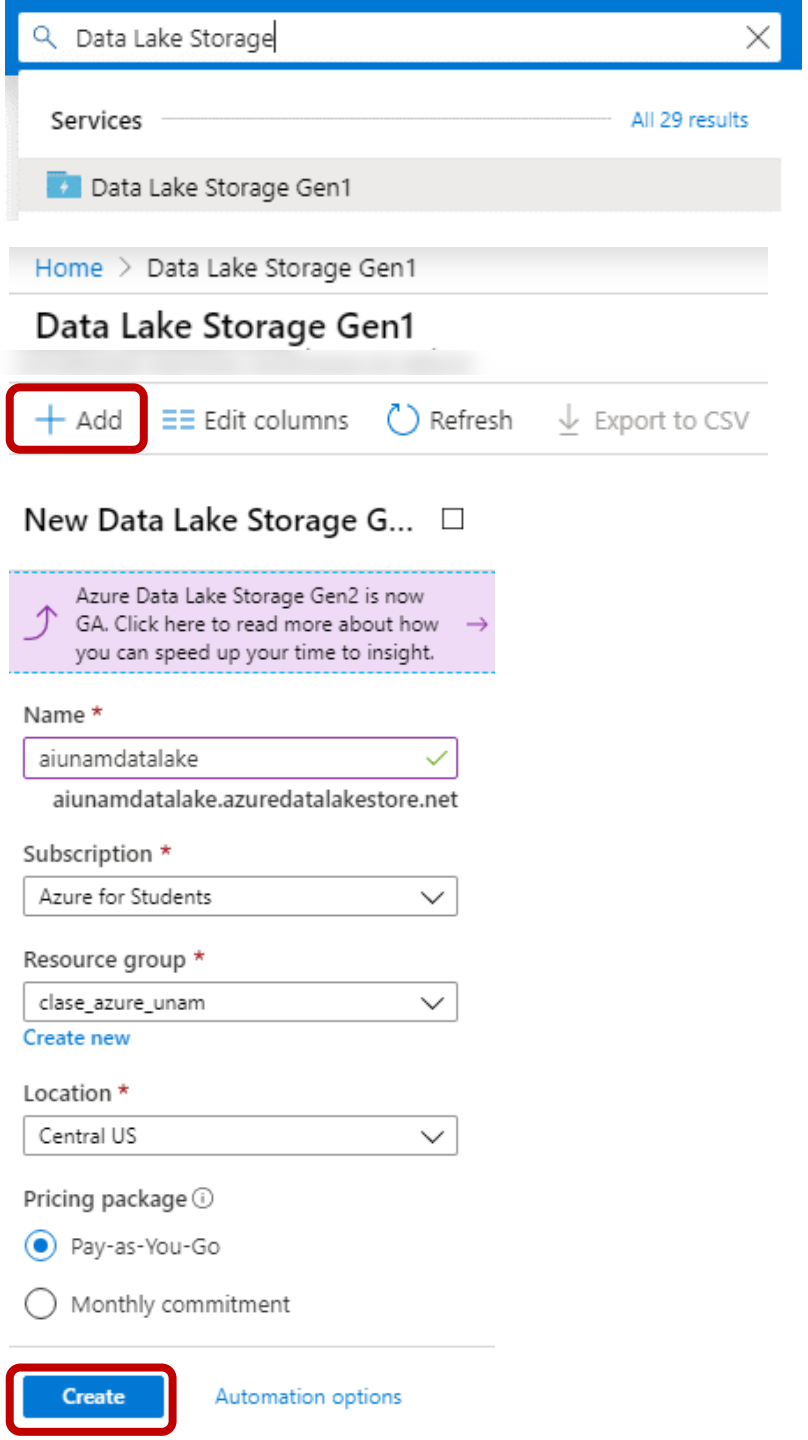
1. Buscar “Data Lake Storage” en el portal de Azure.

2. Seleccionar “add”

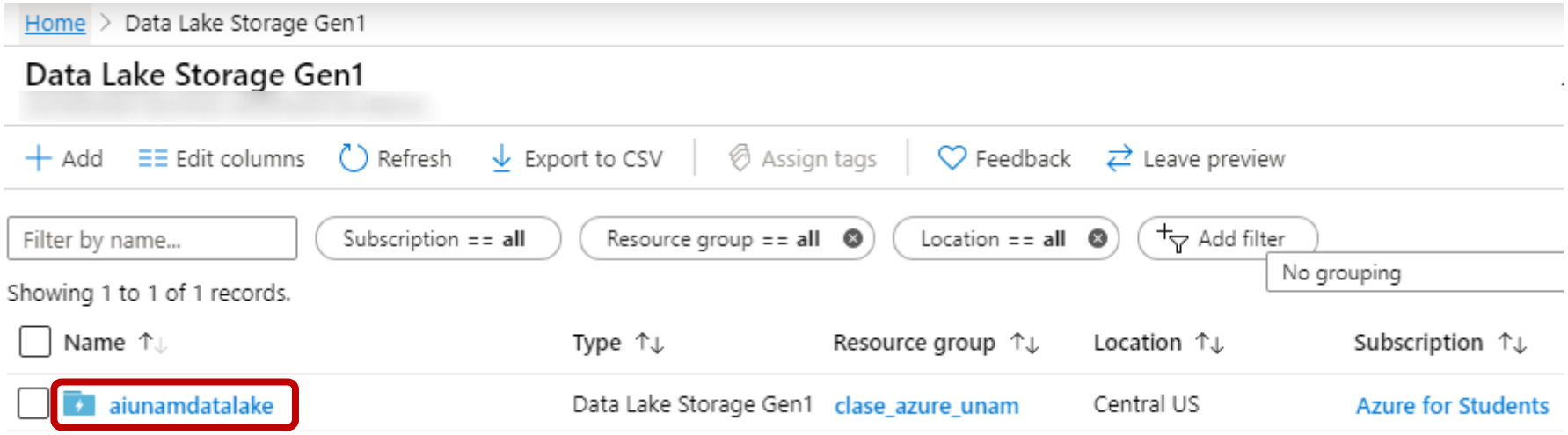
3. Llenar los campos con la información solicitada:

- Name
- Subscription
- Resource group
- Location

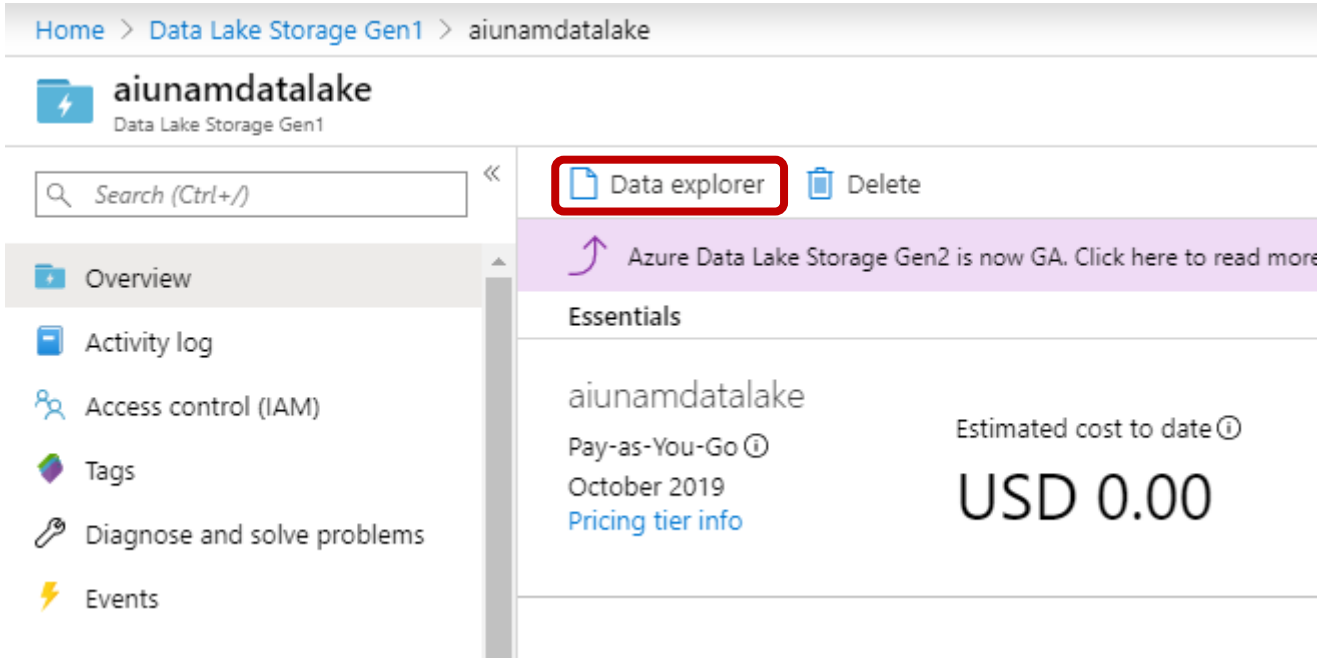
y seleccionar “create”.



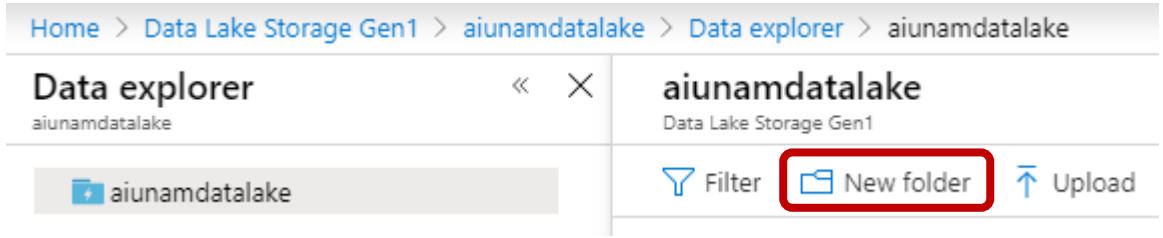
4. Esperar a que finalice el despliegue del servicio y seleccionar el servicio recién creado.



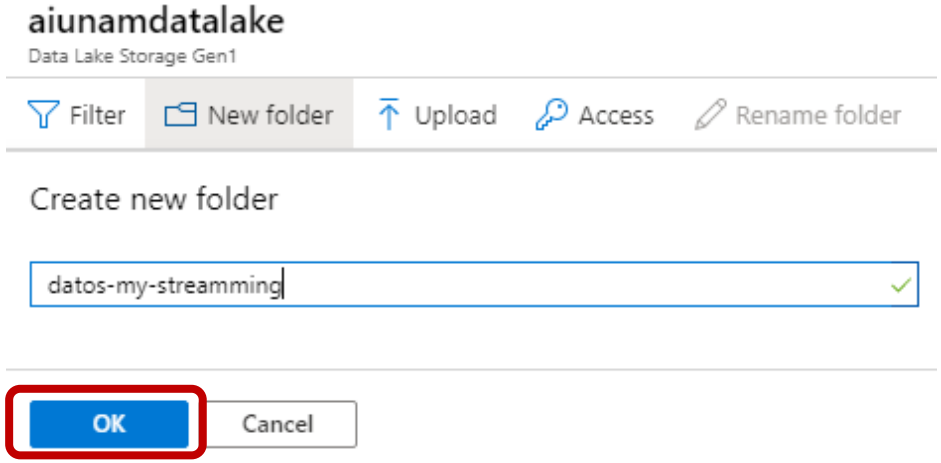
5. Seleccionar “Data Explorer”



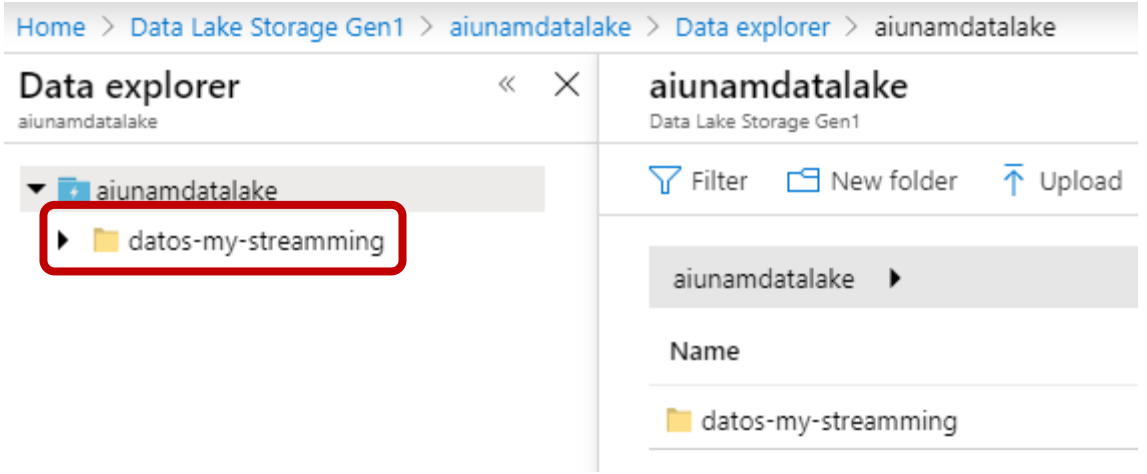
6. Seleccionar “New folder”.



7. Ingresar el nombre y seleccionar “Ok”

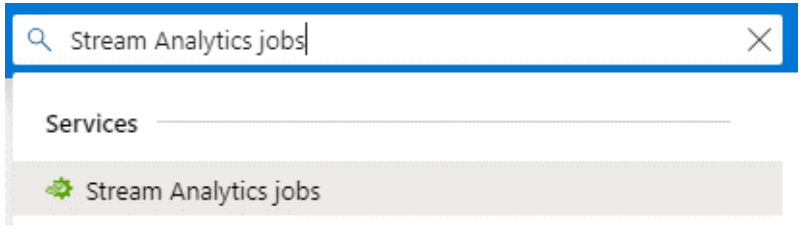


8. Acceder al nuevo folder.

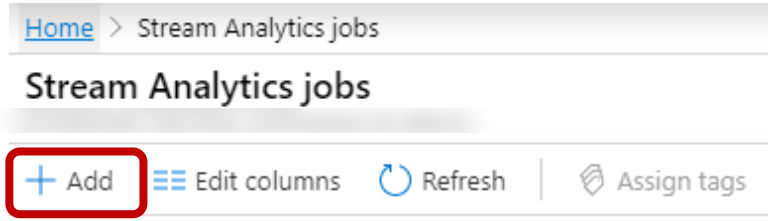


Creación de Analytics Job

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



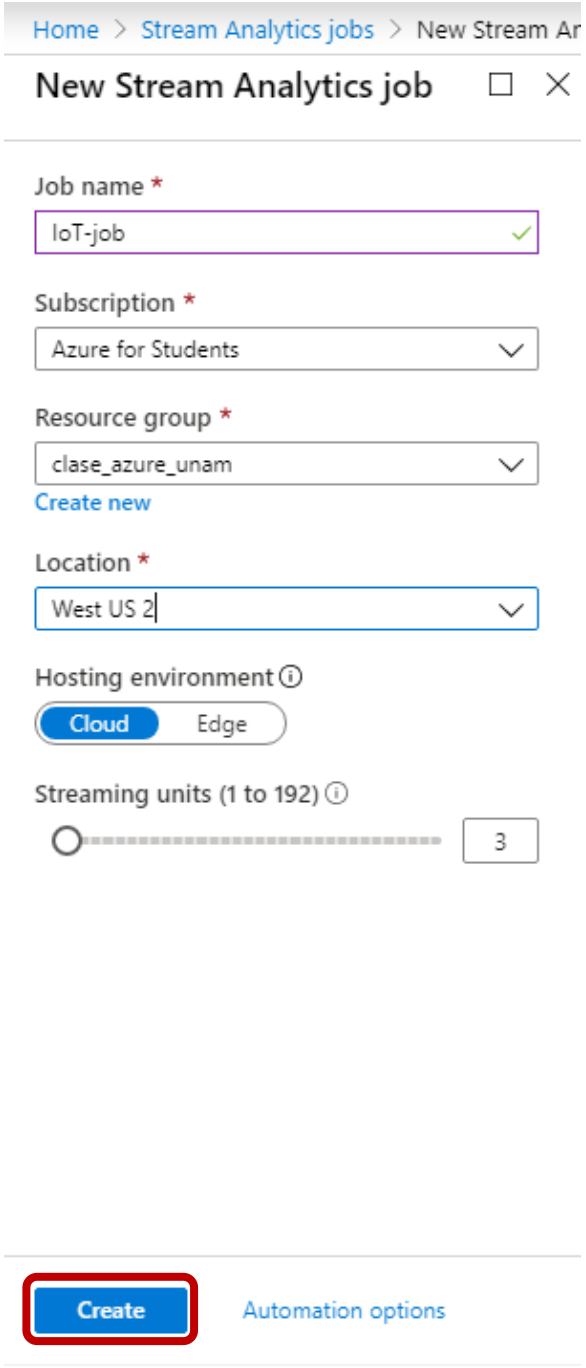
2. Seleccionar “add”



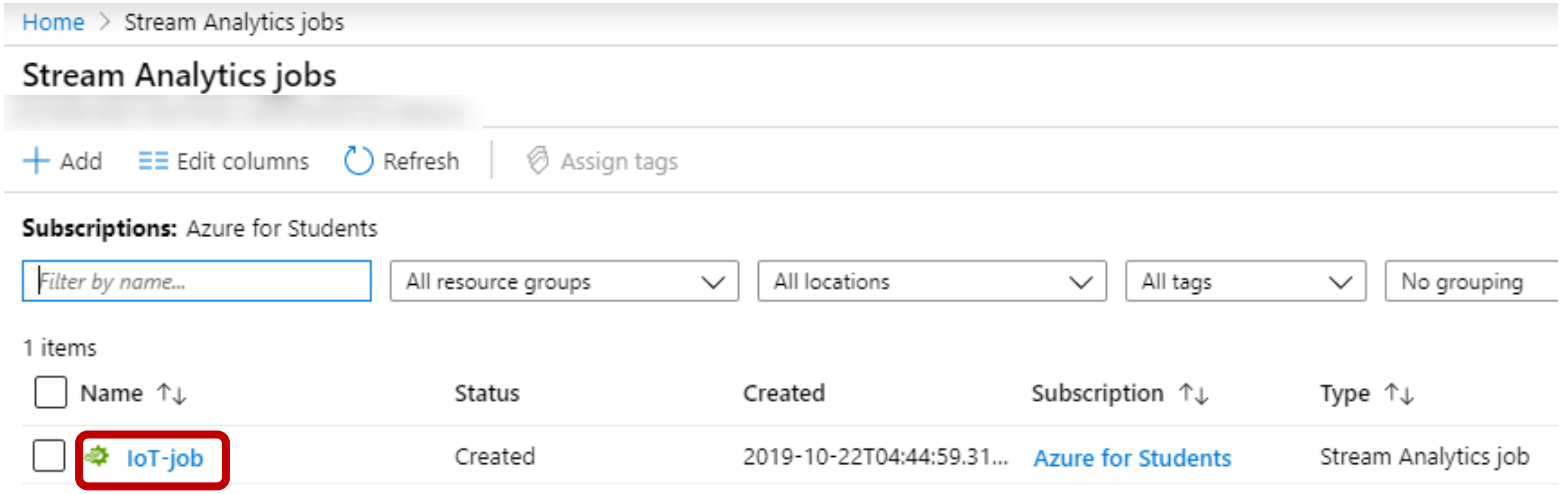
3. Llenar los campos con la información solicitada:

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

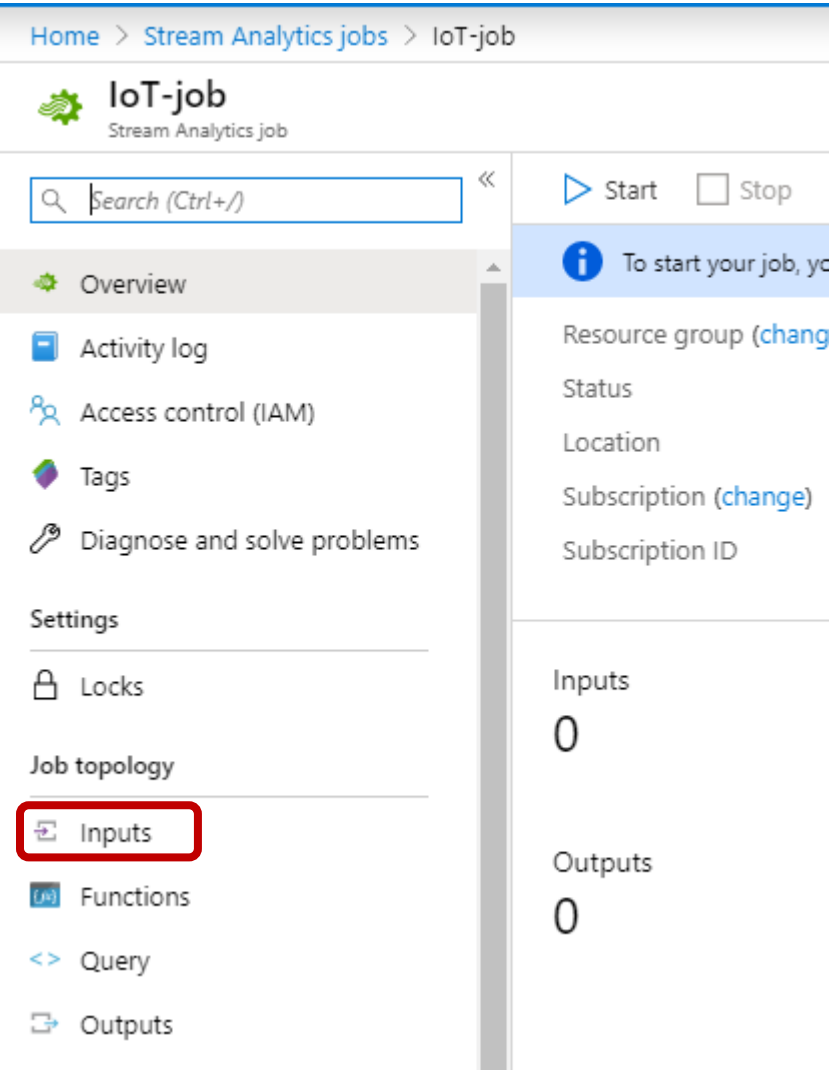
y seleccionar “create”.



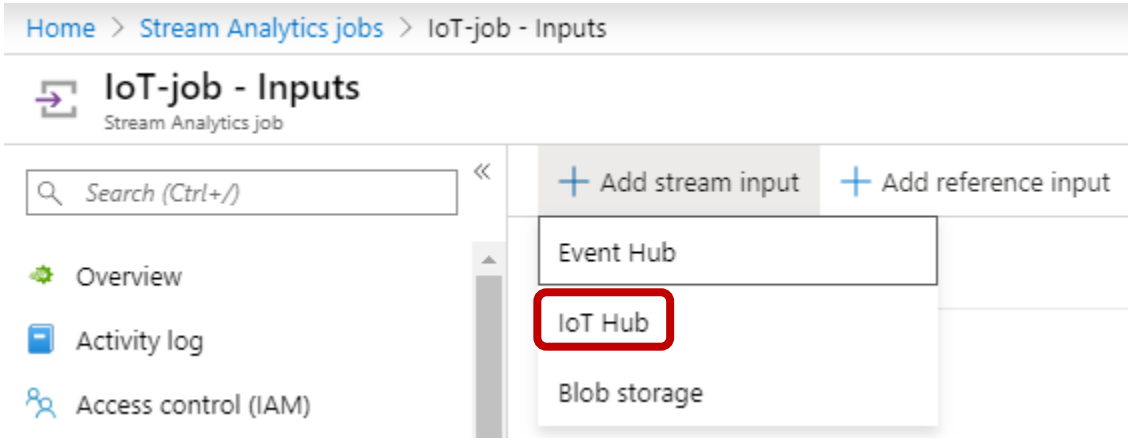
4. Esperar a que finalice el despliegue del servicio y seleccionar el servicio recién creado.



5. Seleccionar “Inputs”



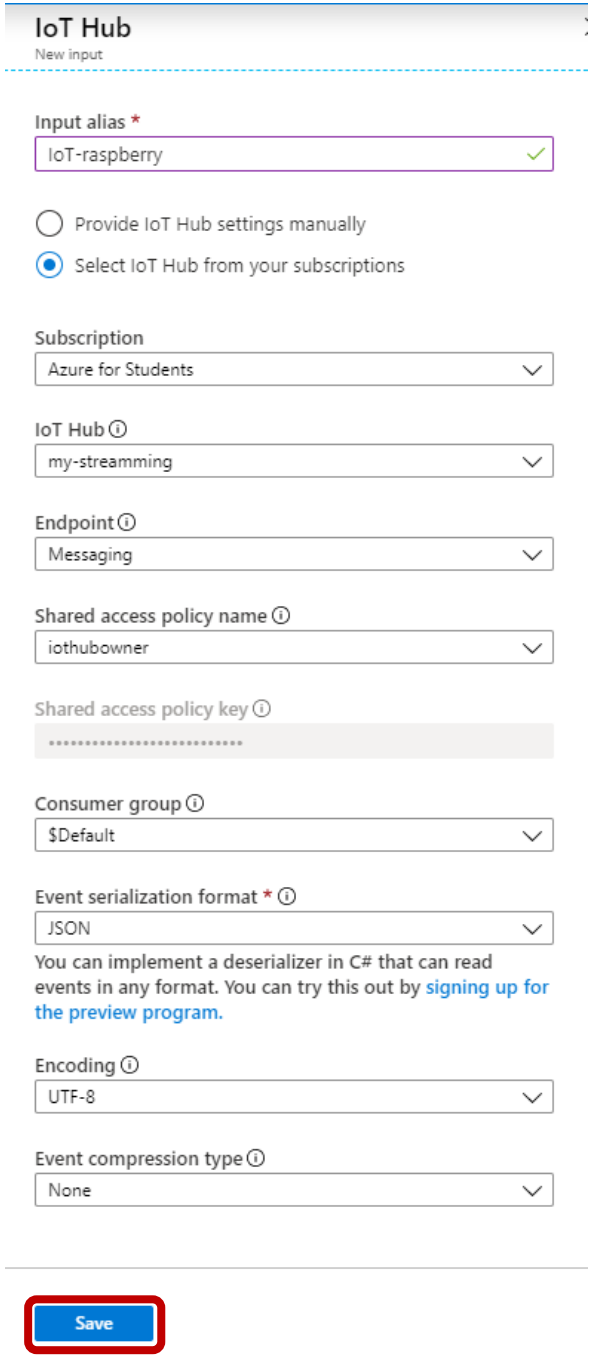
6. Seleccionar “Add stream input” y buscar el servicio “IoT hub”



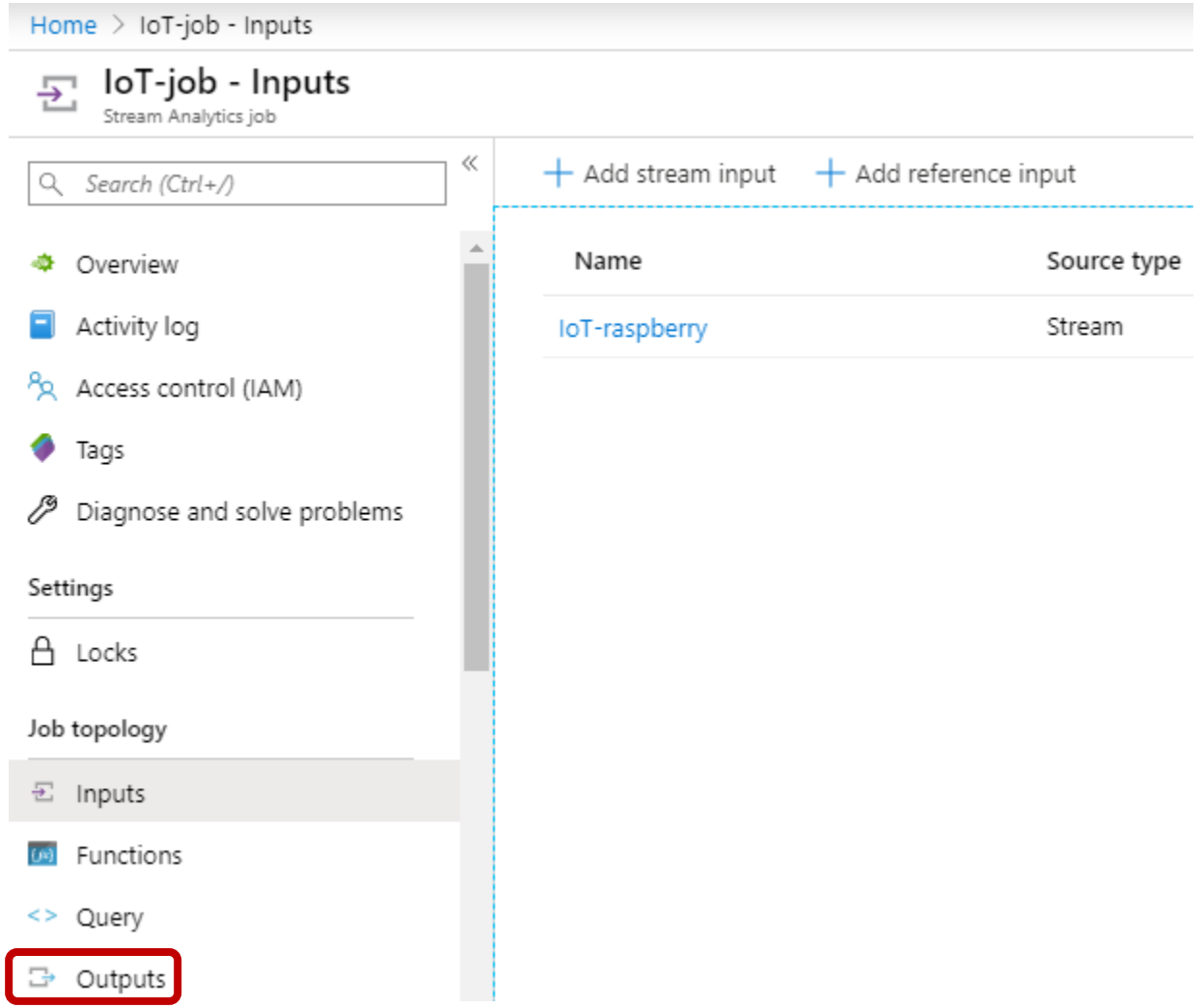
7. Llenar los campos con la información solicitada:

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

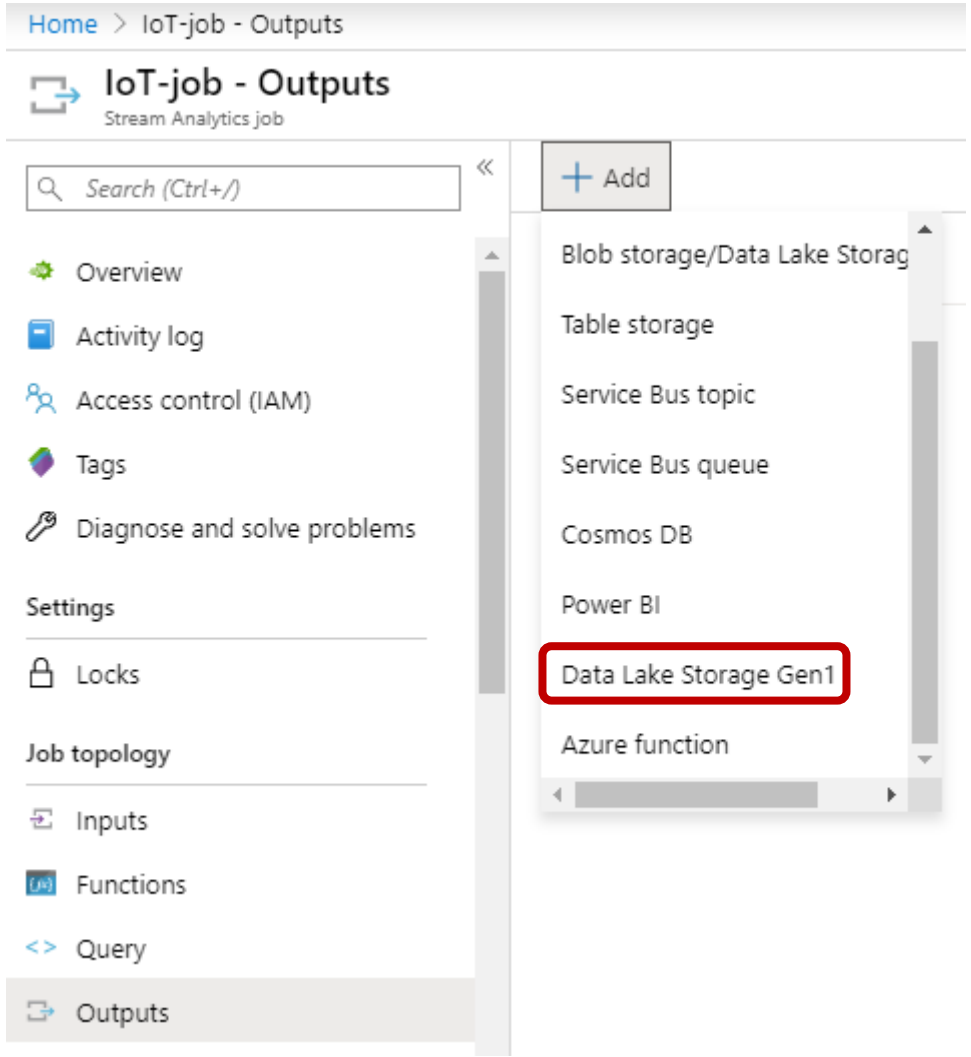
Y seleccionar “save”.

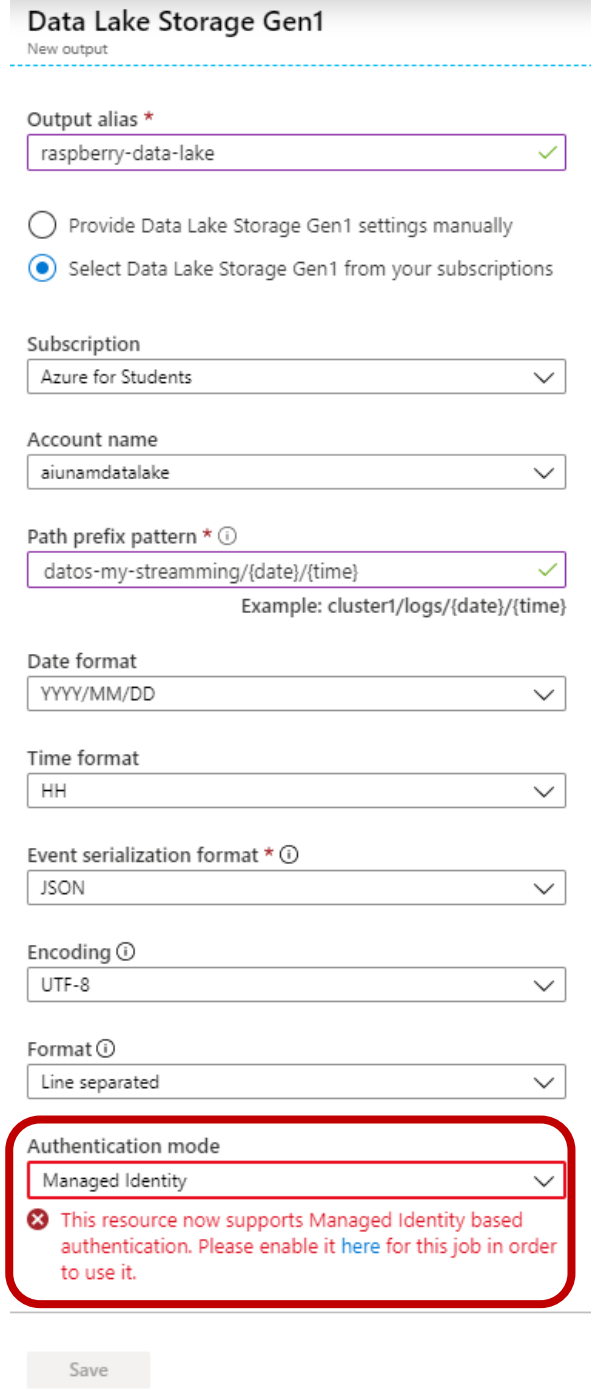


8. Deben poder visualizar el nuevo input generado. Ahora seleccionamos “Outputs”.

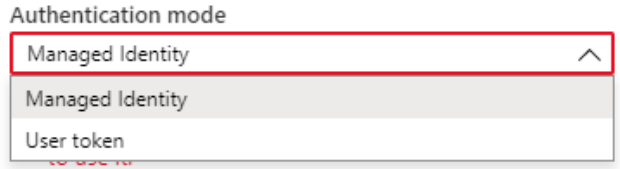


9. Seleccionamos “Add” y buscamos el servicio “Data Lake Storage Gen 1”.

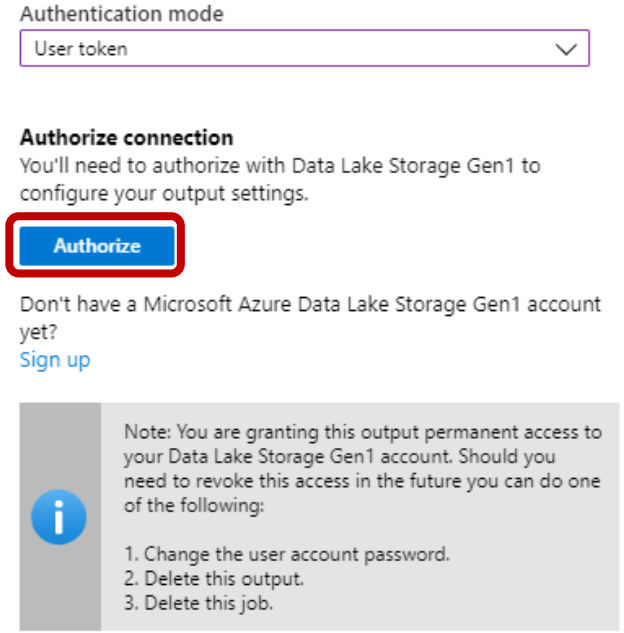




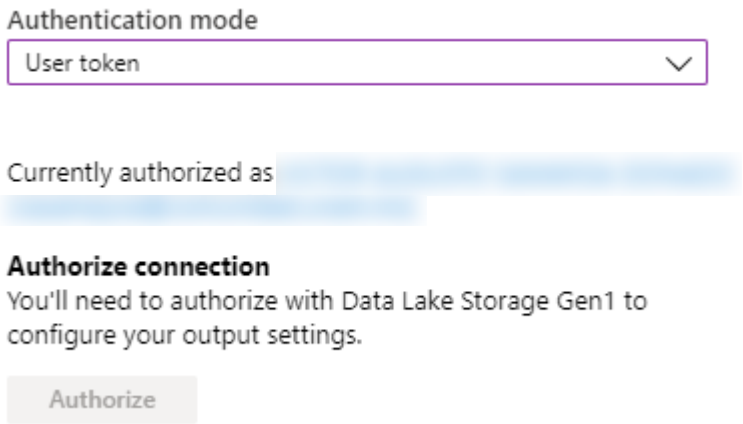
10. En la sección de “Authentication mode”, desplegar la lista y seleccionar “User token”



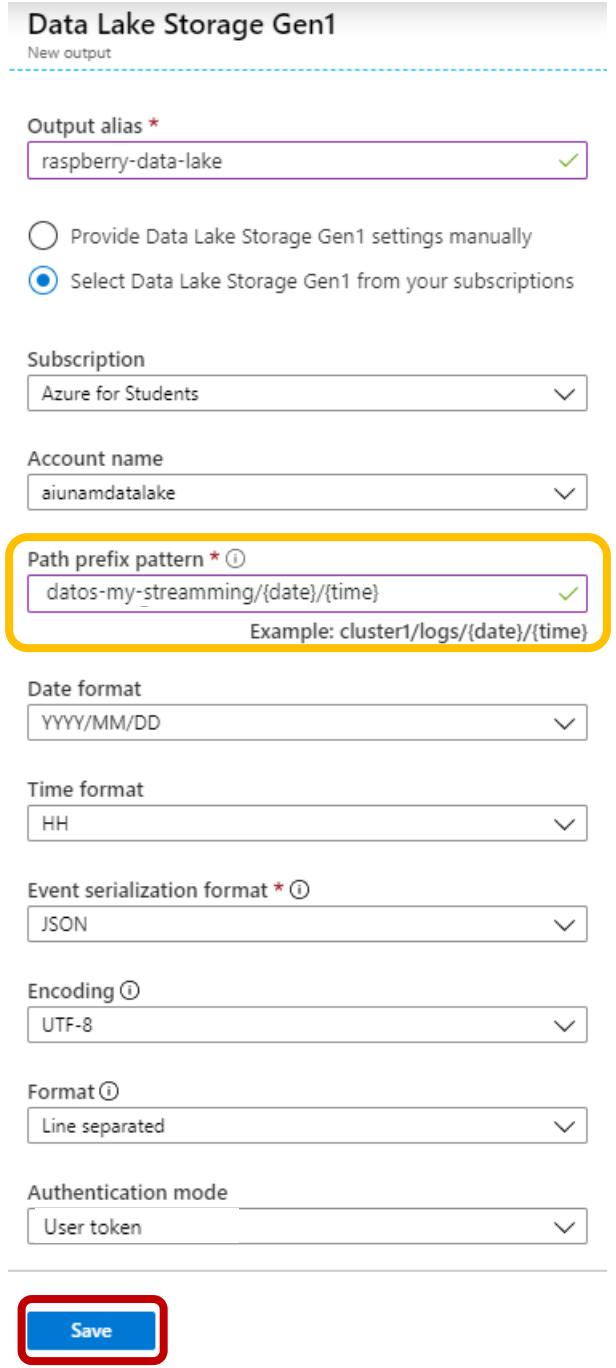
11. Aparecerá el siguiente recuadro, seleccionamos “Authorize”



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



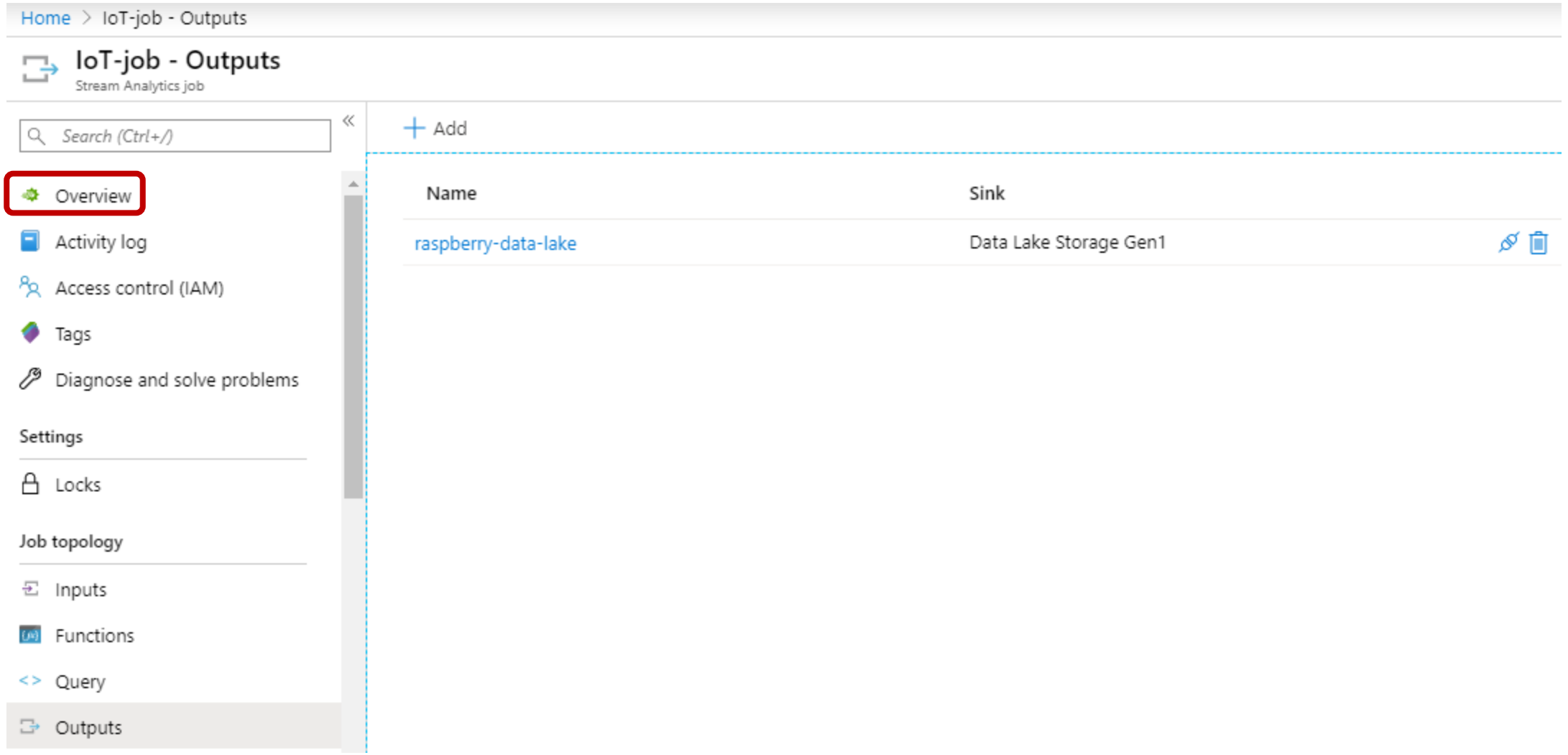
13. Llenar los campos con la información solicitada:



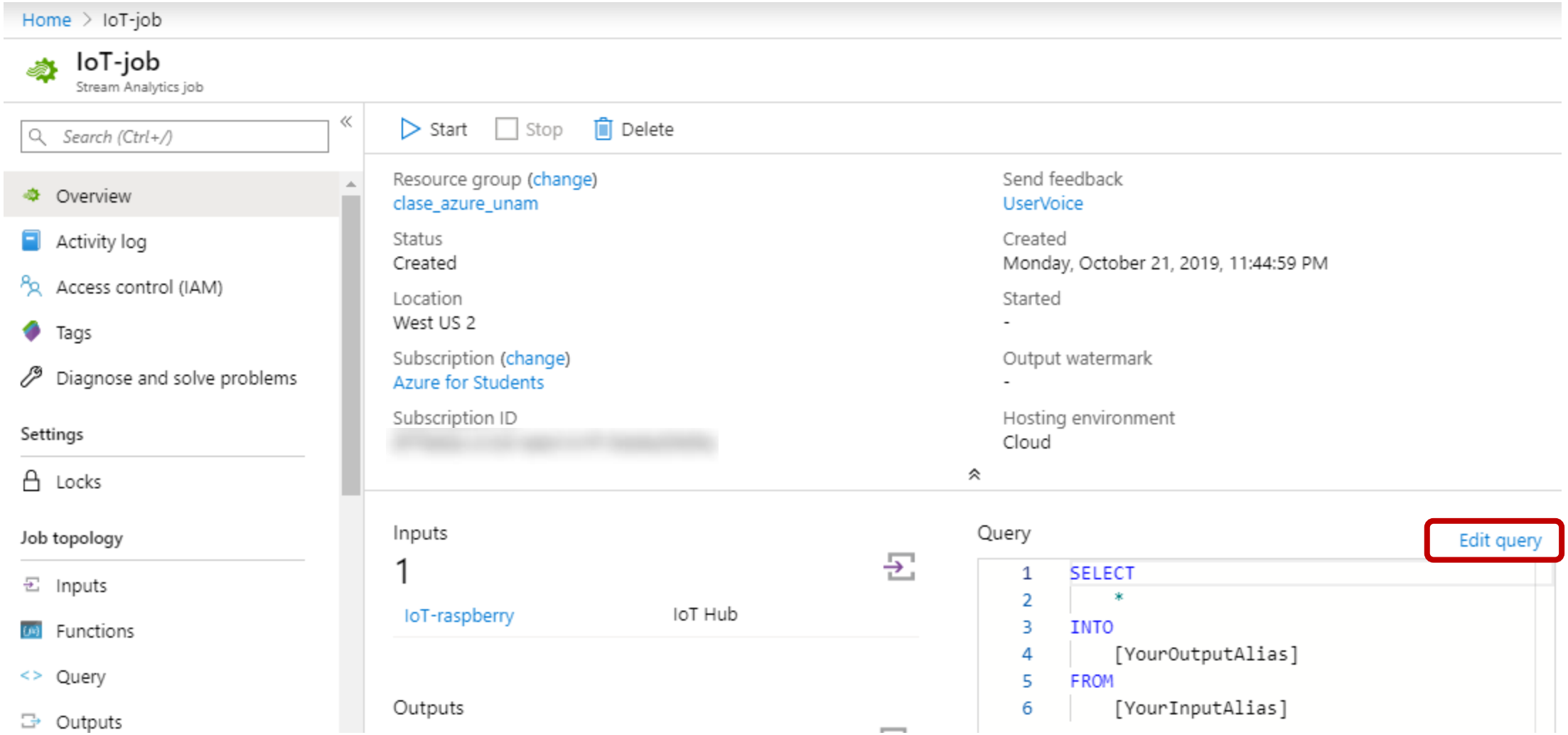
Ya no deben observarse el letrero previo.

- 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”.



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



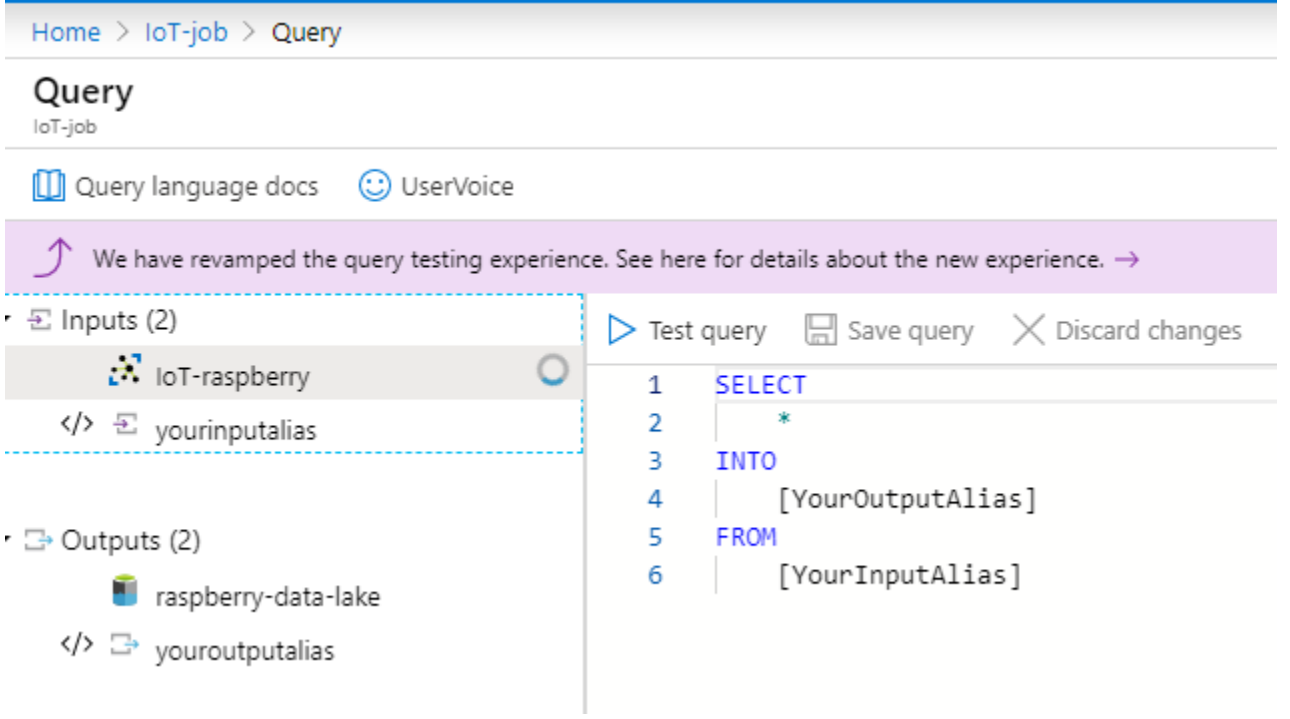
16. En esta sección se crea la consulta (query) que utilizaremos para extraer información del IoT Hub y que se guardará en el data lake.

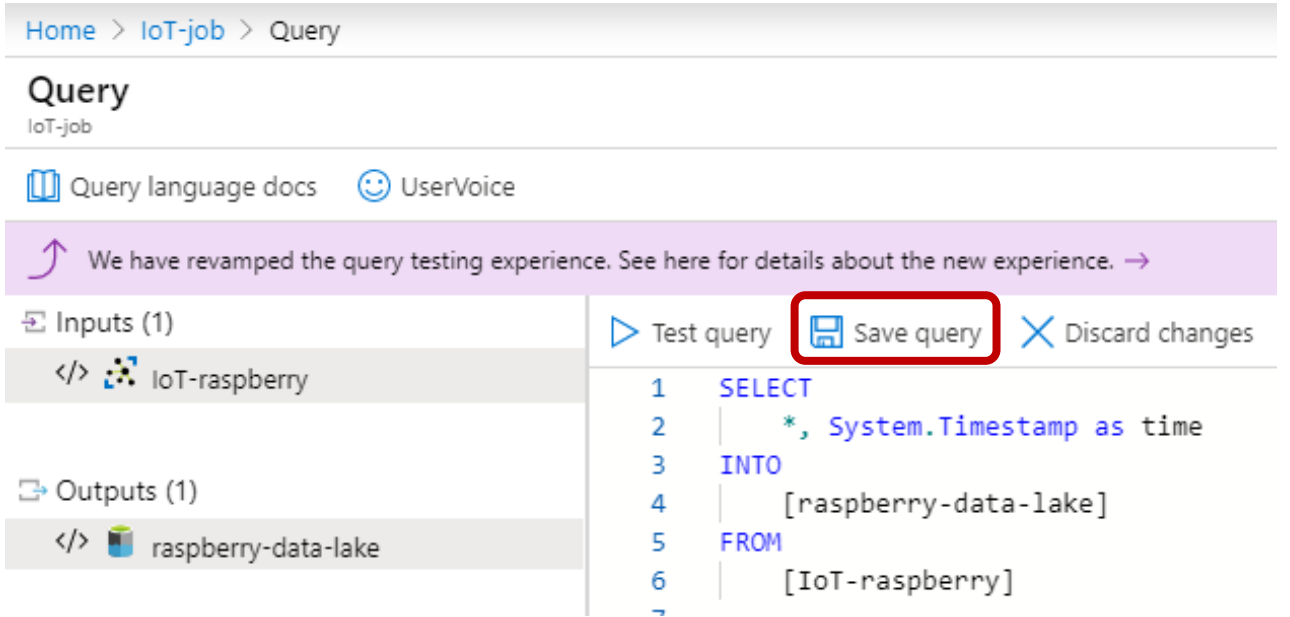
- **SELECT:** Hace referencia a las columnas y/o transformaciones que vamos a realizar. En este caso particular vamos a traer todas las columnas de IoT Hub y además vamos a incorporar un timestamp el cual se llamará “time”.

- **INTO:** Hace referencia al “output”. En nuestro caso particular es el nombre definido previamente como “raspberry-data-lake”.

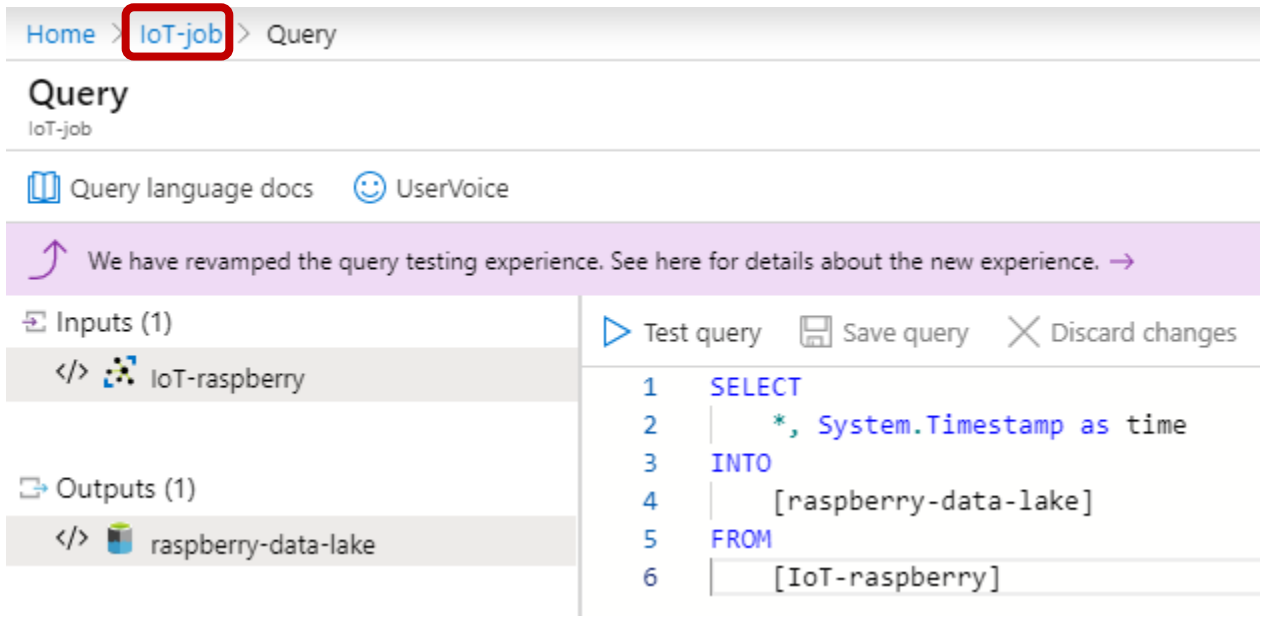
- **FROM:** Hace referencia al “input”. En nuestro caso particular es el nombre definido previamente como “IoT-raspberry”.

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

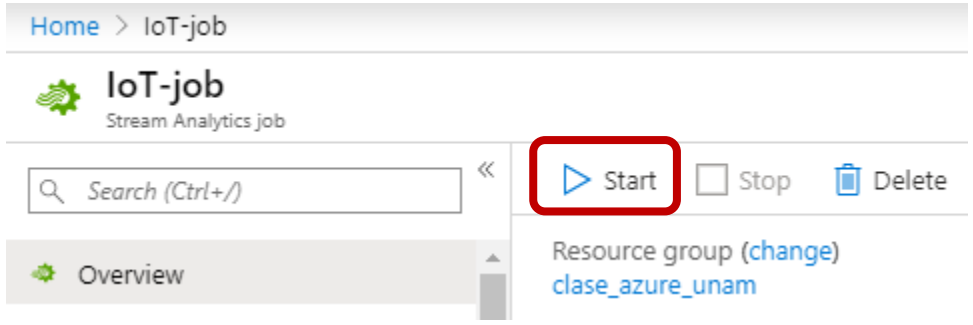




17. Una vez está guardada nuestra consulta, regresamos a la sección anterior seleccionando el nombre del servicio. En este caso “IoT-job”.



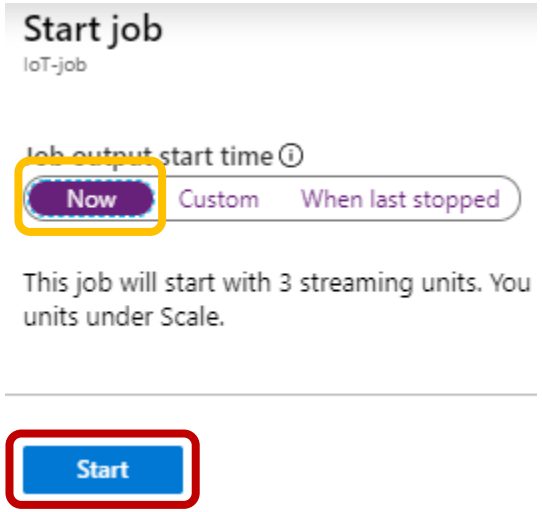
18. Seleccionamos “Start”.

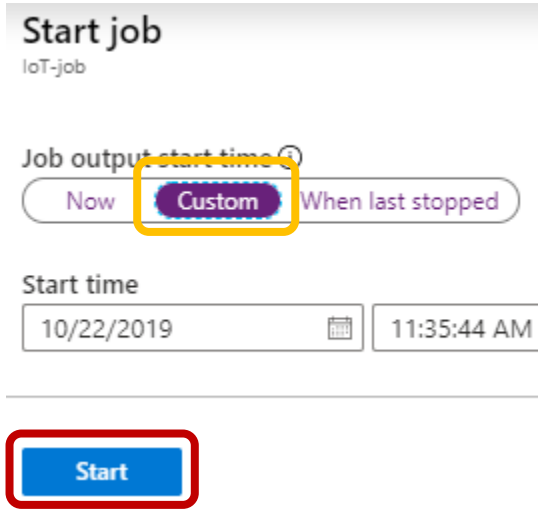


18. Seleccionamos si queremos que el “job” comience con datos desde:

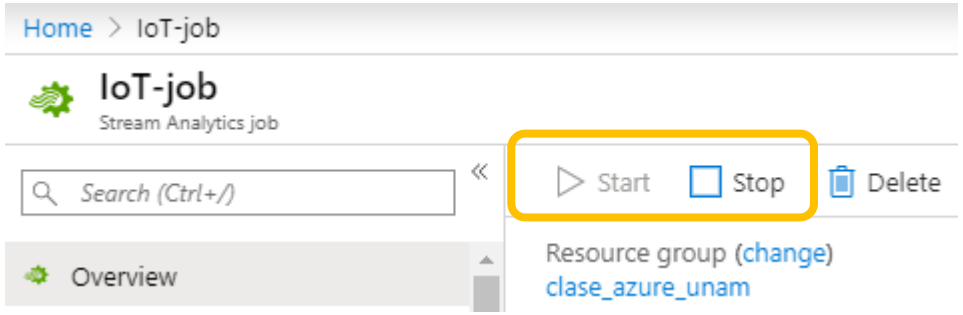
- **Now:** Hace referencia a “este momento”. Solamente guardará los datos en el data lake que vayan ingresando al IoT Hub desde el momento que den “start”.

- **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)

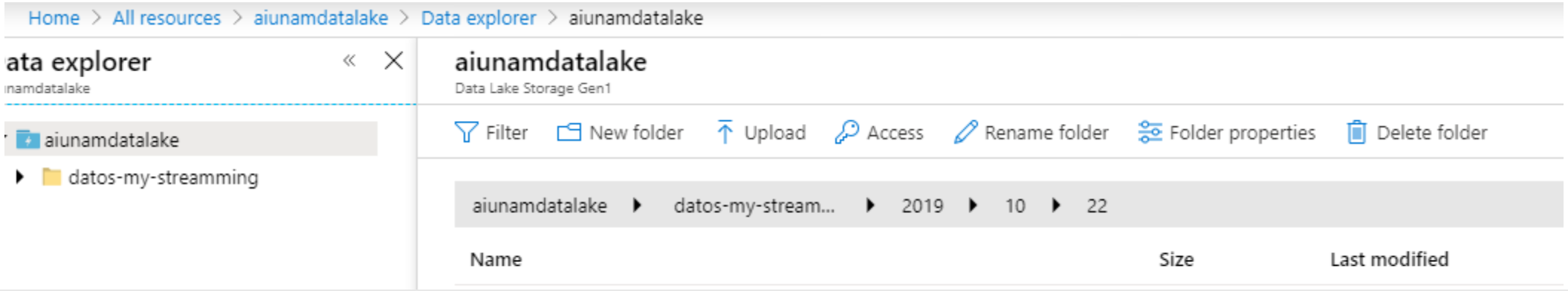


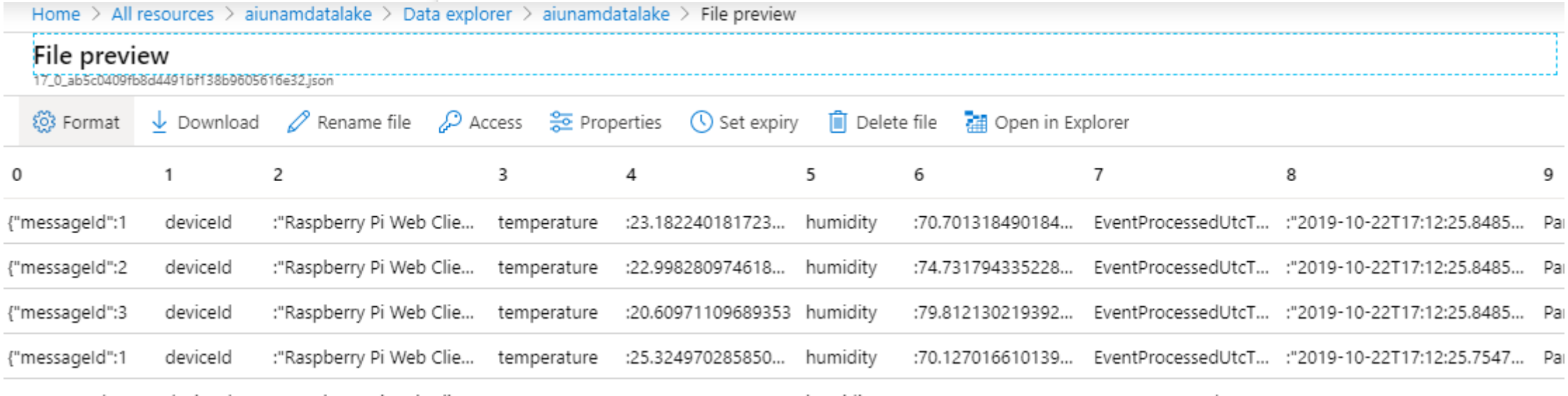


19. Observamos que el proceso está ejecutándose:



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





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 Streamming Job, esperar 10 minutos y reiniciarlo de forma que recuperé la información generada en esos 10 minutos previos.

Microsoft