

Informe:

Práctica 5: Ingesta, almacenamiento y Serveless

Jordi Blasco Lozano
Infraestructuras y Servicios Cloud
Universidad de Alicante

18 de noviembre de 2025

Resumen

Índice

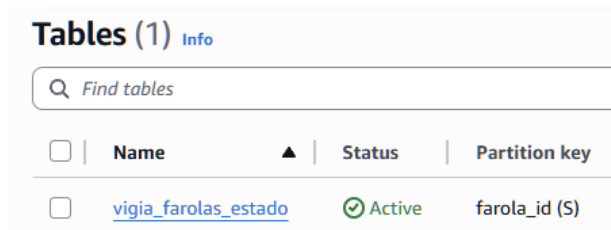
1. Flujo IoT: Ingesta de Telemetría y Almacenamiento Dual	2
1.1. Preparación de Destinos y Configuración IoT	2
1.1.1. Creación de la Tabla DynamoDB	2
1.1.2. Configuración de AWS IoT Core	2
1.2. Creación de la Regla IoT	3
1.2.1. Consulta SQL	3
1.2.2. Acción S3 bucket	3
1.2.3. Accion DynamoDB	4
1.3. Simulación de ingesta con Node-RED	4
1.3.1. Pruebas	5
2. Flujo de Contexto: Web Scraping con Lambda Serverless (RA2)	7
2.1. Creación de la Función Lambda	7
2.2. Orquestación Serverless (EventBridge)	7
3. Flujo Masivo: Transformación y Análisis Serverless	9

1 Flujo IoT: Ingesta de Telemetría y Almacenamiento Dual

1.1 Preparación de Destinos y Configuración IoT

1.1.1 Creación de la Tabla DynamoDB

Se ha creado una tabla en DynamoDB llamada `vigia_farolas_estado` para almacenar el último estado reportado por cada farola junto con métricas del consumo y el momento de medir estas métricas. La clave de la tabla será la id de cada farola. Solamente indicando como queremos la clave de partición nuestra tabla será funcional.



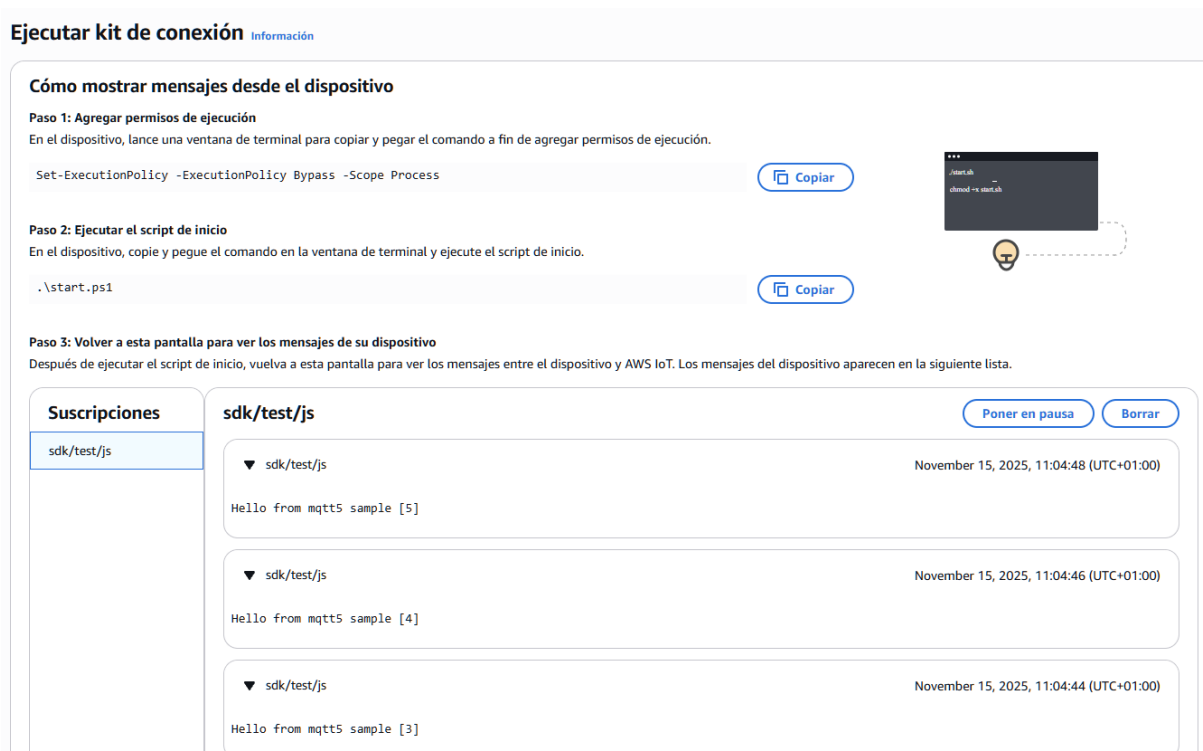
The screenshot shows the AWS DynamoDB console interface. At the top, it says 'Tables (1) Info'. Below that is a search bar labeled 'Find tables'. A table lists the details of the 'vigia_farolas_estado' table. It has a checkbox, the name 'vigia_farolas_estado', a status of 'Active' with a green checkmark, and a partition key of 'farola_id (S)'.

<input type="checkbox"/>	Name	Status	Partition key
<input type="checkbox"/>	vigia_farolas_estado	Active	farola_id (S)

Figura 1: Tablas DynamoDB

1.1.2 Configuración de AWS IoT Core

- **Conexión del dispositivo:** Mediante el asistente de AWS IoT Core registré un dispositivo nuevo. Durante este proceso, se generaron y descargaron los certificados de seguridad (`farolasestado.cert.pem`, la clave privada correspondiente y publica correspondiente). Esto nos servirá para realizar la prueba de conexión y para conectarnos mediante Node-RED posteriormente.
- **Prueba de conexión:** Para verificar la conectividad, se ejecutó el script de prueba `start.ps1` proporcionado por AWS. El script instaló las dependencias necesarias y envió con éxito cinco mensajes de prueba, que fueron validados tanto en la terminal local como en el cliente de pruebas MQTT en el tema `sdk/test/js` en el mismo panel de conexión de dispositivos IoT.



The screenshot shows the 'Ejecutar kit de conexión' page in the AWS IoT console. It has a title bar 'Ejecutar kit de conexión' with an 'Información' link. The main content area is titled 'Cómo mostrar mensajes desde el dispositivo' and contains three steps: 'Paso 1: Agregar permisos de ejecución', 'Paso 2: Ejecutar el script de inicio', and 'Paso 3: Volver a esta pantalla para ver los mensajes de su dispositivo'. Each step has a description and a 'Copiar' button. To the right of the steps is a terminal window showing the execution of the 'start.ps1' script. Below the steps is a 'Suscripciones' section with a table showing the 'sdk/test/js' topic. The table has columns for the topic name, the number of messages received, and the time received. The messages are 'Hello from mqtt5 sample [5]', 'Hello from mqtt5 sample [4]', and 'Hello from mqtt5 sample [3]'. There are 'Poner en pausa' and 'Borrar' buttons at the top right of the table.

Ejecutar kit de conexión Información

Cómo mostrar mensajes desde el dispositivo

Paso 1: Agregar permisos de ejecución
En el dispositivo, lance una ventana de terminal para copiar y pegar el comando a fin de agregar permisos de ejecución.

Set-ExecutionPolicy -ExecutionPolicy Bypass -Scope Process

Paso 2: Ejecutar el script de inicio
En el dispositivo, copie y pegue el comando en la ventana de terminal y ejecute el script de inicio.

.\start.ps1

Paso 3: Volver a esta pantalla para ver los mensajes de su dispositivo
Después de ejecutar el script de inicio, vuelva a esta pantalla para ver los mensajes entre el dispositivo y AWS IoT. Los mensajes del dispositivo aparecen en la siguiente lista.

Suscripciones

Topic	Messages	Time
▼ sdk/test/js	Hello from mqtt5 sample [5]	November 15, 2025, 11:04:48 (UTC+01:00)
▼ sdk/test/js	Hello from mqtt5 sample [4]	November 15, 2025, 11:04:46 (UTC+01:00)
▼ sdk/test/js	Hello from mqtt5 sample [3]	November 15, 2025, 11:04:44 (UTC+01:00)

Figura 2: Test start.ps1

■ Cambio en la Política:

Automáticamente se asocia una política de seguridad, esta política es bastante restrictiva la cambiaremos para que se permitan la conexiones y publicaciones a cualquier tema, yo lo haré en el tema "smartcity/consumo/test". Usamos 'test' porque en ningún momento serán farolas reales por lo que en un supuesto despliegue de la aplicación deberíamos de cambiar el tema. Cabe recalcar que para activar la política debemos de crear una nueva versión seleccionarla como activa y eliminar la versión anterior si no la vamos a volver a usar.

Versión activa: 2 [Información](#)

Efecto de la política	Acción de la política
Allow	iot:Connect
Allow	iot:Publish
Allow	iot:Receive
Allow	iot:Subscribe

Figura 3: Política activa

1.2 Creación de la Regla IoT

1.2.1 Consulta SQL

Antes de definir la regla, se creó un bucket en S3 para el almacenamiento histórico de la telemetría. La regla de IoT se configuró para procesar los mensajes que llegan al tema `smartcity/consumo/test`. Para la consulta SQL tuve que cambiar un par de instrucciones ya que no me funcionaba correctamente el `processed_time`. Luego veremos lo que use en la función de Node-RED pero por ahora analizaremos la consulta SQL que he utilizado.

Bloque 1: Consulta SQL para la regla IoT

```

1 SELECT
2     consumption_kw,
3     timestamp,
4     concat(day, '-', month, '-', year, ' ', hour, ':', minute, ':', seconds) AS
5     processed_time
6 FROM
7     'smartcity/consumo/test'
```

Como vemos en el SQL hemos quitado el `home_id` y hemos cambiado el `processed_time`, esto se debe a que AWS no me dejaba utilizar DynamoDBv2 con plantilla y al usar DynamoDB legacy la `clave_id` la he mapeado directamente en el apartado correspondiente de DynamoDB, de esta forma ya no me apareciera doble en la tabla de Dynamo (como clave y como atributo). Posteriormente cambie el `processed_time` porque no hubo forma de procesar el tiempo con la función propuesta. Esta ha sido la forma más sencilla de obtener el tiempo procesado. Podría haberlo procesado al crear el mensaje cambiando la función del Node-RED o después al recibir el mensaje, y he decidido hacerlo después para que las carpetas del S3 pudieran mapear mucho mas facil el tiempo, y así crear las carpetas sin errores.

1.2.2 Acción S3 bucket

En esta acción también cambiamos un apartado, la "clave", en el enunciado se pedía utilizar un almacenamiento en carpetas clasificadas por año por mes por día, el problema que le vi principalmente fue que en un sistema real podríamos tener diferentes farolas por lo que utilice una ultima clasificación que en la carpeta días tuviera carpetas por "farola_id". Dentro de esta ultima carpeta ya tenemos cada archivo ordenado por la hora, minuto y segundo usando "hh:mm:ss-data.json" en vez de con timestamp (que a mi se me ponía en milisegundos). He usado la siguiente clave con los parámetros obtenidos directamente desde sus variables del mensaje (en vez de desde el timestamp).

Bloque 2: Clave del S3

```

1 farolas/year=${year}/month=${month}/day=${day}/${home_id}/${hour}:${minute}:${seconds}-data.json
```

1.2.3 Accion DynamoDB

En esta otra acción se pedía usar DynamoDBv2, despues de probar mil formas cambie a la versión a la de DynamoDB normal en la cual nosotros mismos debemos de vincular la variable clave para usarla como clave de particion de nuestra tabla. Lo hice manualmente porque no sabia como insertar la plantilla de atributos que proporciona el enunciado. Y no conseguí que las claves se vincularan automáticamente, por lo que lo hice manualmente con DynamoDB de la siguiente forma.

The screenshot shows the configuration interface for a DynamoDB action in the AWS IoT console. The form is titled 'DynamoDB' and includes an 'Eliminar' button. The configuration steps are as follows:

- Nombre de la tabla:** 'vigia_farolas_estado'. There are buttons for 'Ver' and 'Crear una tabla de DynamoDB'.
- Clave de partición:** 'farola_id'. A note states: 'La clave de partición (también denominada clave hash) debe coincidir con la clave de partición de la tabla de DynamoDB que ha creado.'
- Tipo de clave de partición:** 'STRING'. A note states: 'El tipo de clave de partición (también denominada clave hash) puede ser STRING o NUMBER. El valor predeterminado es STRING.'
- Valor de clave de partición:** '\${home_id}'. A note states: 'El valor de clave de partición (también denominada clave hash) admite plantillas de sustitución que proporcionan datos en tiempo de ejecución.'
- Clave de ordenación:** 'MySortKey'. A note states: 'La clave de ordenación (también denominada clave de rango) debe coincidir con la clave de ordenación de la tabla de DynamoDB que ha creado.'
- Tipo de clave de rango:** (Empty). A note states: 'El tipo de clave de ordenación (también denominada clave de rango) puede ser STRING o NUMBER. El valor predeterminado es STRING.'
- Valor de clave de rango:** 'MysortKeyValue'. A note states: 'El valor de clave de ordenación (también denominada clave de rango) admite plantillas de sustitución que proporcionan datos en tiempo de ejecución.'
- Escritura de datos del mensaje en esta columna:** 'payload'. A note states: 'Escritura de datos del mensaje en esta columna: opcional'.
- Operación:** 'INSERT'. A note states: 'Operación: opcional. La operación puede ser INSERT, UPDATE o DELETE. El valor predeterminado es INSERT.'
- Rol de IAM:** 'LabRole'. There are buttons for 'Ver' and 'Crear un nuevo rol'. A note states: 'AWS IoT creará automáticamente una política con el prefijo "aws-iot-rule" bajo el rol de IAM seleccionado.'

Figura 4: Configuración de la acción DynamoDB

1.3 Simulación de ingesta con Node-RED

Node-RED permite conectarse mediante MQTT a nuestro tema de 'thing' de IoT para simular el envío de datos de las farolas. El flujo consta de tres nodos:

- **Inject:** Un disparador manual para iniciar el flujo, cada 10 segundos se envia una señal al siguiente nodo para que mande un mensaje.
- **Function:** Un nodo que construye el mensaje JSON con los datos de la farola (ID, consumo y el tiempo) y establece el tema de destino en `smartcity/consumo/test`. Este nodo consta de una serie de variables que hemos randomizado para simular la farola. He utilizado 6 variables de más para el tiempo, de forma que tengamos en variables separadas el año, el mes, el día, la hora, los minutos y los segundos. Consiguiendo que los pasos anteriores hayan sido más sencillos de implementar sabiendo que no tenemos que parsear ni usar funciones diferentes para obtener los datos del tiempo.

Bloque 3: Funcion constructora del mensaje

```
1 const now = new Date();
2 msg.topic = "smartcity/consumo/test";
3 msg.payload = {
4   home_id: "farola-001",
5   consumption_kw: +(Math.random() * 2).toFixed(2),
6   timestamp: now.getTime(),
7   year: now.getFullYear(),
8   month: String(now.getMonth() + 1).padStart(2, "0"),
9   day: String(now.getDate()).padStart(2, "0"),
10  hour: String(now.getHours()).padStart(2, "0"),
11  minute: String(now.getMinutes()).padStart(2, "0"),
12  seconds: String(now.getSeconds()).padStart(2, "0"),
13 };
14 return msg;
```

- **MQTT Out:** Este nodo está configurado para publicar el mensaje en el endpoint de AWS IoT (a1paa0c5brn8mp-ats.iot.us-east-1.amazonaws.com) a través del puerto 8883 con TLS. La autenticación se realizó utilizando los mismos certificados de dispositivo que en la prueba de conexión inicial. Usamos TLS el cual contiene todos los certificados necesarios para conectarnos al endpoint que anteriormente hemos introducido.

Con estos tres nodos configurados tenemos ya el flujo funcional, y en darle a instanciar tendremos el sistema enviando mensajes cada 10 segundos al tema de MQTT de nuestro objeto. Para parar el sistema es tan simple como deshabilitar el nodo del disparador y volver a darle a instanciar.

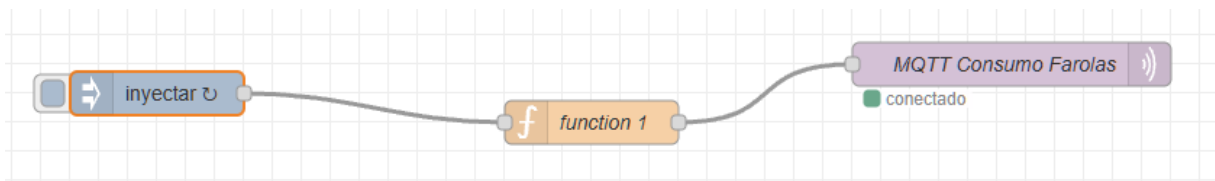


Figura 5: Flujo Node-RED

1.3.1 Pruebas

- **Cliente de prueba MQTT:** Para comprobar que cada una de las acciones anteriores de la regla funcionen debemos de comprobarlo lanzando el Node-RED. El primer paso es comprobar que los mensajes son enviados al tema correspondiente, para esto nos vamos al cliente de prueba de MQTT dentro de la pestaña de IoT de AWS y nos subscribimos a `smartcity/consumo/#` cuando nos subscribamos nos empezaran a salir mensajes como estos.

Figura 6: Cliente de prueba de MQTT

- **S3:** Respecto al guardado de mensajes historicos debemos de comprobarlo dentro del S3 de la practica, nos dirigimos a la sucesiones de carpetas que tenemos configurada y dentro de la ultima vemos como tenemos estos datos. Los datos de dentro de cada archivo tambien corresponden con el .json definido en el SQL, ej:

```
1 {"consumption_kw":0.3,"timestamp":1763374077358,"processed_time":"17-11-2025 11:07:57"}
```



Figura 7: S3 bucket

- **DynamoDB** Finalmente debemos de comprobar la tabla de DynamoDB, dentro de `vigia_farolas_estado`, y en explorar elementos de la tabla vemos como nuestra tabla contiene una `farola001` con el json correspondiente. Este json se va actualizando cada 10 segundos y muestra siempre el estado actual de la farola. Si tuviéramos mas farolas se mostraría una tabla con el estado actual de cada una de las farolas.

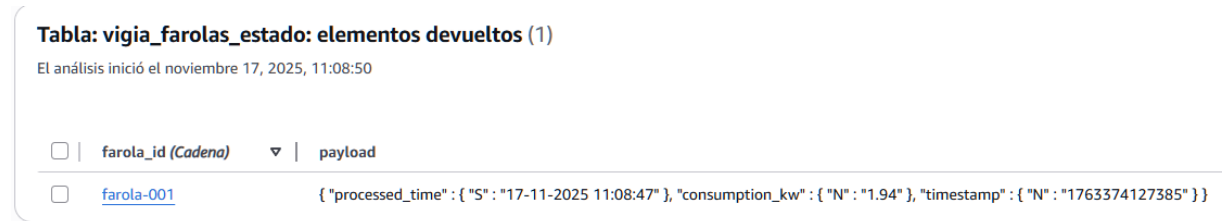
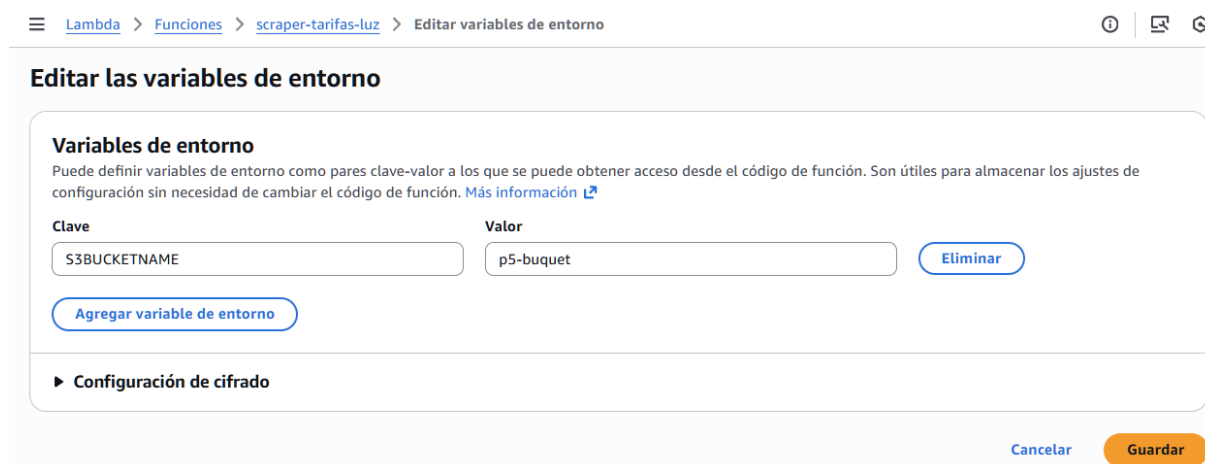


Figura 8: Tabla DynamoDB elementos

2 Flujo de Contexto: Web Scraping con Lambda Serverless (RA2)

2.1 Creación de la Función Lambda

Para crear la función Lambda debemos de ingresar al servicio Lambda y crear una función de python desde cero con los permisos de Lab-Role para poder ejecutar correctamente el proceso y realizar el PutObject al S3. Una vez dentro se nos abrirá un editor de código, que en si es un fork web de visual Studio Code. Debemos de pasar la variable por entorno, para esto dentro del panel de configuración ingresamos a variables de entorno y copiamos el nombre del bucket con el nombre de la variable del código proporcionado.



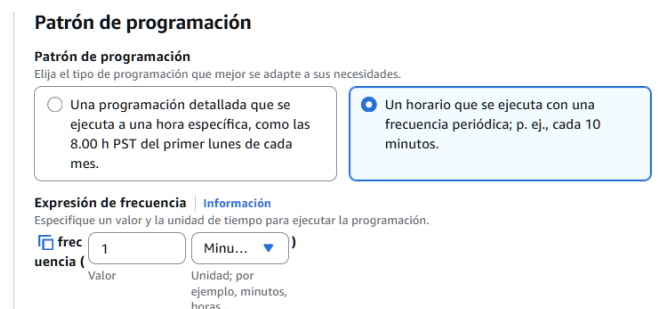
The screenshot shows the 'Editar las variables de entorno' (Edit environment variables) page in the AWS Lambda console. The breadcrumb trail is 'Lambda > Funciones > scraper-tarifas-luz > Editar variables de entorno'. The page title is 'Editar las variables de entorno'. Below the title, there is a section 'Variables de entorno' with a description: 'Puede definir variables de entorno como pares clave-valor a los que se puede obtener acceso desde el código de función. Son útiles para almacenar los ajustes de configuración sin necesidad de cambiar el código de función. Más información'. Below this, there is a table with two columns: 'Clave' (Key) and 'Valor' (Value). The table contains one row with the key 'S3BUCKETNAME' and the value 'p5-buquet'. To the right of the value is an 'Eliminar' (Delete) button. Below the table is an 'Agregar variable de entorno' (Add environment variable) button. At the bottom of the table is a 'Configuración de cifrado' (Encryption configuration) section. At the bottom right of the page are 'Cancelar' (Cancel) and 'Guardar' (Save) buttons.

Figura 9: Variable de entorno

Cuando tengamos la variable de entorno copiamos el código que simulará nuestra ingesta de datos de contexto aletorizando la tarifa de luz del día `tarifakwh`. Tras copiar el código ya podemos desplegarlo.

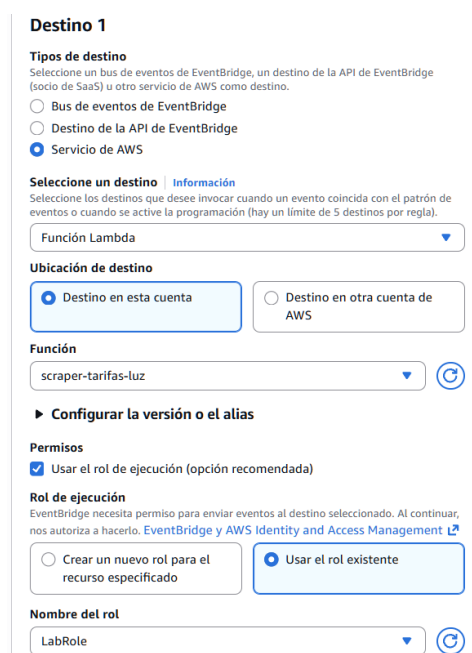
2.2 Orquestación Serverless (EventBridge)

Al igual que hemos hecho en el apartado anterior con la ingesta de telemetría mediante la simulación por Node-RED, en este apartado también tenemos que simular una entrada programada cada cierto tiempo. En este caso mandaremos una señal a nuestra función Lambda cada 1 minuto. Esto lo haremos mediante el servicio de Amazon EventBridge, este nos permitirá crear un evento programado al destino de AWS que nosotros queramos. Lo configuraremos de la siguiente forma eligiendo nuestra función Lambda y poniendo una programación que se repita cada minuto.



The screenshot shows the 'Patrón de programación' (Pattern of programming) page in the AWS EventBridge console. The page title is 'Patrón de programación'. Below the title, there is a section 'Patrón de programación' with a description: 'Elija el tipo de programación que mejor se adapte a sus necesidades.' Below this, there are two radio buttons. The first is 'Una programación detallada que se ejecuta a una hora específica, como las 8.00 h PST del primer lunes de cada mes.' The second is 'Un horario que se ejecuta con una frecuencia periódica; p. ej., cada 10 minutos.' The second option is selected. Below the radio buttons, there is a section 'Expresión de frecuencia' with a description: 'Especifique un valor y la unidad de tiempo para ejecutar la programación.' Below this, there is a 'frecuencia' (frequency) section with a value of '1' and a unit of 'Minu...' (minutes).

Figura 10: Frecuencia Event Bridge



The screenshot shows the 'Destino 1' (Destination 1) page in the AWS EventBridge console. The page title is 'Destino 1'. Below the title, there is a section 'Tipos de destino' (Destination types) with a description: 'Seleccione un bus de eventos de EventBridge, un destino de la API de EventBridge (socio de SaaS) u otro servicio de AWS como destino.' Below this, there are three radio buttons: 'Bus de eventos de EventBridge', 'Destino de la API de EventBridge', and 'Servicio de AWS'. The third option is selected. Below the radio buttons, there is a section 'Seleccione un destino' (Select a destination) with a description: 'Seleccione los destinos que desee invocar cuando un evento coincida con el patrón de eventos o cuando se active la programación (hay un límite de 5 destinos por regla).' Below this, there is a dropdown menu set to 'Función Lambda'. Below the dropdown menu, there is a section 'Ubicación de destino' (Destination location) with two radio buttons: 'Destino en esta cuenta' (Destination in this account) and 'Destino en otra cuenta de AWS'. The first option is selected. Below the radio buttons, there is a section 'Función' (Function) with a dropdown menu set to 'scraper-tarifas-luz'. Below the dropdown menu, there is a section 'Configurar la versión o el alias' (Configure the version or alias). Below this, there is a section 'Permisos' (Permissions) with a checked box 'Usar el rol de ejecución (opción recomendada)' (Use the execution role (recommended)). Below the checked box, there is a section 'Rol de ejecución' (Execution role) with a description: 'EventBridge necesita permiso para enviar eventos al destino seleccionado. Al continuar, nos autoriza a hacerlo. EventBridge y AWS Identity and Access Management'. Below this, there are two radio buttons: 'Crear un nuevo rol para el recurso especificado' (Create a new role for the specified resource) and 'Usar el rol existente' (Use the existing role). The second option is selected. Below the radio buttons, there is a section 'Nombre del rol' (Role name) with a dropdown menu set to 'LabRole'.

Figura 11: Destino Event Bridge

Si nos fijamos en el código de nuestra función, el guardado de datos se sobrescribiera al poner como nombre DD-MM-YYYY-data.json. Pero podremos comprobar que se sobrescriba correctamente comprobando el contenido del archivo del S3 cada minuto. Nuestros archivos `-data.json` tendrán la estructura de este ejemplo.

```
1 {"timestampscraper": "2025-11-17T18:20:11.307175", "tarifakwh": 0.1833, "ciudad": "Neo-Tech"}
```

Nuestra carpeta dentro del bucket dentro esta otra forma.

tarifas/

Objetos | Propiedades

Objetos (1) Copiar URI de S3

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para concederles permisos de forma explícita. [Más información](#)

🔍 *Buscar objetos por prefijo*

<input type="checkbox"/>	Nombre	Tipo	Última modificación	Tamaño	Clase de almacenamiento
<input type="checkbox"/>	2025-11-17-data.json	json	17 Nov 2025 7:40:11 PM CET	93.0 B	Estándar

Figura 12: S3 tarifas

Finalmente en los registros de cloudWatch podemos ver las invocaciones a la función mas recientes.

Invocaciones recientes							
#	Timestamp	RequestId	LogStream	DurationInMS	BilledDurationInMS	MemorySetInMB	MemoryUsedInMB
1	2025-11-17T18:58:11.013Z	575619ea-a83b-471c-9d6f-37f6058545f0	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	208.84	209.0	128.0	86.0
2	2025-11-17T18:57:10.973Z	c6ac8c9f-68dd-4933-af51-a44ff7ae7fe7	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	212.12	213.0	128.0	86.0
3	2025-11-17T18:56:10.998Z	227f0914-51f4-486a-86bd-a33f9685c66f	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	234.47	235.0	128.0	86.0
4	2025-11-17T18:55:11.153Z	3d3c76bb-5ebc-4971-818a-1d2d424088bd	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	228.2	229.0	128.0	86.0
5	2025-11-17T18:54:10.952Z	75f3bae5-0c4a-45cd-905f-4ffc487708f2	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	278.49	271.0	128.0	86.0
6	2025-11-17T18:53:10.897Z	3dcace58-048f-499c-8faa-8381f8bd7b81	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	242.78	243.0	128.0	86.0
7	2025-11-17T18:52:11.173Z	0ce625f9-53d6-4f0f-b4eb-089ca92fae0f	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	229.41	230.0	128.0	86.0
8	2025-11-17T18:51:10.889Z	b116b073-292c-48fb-8908-d2524f95e4ec	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	221.17	222.0	128.0	86.0
9	2025-11-17T18:50:11.634Z	78b36c11-ff32-4986-846c-72811370e9b3	2025/11/17/[SLATEST]d6aae628ca1c4256887b1a572de74437	212.55	213.0	128.0	86.0

Figura 13: S3 tarifas

3 Flujo Masivo: Transformación y Análisis Serverless