BIG DATA

ALMACENAMIENTO DATOS EN AWS - DYNAMODB

EDUARD LARA

□ Es una clase de almacenamiento de Amazon S3
 □ Seguro, duradero y de bajo coste.
 □ Sirve para guardar gran cantidad de datos a escala de petabytes durante un tiempo prolongado
 □ Permite hacer copias de seguridad a largo plazo.
 □ Permite guardar nuestros datos a bajo coste
 □ Es un almacenamiento de ficheros en frío, es decir, no se pueden consultar en tiempo real.

Almacén (Vault)

- ☐ En Glacier un almacén (Vault) es un contenedor para guardar archivos.
- □ Cuando se crea un almacén o un contenedor se debe de especificar un **nombre** y seleccionar una **región** en la que va a estar creado.
- ☐ Se pueden crear **políticas** para tus almacenes. Por ej:
 - □ bloquear notificaciones de los datos,
 - □ control de acceso por IAM,
 - □ no permitir borrar ficheros durante un año, etc..

Almacén bloqueado (Vault lock)

- ☐ En Glacier un almacén bloqieado (vault lock) sirve para construir almacenes de datos con una configuración específica que no se podrá cambiar.
- ☐ Esto es útil para cumplir normativas en cuanto a almacén seguro de datos.
- Las políticas son creadas utilizando con IAM. Una vez creadas para ese almacén, se tienen 24 horas para revisarlas. Pasado ese tiempo ya no se podrán modificar y quedará así de forma permanente.
- ☐ Asi nos aseguramos que cumplimos la normativas, porque ese almacén está bloqueado con esas propiedades y no se podrá cambiar.

Archivos

- □ Los archivos que se pueden almacenar en Glacier son de cualquier tipo de datos: fotos, videos, documentos, etc.
- ☐ Cada archivo tiene un identificador único y una descripción opcional que se pone al principio cuando se crea el archivo.

Trabajos

- □ Los **trabajos** en Glacier pueden servir para consultar archivos, para recuperar un archivo o para obtener el inventario de un almacén.
- ☐ Un almacén o un contenedor puede tener varios trabajos en ejecución al mismo tiempo.

Notificaciones

- □ Como los trabajos en Glacier tardan un tiempo en completarse, Glacier tiene un mecanismo de notificaciones para avisar cuando se completa un trabajo
- □ Podemos configurar un almacén para que envíe una notificación a un tema del SNS cuando los trabajos se completan
- □ Podemos utilizar los temas de SNS para enviar una notificación a un tema (a un grupo de personas, o la dirección de correo electrónico, etc..)
- □ Sólo va a haber una configuración de notificaciones por almacén.

□ DynamoDB es un servicio de base de datos no SQL que está completamente administrado, no hay limitaciones de almacenamiento. ☐ Permite gestionar los permisos mediante accesos con IAM. ☐ Es una colección de tablas. Las tablas son la estructura de más alto nivel en estas bases de datos DynamoDB. □ Los requerimientos de rendimiento: el número de lecturas o escrituras por segundo se especifica a nivel de tablas ☐ Tenemos dos requerimientos: □ WCU = Write Capacity Units = número de bloques de un 1KB que puede escribir por segundo □ RCU = Read Capacity Units = el número de bloques de 4KB que pueden leer por segundo.

- □ El coste de nuestra base de datos DynamoDB viene determinada por los valores que pongamos de WCU y RCU en cada tabla
- ☐ Es un sistema distribuido, es decir, que cada vez que grabamos un dato se distribuye por otras zonas geográficas.
- ☐ Por esta razón es eventualy consistent reads, es decir, es de lecturas consistentes y eventuales.
- ☐ Pero podemos forzar para que sea strong consistent reads, es decir, que sean las lecturas consistentes de forma permanente o fuerte.

- □ Como es una base de datos de tipo no SQL, no tiene esquemas definidos para cada una de las tablas
- □ Las tablas son flexibles, podemos añadir más columnas en cualquier momento.
 - □Items = filas de las tablas
 - \square Atributos = columnas de las tablas.
- □ Cada fila o cada item tiene un número variable de atributos o columna rellenos. Podemos crear una fila con cuatro columnas rellenas y otra fila solamente con dos.
- □ Podemos rellenar un número arbitrario de columnas o atributos.

☐ Tenemos dos atributos especiales: □ Partition key o Hash Key, clave de partición, contiene un valor único para cada ítem o cada fila de la tabla. □ Sort Key, clave de ordenación, contiene un segundo valor que puede servir para ordenar los datos. Por ejemplo, un campo de fecha y tiempo ☐ Tipos de atributos: □ String, number, binary, boolean, null, document, set ☐ Integraciones con DynamoDB: □Integración directa con EMR, S3, Data Pipeline, lambda, Kinesis Streams

3. DYNAMODB. PARTICIONES

□ Las particiones en DynamoDB son los nodos de almacenamiento y procesamiento de DynamoDB. ☐ Inicialmente una tabla equivale a una partición. Todos los datos de la tabla se almacenan en una partición. ☐ Una partición puede almacenar hasta un máximo de 10GB de datos. Una partición gestiona hasta un máximo de 3000 RCU y 1000 WCU ☐ El numero de particiones afecta al rendimiento del sistema, es decir, hay que diseñar las tablas y aplicaciones para evitar problemas de entrada/salida ☐ Si se superan alguno de los límites 10GB, 3000 RCU o 1000 WCU los datos son divididos en varias particiones

3. DYNAMODB. PARTICIONES

☐ Los datos se distribuyen por las distintas particiones, basándose en el atributo partition key, □ El número de particiones es gestionado por DynamoDB según nuestros requisitos de rendimiento RCU y WCU ☐ Si aumentamos los valores para RCU o WCU entonces Dynamo DB automáticamente aumentará el nº particiones ☐ Si bajamos los valores de RCU o WCU, DynamoDB ya no reducirá el nº de particiones, con lo cual tendremos un exceso de almacenamiento sin usar en cada partición. □ Como el nº de particiones afecta al rendimiento, hay que ser cuidadoso a la hora de tocar los requerimientos de rendimiento RCU y WCU.

- □ Vamos a ver los índices secundarios globales y locales
 □ DynamoDB ofrece 2 operaciones recuperación de datos:
 □ Scan se utiliza para revisar la estructura de una tabla, los atributos, etc
 □ Query: sirve para obtener datos de la tabla mediante partition key o el sort Key e indicando una partición.
 □ Puedes obtener una única fila de la tabla pasándole un
- valor concreto en la partition key o puedes obtener múltiples filas de la tabla pasándole una partition key y una sort key.

- □ Los índices mejoran el rendimiento de las consultas a las tablas. Existen dos tipos de índices:
 - □LSI: el índice secundario global
 - $\square GSI$: el índice secundario local.

LSI Índice secundario local

- ☐ Este índice sólo se puede crear en el momento en el que se crea la tabla.
- ☐ Contiene la clave de partición, la clave de ordenación, la nueva clave de ordenación y también podemos incluir otros atributos opcionales de la tabla que queramos introducir (atributos proyectados)

□ Los datos de la tabla son copiados de forma asíncrona a todos los índices locales secundarios de esa tabla. □ LSI comparte el RCU y WCU con la tabla ☐ Es un índice independiente, □ Sólo contendrá aquellas filas que tengan rellena la información de la nueva clave de ordenación o de aquellos atributos opcionales incluidos en la clave, es decir, los atributos proyectados, □ Tiene un menor número de filas, por tanto, las consultas serán más eficientes.

Índice secundarios Global

- □ Los índices secundarios globales (Global secondary index GSI) se pueden crear en cualquier momento. Son una alternativa a la clave de partición y a la de ordenación
- □ Las opciones para los atributos de proyección son:
 - □KEYS_ONLY: solo incluye clave partición y ordenación
 - □INCLUDE: atributos proyectados seleccionados
 - □ALL: todos los atributos son seleccionados para esta clave global
- ☐ GSI puede definir su propio RCU y WCU, al igual que una tabla. Los cambios escritos en la tabla son copiados de forma asíncrona a este índice secundario global.

5. DYNAMODB. STREAMS

- □ Los streams son un conjunto ordenado de actualizaciones realizadas sobre una tabla de DynamoDB.
- □ Cuando un stream está habilitado en una tabla, los registros que se modifican en esta tabla son almacenados durante 24 horas.
- ☐ Se pueden habilitar los stream en una tabla vía consola o mediante una API.
- □ Los datos almacenados sólo pueden ser leídos por un endpoint o por llamadas a una API.

5. DYNAMODB. STREAMS

- ☐ Se pueden configurar distintas formas:
 - □KEYS_ONLY: se almacena en los streams los atributos que son clave
 - □NEW_IMAGE, sólo almacena en los streams la fila entera que se ha actualizado
 - □OLD_IMAGE: se almacena la fila entera previa al cambio.
 - □NEW_AND_OLD_IMAGES: se almacena la fila previa al cambio y también como ha quedado después de realizar el cambio.

5. DYNAMODB. STREAMS

Casos de uso:

- □ Replicación de datos, por ejemplo la replicación de datos entre regiones de AWS para hacer copias de seguridad.
- ☐ Trigers o los lanzadores de eventos con lambda.
 - □Una función Lambda se lanza cuando se añade una nueva fila en una tabla para analizar sus datos.
 - □ Lanzar una función lambda cuando ocurra un evento o algún evento en nuestra aplicación, por ejemplo un nuevo registro, etc...

6. RENDIMIENTO EN DYNAMODB

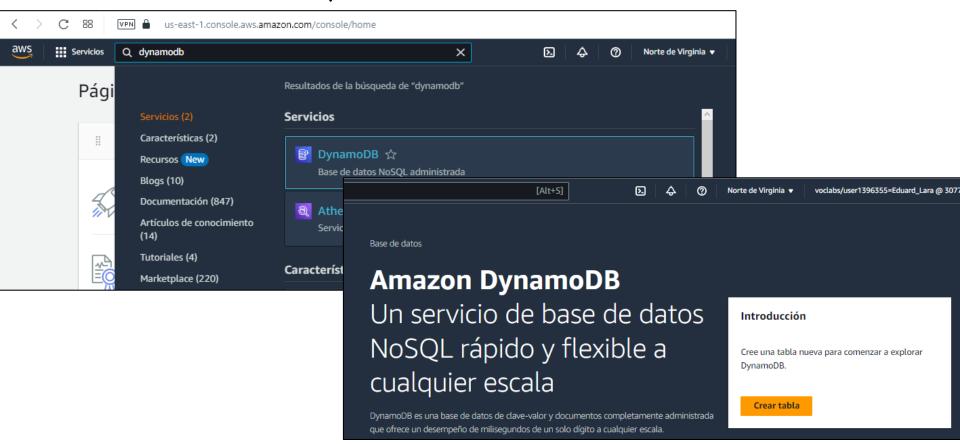
□ Las particiones van a definir el rendimiento que tendrá nuestra base de datos en DynamoDB. ☐ Hay dos fórmulas para el calculo del nº de particiones. □ La primera de ellas tiene en cuenta el rendimiento: Número de particiones = (RCU deseado / 3000 RCU) + (WCU deseado /1000 WCU) □En la segunda fórmula, sería según la capacidad: Número de particiones = (Tamaño de los datos GB /10) ☐ El número de particiones totales será el valor máximo

de estas dos fórmulas.

6. RENDIMIENTO EN DYNAMODB

☐ Las lecturas y las escrituras en nuestra base de datos son distribuidas entre las distintas particiones de forma equitativa. ☐ Se distribuirán de forma equitativa si tenemos bien definida nuestra clave de partición. Así los datos irán de forma equitativa a las distintas particiones, e iremos las mismas veces a cada una de las particiones. ☐ Esta clave de partición debe ser un atributo que debe tener muchos valores diferentes y que tenga un patrón uniforme de lecturas escrituras en todas las particiones. ☐ Si no encontramos ese atributo único, podemos crear esta tabla mediante la unión de varios atributos.

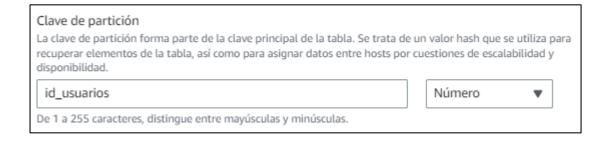
- Paso 1. Vamos a crear nuestra primera tabla con datos en DynamoDB. Vamos al servicio DynamoDB y le damos a crear tabla.
- Bueno, y este es el primer laboratorio sencillo para que veais como se crea una tabla y como se introducen

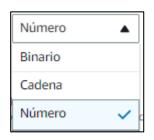


Paso 2. No es necesario crear una base de datos, ya que eso lo gestiona internamente DynamoDB vamos directamente a la creación de la tabla, así que ponemos el nombre a nuestra tabla. Se va a llamar usuarios.



Paso 3. La clave principal o clave de partición va a ser el identificador id_usuarios. Debemos definir el tipo de datos. Puede ser de tipo cadena, binario o numero. En este caso le vamos a poner un número,





Paso 4. En configuración de la tabla vamos a desactivar el check de configuración predeterminada.



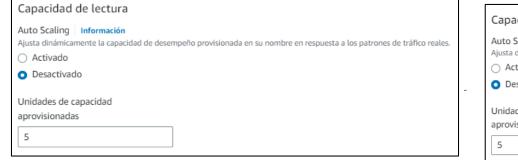
Paso 5. En cuanto al Modo de capacidad de lectura/escritura, vamos a dejar por defecto Aprovisionado



Paso 6. Desactivamos el Auto-Scaling de momento. Hace que cuando nuestra capacidad de lectura supera al 70%, va añadiendo más capacidad, más unidades de 1 en 1 hasta un máximo de 10. En este caso, tanto en lectura como escritura, es muy recomendable hacer Auto-Scaling.

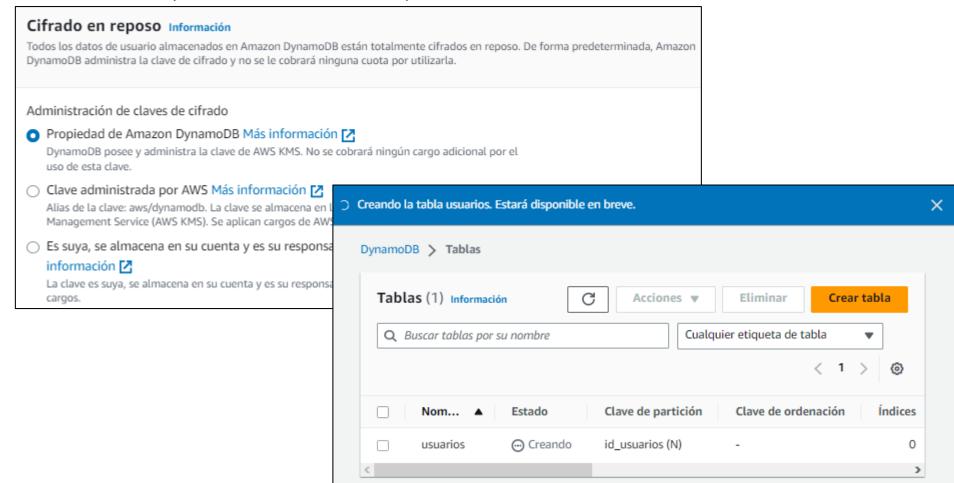
Capacidad de lectura									
Auto Scaling Información Ajusta dinámicamente la capacidad de desempeño provisionada en su nombre en respuesta a los patrones de tráfico reales. Activado									
O Desactivado									
b)									
70									
-									

En este caso, como es una tabla de ejemplo lo vamos a desactivar

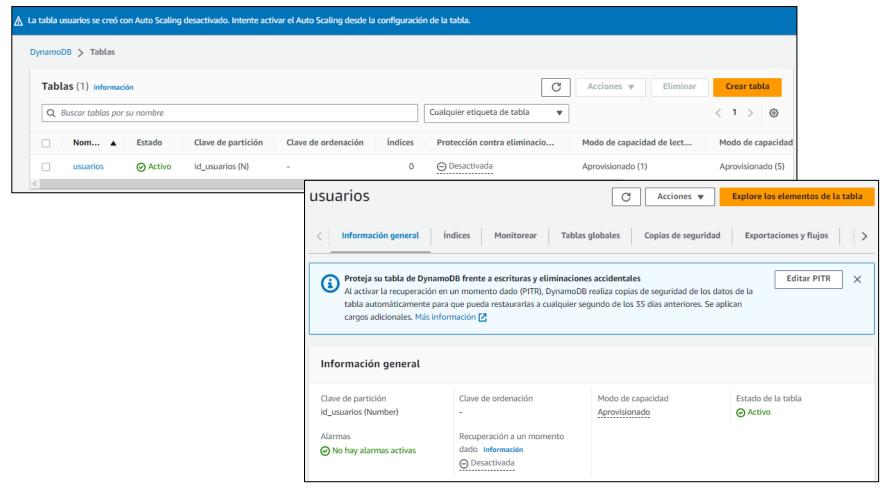


П								
	Capacidad de escritura							
Auto Scaling Información Ajusta dinámicamente la capacidad de desempeño provisionada en su nombre en respuesta a los patrones de tráfi								
	○ Activado							
	Desactivado							
	Unidades de capacidad aprovisionadas							
	5							

Paso 7. El cifrado en reposo, dejamos el predeterminado "Propiedad de Amazon DynamoDB" y le damos al botón de crear. Esto creará nuestra primera tabla en DynamoDB.



Paso 8. Una vez creada, nos indica la información principal, el nombre de la tabla, el identificador, el tipo de dato, etc, todos los datos que hemos creado



Paso 9. Para añadir elementos nuevos a nuestra tabla, vamos al botón Acciones y le damos a Crear elemento. Esto permite añadir filas a

Acciones

Editar capacidad

Eliminar tabla
Crear elemento
Crear índice
Crear réplica

Actualizar la clase de tabla

Explore los elementos de la tabla

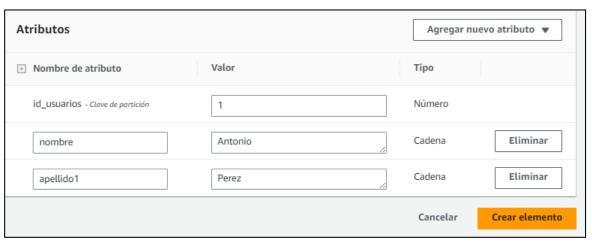
nuestra tabla.



Paso 10. Al identificador del primer usuario le ponemos 1. Añadimos un atributo de tipo cadena que será el nombre.



Paso 11. Añadimos otro atributo de tipo cadena, apellido1. Y lo dejamos así,



Paso 12. Le damos a guardar y esto añadirá una fila a nuestra tabla con la información que hemos introducido: Usuario numero 1, Apellido Pérez y Nombre Antonio.



Paso 13. Creamos un segundo elemento con los siguientes valores:

id_usuario 2

nombre Luis

apellido2 Gonzalez

Edad 35 numero

Le damos a guardar. Y añade la fila número 2.



Paso 14. En este caso ha dejado el apellido1 que no sabíamos en blanco ha añadido apellido2 y ha añadido la edad. En DynamoDB se pueden dejar valores de las columnas sin datos no hay ningún problema y podemos ir añadiendo números de columna según vayamos necesitando, con lo cual vemos que muy flexible a la hora de introducir nuevas filas con datos en nuestra tabla de DynamoDb

Eleme	entos devueltos (2	2)						
	id_usuarios	▼ apellio	01 ▽	apellido2	▽	NewValue	▽	nombre
	2			Gonzalez		35		Luis
	1	Perez						Antonio

Paso 15. En la pestaña Información General podemos ver Métricas de capacidad de tabla que es información sobre nuestras tablas, gráficas de rendimiento



Paso 16. En monitorear podemos añadir alarmas

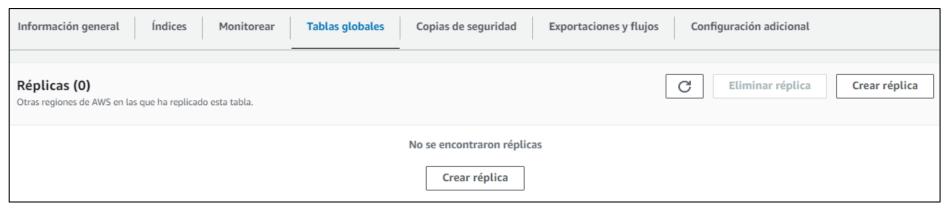


Paso 17. En configuracion adicional podemos cambiar la capacidad de lectura y escritura de nuestra tabla,



Paso 18. Vemos índices, tablas globales, copia de seguridad, etc Tenemos todas las posibilidades de cambiar nuestra configuración de la tabla.

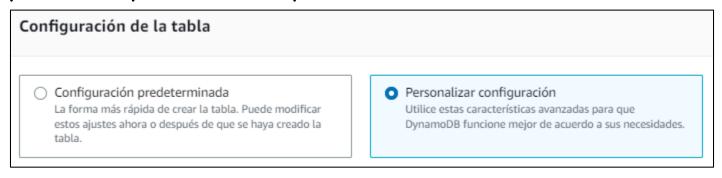




Paso 19. Queremos crear otra nueva tabla, le damos al botón Crear tabla y creamos la tabla artículos y la clave principal esta tabla va a ser primero el identificador de usuarios del que el usuario va a recibir el artículo y el número o identificador del artículo. en este caso todos van a ser numéricos. Vamos a crear una clave principal formada por dos partes una, la clave de partición, que será que el usuario va a crear el artículo y a su vez va añadir una clave de ordenación que va a ser el identificador del artículo de ese usuario.

Nombre de la tabla Se utilizará para identificar su tabla. articulos Entre 3 y 255 caracteres. Solo se pueden usar letras, números, guiones bajos (_), guiones (-) y puntos (.).	
Clave de partición La clave de partición forma parte de la clave principal de la tabla. Se tabla, así como para asignar datos entre hosts por cuestiones de esca id_usuario De 1 a 255 caracteres, distingue entre mayúsculas y minúsculas.	
Clave de ordenación - <i>opcional</i> Puede utilizar una clave de ordenación como segunda parte de la cla buscar entre todos los elementos que comparten la misma clave de de la cla articulo De 1 a 255 caracteres, distingue entre mayúsculas y minúsculas.	ve principal de una tabla. La clave de ordenación le permite ordenar o partición. Número

Paso 20. Quitamos la opción de configuración predeterminada. En configuración de capacidad de lectura/escritura dejamos el modo de capacidad Aprovisionado por defecto.





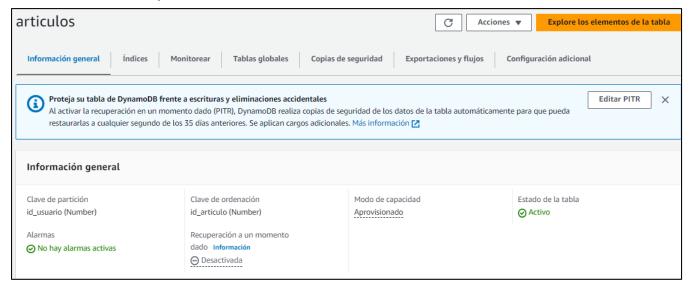
Paso 21. El auto Scaling de momento Lo desactivamos. El cifrado en reposo lo dejaos en predeterminado, y le damos a crear nuestra segunda tabla en dynamoDB

Capacidad de lectura					
Auto Scaling Información Ajusta dinámicamente la capacidad de desempeño provisionada en s Activado	su nombre en respuesta a los patrones de tráfico reales.				
 Desactivado 					
Unidades de capacidad aprovisionadas					
Capacidad de escritura	Cifrado en reposo Información Todos los datos de usuario almacenados en Amazon DynamoDB están totalmente cifrados en reposo. De forma predeterminada, Amazon DynamoDB administra la clave de cifrado y no se le cobrará ninguna cuota por utilizarla.				
Auto Scaling Información Ajusta dinámicamente la capacidad de desempeño provisionada en s Activado	Administración de claves de cifrado Propiedad de Amazon DynamoDB Más información DynamoDB posee y administra la clave de AWS KMS. No se cobrará ningún cargo adicional por el uso de esta clave.				
Desactivado Unidades de capacidad aprovisionadas	 ○ Clave administrada por AWS Más información Alias de la clave: aws/dynamodb. La clave se almacena en la cuenta y es administrada por AWS Key Management Service (AWS KMS). Se aplican cargos de AWS KMS. ○ Es suya, se almacena en su cuenta y es su responsabilidad administrarla. Más 				
5	información ☑ La clave es suya, se almacena en su cuenta y es su responsabilidad administrarla. AWS KMS cobra cargos.				

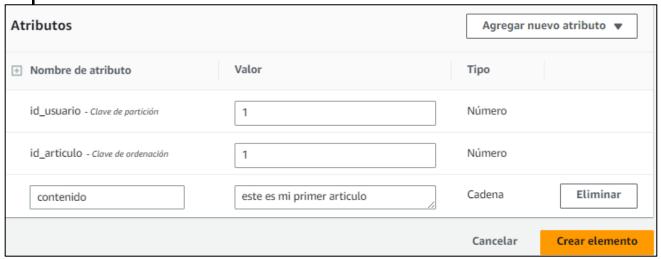
Paso 22. se ha creado la tabla de artículos,



En la pestaña información general esta la información de la tabla, el nombre de la tabla, los dos identificadores que forman la clave principal de la tabla, etc..



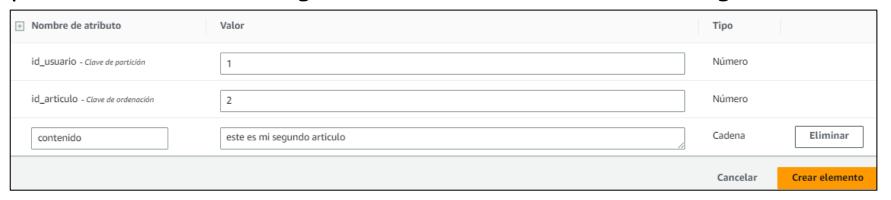
Paso 23. Añadir elementos. Tenemos que poner datos obligatorios para las dos identificadores, en este caso para el usuario1 va a escribir el artículo número uno. Vamos a poner el contenido y ponemos "este es mi primer artículo".



Le vamos a guardar y vemos que ha añadido esta primera fila al usuario 1. Ha añadido un artículo y este es el contenido.

id_usuario	\triangledown	id_articulo	▽	contenido
1		1		este es mi primer articulo

Paso 24. Creamos otro elemento. Indicamos que usuario1 ha añadido el artículo2. Añadimos otro atributo tipo una cadena llamado contenido y ponemos "Este es mi segundo artículo". Hacemos click en guardar.



Vemos que ha añadido la segunda fila, donde hay un identificador de usuario y una columna de ordenación que va ordenando por el identificador de artículo. Esta tabla en la clave principal tiene dos columnas y la anterior tabla tenía solamente una columna.

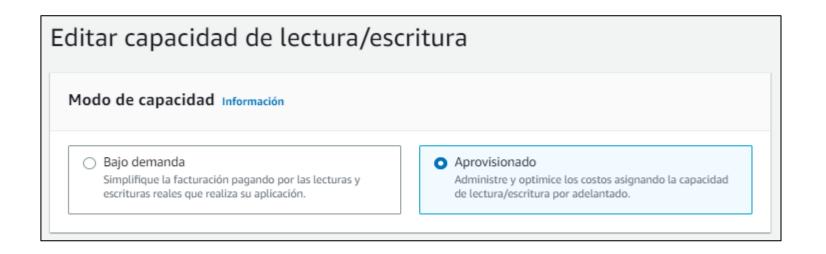
id_usuario	\triangledown	id_articulo	\triangledown	contenido
1		1		este es mi primer articulo
1		2		este es mi segundo articulo

Paso 1. Vamos a hacer ahora un laboratorio sobre dynamoDB dentro de la parte de capacidades y rendimiento de nuestras tablas. Vamos a dynamoDB, al apartado de tablas. Vemos las dos tablas creadas: artículos y usuarios. Vamos a la tabla de artículos.

Esta es la forma de configurar el rendimiento o la capacidad de nuestras tablas en dynamoDB



Paso 2. En la pestaña configuración adicional vamos a capacidad de lectura/escritura. Hay dos formas de definir nuestra capacidad de lectura y escritura. Por un lado, podemos elegir la de Aprovisionado, es decir, apta para la capa gratuita. Aprovisionado significa que nosotros decidimos cuántas unidades de lectura o escritura simultánea se pueden hacer en nuestra tabla.



Paso 3. En este caso tenemos puesto una capacidad lectura y escritura de 1 unidad simultáneas El coste es de 0.59\$/mes

▼ Calculadora de cap	acidad				
Tamaño medio de elemento	(KB)				
1					
Lectura de elemento/seguno	do	Consistencia de lectura			
1		Eventualmente consistente			
Escritura de elemento/segur	ndo	Consistencia de escritura			
1		Estándar ▼			
Capacidad de lectura (en	Capacidad de escritura	Región	Costo estimado		
unidades)	(en unidades)	us-east-1	0,59 US\$/mes		
1	1				

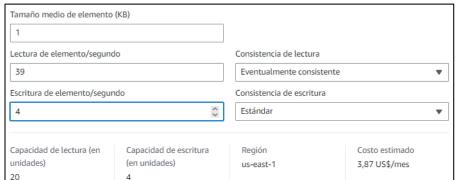
Si ponemos una capacidad de 5 lecturas simultaneas y 5 escrituras simultaneas sale un coste de 2,91\$/mes

Tamaño medio de elemento	(KB)					
Lectura de elemento/segund	lo	Consistencia de lectura				
10		Eventualmente consistente				
Escritura de elemento/segun	do	Consistencia de escritura				
5		Estándar ▼				
Capacidad de lectura (en	Capacidad de escritura	Región	Costo estimado			
unidades)	(en unidades)	us-east-1	2,91 US\$/mes			
5	5		_,			

Paso 4. Podemos ir variando, por ejemplo, 10 lecturas y 10 escrituras,

eniendo un coste por				
1				
Lectura de elemento/segur	ndo	Consistencia de lectur	ra .	
19	\$	Eventualmente cons	istente	•
Escritura de elemento/segu	ındo	Consistencia de escrit	ura	
10		Estándar		•
Capacidad de lectura (en	Capacidad de escritura	Región	Costo estimado	
unidades)	(en unidades)	us-east-1	5,81 US\$/mes	
10	10	us-cast* I	3,01 03\$/IIIes	

Paso 5. Si ponemos muchas lecturas 20, pero pocas escrituras 4, obtenemos un coste aproximado de lo que nos va a costar 3,87\$ por mes mantener esta tabla, según nuestras capacidades de 20 lecturas y 4 escrituras simultaneas

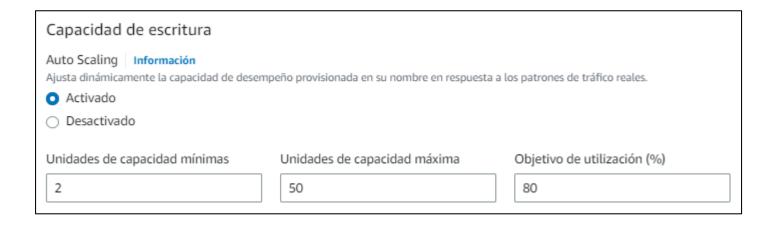


Paso 6. Luego tenemos el Auto Scaling. Si tenemos una punta de lecturas por encima de 20, nos dará un error. Tenemos que decirle que cuando supere, por ejemplo, el 70% o el 80%, 90\$ de estas 20 añada 5 unidades más, así hasta un máximo de 100 unidades.

Se incrementará este número 20 de 5 en 5, cada vez que supere el 70% de utilización hasta un máximo de 100.

Capacidad de lectura		
,	eño provisionada en su nombre en respuesta a l	os patrones de tráfico reales.
 Activado 		
○ Desactivado		
Unidades de capacidad mínimas	Unidades de capacidad máxima	Objetivo de utilización (%)
5	100	70

Paso 7. Lo mismo para el auto-escalado de la escritura. Podemos aumentarlo si supera el umbral, por ejemplo, el 80%, lo vamos aumentando de 2 en 2 hasta un máximo por ejemplo de 50.



Paso 8. A guardar esta configuración, haciendo click en el botón Guardar Cambios, se produce un error de permisos, indicando que la función autoscaling no esta dispoible

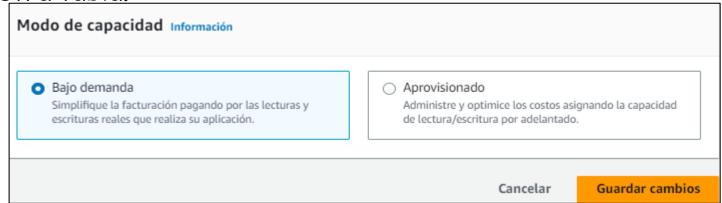
Se ha producido un error al guardar la configuración de capacidad de la tabla articulos

User: arm:aws:sts::307729659735:assumed-role/voclabs/user1396355=Eduard_Lara is not authorized to perform: application-autoscaling:RegisterScalableTarget on resource: arm:aws:application-autoscaling:us-east-1:307729659735:scalable-target/* because no identity-based policy allows the application-autoscaling:RegisterScalableTarget action

Si se hubiera podido guardar, nuestra tabla articulos, hubiera adoptado la configuración de capacidad: 20 de lectura, 4 de escritura y autoscaling. Es decir, incrementa cuando suoera el 70% de 20, subinedo de 5 a 5 unidades hasta llegar a más de 100.

NOTA: Lo importante es que podemos configurar nosotros la capacidad de la tabla en función de la carga de trabajo que podemos tener.

Paso 9. Hay otra forma de configurar la capacidad de las tablas que sería bajo demanda: No tenemos que planificar cuántas unidades o cuántas capacidades de lectura, escritura o auto-scaling. Simplemente configuramos bajo demanda, y ya no configuramos nada más, sino que Amazon va a configurarlo en función del uso real que le demos a nuestra tabla.



Se ha producido un error al guardar la configuración de capacidad de la tabla articulos

User: arn:aws:sts::307729659735:assumed-role/voclabs/user1396355=Eduard_Lara is not authorized to perform: application-autoscaling:DeregisterScalableTarget on resource: arn:aws:application-autoscaling:DeregisterScalableTarget action

Paso 10. Modo bajo demanda: Si de repente vienen 200 peticiones de lectura, se atenderán las 200 en ese momento y nos cobrará por ello. Si sólo hay 2 lecturas, pagaremos exactamente por lo que utilicemos sin tener que provisionar nosotros cual es nuestra capacidad.

Amazon se encarga de gestionar las capacidades para que siempre funcione y así atienda bien a los picos. Cuando hay poca demanda también baja las capacidades y solo nos cobra por el uso real.

Esta es la forma más cómodo a la hora de no tener ningún error Nuestra tabla en cuanto a capacidad, no tendrá problema, será flexible en función de la carga que tenga.

Esta es quizá la forma más eficiente de configurar nuestra tabla para no tener ningún tipo de problemas.

Paso 1. En este laboratorio vamos a ver que operaciones podemos realizar sobre nuestras tablas en DynamoDB

Inserción. En la tabla artículos, que teóricamente debe contender 2 elementos, podemos añadirle un nuevo elemento, pro ejemplo el usuario2 va a escribir su primer artículo, con el contenido "Este es el primer artículo del usuario2". Le damos a guardar

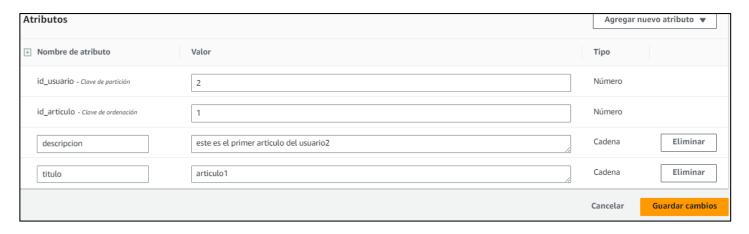
Atributos		Agregar n	uevo atributo 🔻
Nombre de atributo	Valor	Tipo	
id_usuario - Clave de partición	2	Número	
id_articulo - Clave de ordenación	1	Número	
contenido	este es el primer articulo del usuario2	Cadena	Eliminar
		Cancelar	Guardar cambios

id_usuario	▽	id_articulo	▽	contenido
2		1		este es el primer articulo del usuario2
1		1		este es mi primer articulo
1		2		este es mi segundo articulo

Paso 2. Edición o Modificación. Podemos editar el contenido y modificar el texto o agregar otro atributo



Por ejemplo hemos agregado el atributo titulo y cambiado el atributo contenido por descripción

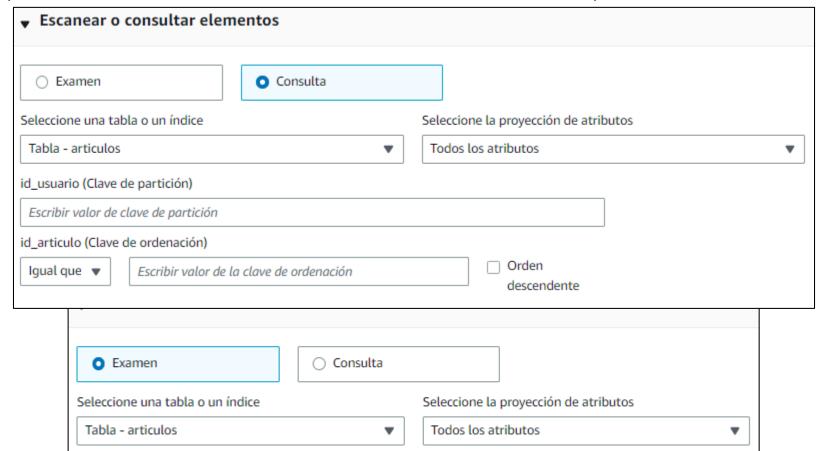


Paso 3. Edición o Modificación. Vemos que la parte del contenido esta eliminada. En cambio, aparece la descripción y el título.

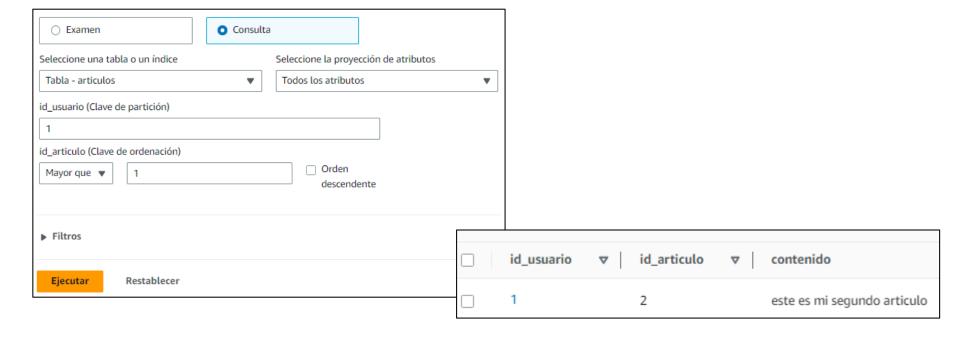
	id_usuario	▼ id_articulo	▼ c	contenido	▼	descripcion	▽	titulo
~	2	1				este es el primer articulo del usuario	2	articulo1
	1	1	e	este es mi primer articulo				
	1	2	e	este es mi segundo articulo				

54

Paso 4. Búsquedas. También podemos hacer búsquedas. Por ejemplo, podemos buscar mediante un examen o una consulta. Sí queremos buscar sin especificar un valor para la clave, utilizamos el examen y si ponemos un valor a nuestra clave, usaremos la opción de consulta.



Paso 5. Búsquedas La consulta nos permite buscar por un identificador de usuario concreto. Ponemos que el identificador sea igual a 1, donde no podemos cambiar el igual. Y en la parte de ordenación si podemos indicar igual, menor, mayor, etc Indicaremos que queremos todos los artículos del usuario1 cuyo número de artículos sea mayor 1. Sólo mostrara una unica fila. Si ejecutamos nos aparece una única fila



Paso 6. Búsquedas. Podemos modificarlo y ponerle que sea mayor o igual que uno, así también saldrá el uno, saldrán dos filas o dos elementos de identificador número uno.



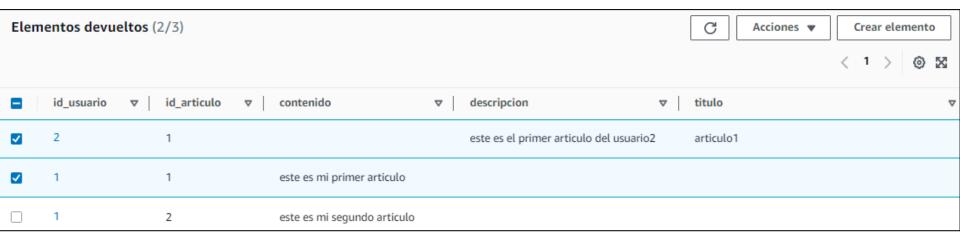
id_usuario	\triangledown	id_articulo	▽	contenido
1		1		este es mi primer articulo
1		2		este es mi segundo articulo

Paso 7. Búsquedas. Si queremos buscar por el contenido, tendríamos que usar el examen. Seleccionamos la tabla artículos y añadimos un filtro. buscamos si hay alguna fila o elemento en mi tabla cuyo contenido contenga la palabra mi.

▼ Escanear o consultar elementos					
• Examen Consulta					
Seleccione una tabla o un índice	Seleccione la proyección de atributos				
Tabla - articulos ▼	Todos los atributos ▼				
▼ Filtros Nombre de atributo Tipo Condición contenido Cadena ▼ Contiene Agregar filtro	Valor ▼ mi Eliminar				
Ejecutar Restablecer					

id_usuario	\triangledown	id_articulo	▽	contenido
1		1		este es mi primer articulo
1		2		este es mi segundo articulo

Paso 8. Borrado. También podemos borrar todos los elementos de una tabla. Vamos a examen, eliminados el filtro de búsqueda. Obteneos todos los elementos, los podemos marcar todos y darle a borrar y esto eliminará todas las filas o todos los elementos de nuestra tabla.



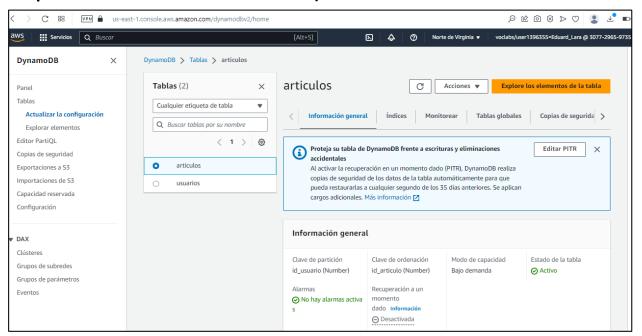
59

Paso 9. Borrado Tabla. Si hay muchos elementos en nuestra tabla, millones, y queremos borrar todos los elementos, por lo más sencillo sería borrar la tabla, marcamos la tabla que queremos borrar artículos y le vamos a eliminar tabla. Esto eliminaría toda la tabla, como todas las filas que contiene y luego podríamos volver a crearla.



Paso 1. Veremos el uso de DAX, que es una caché de memoria que reduce los tiempos de acceso a nuestras tablas.

Ahora mismo el tiempo de respuesta a cualquier acceso en la tabla artículos es del orden de milisegundos (1-9). Si queremos reducirlo a microsegundos, necesitaríamos crear una caché, que es una especie de memoria anterior a la tabla donde los usuarios acceden sin tener que ir a la propia tabla. Eso sería lo que se hace mediante DAX.



Paso 2. Vamos a hacer una cache para la tabla artículos. Vamos al Panel, y hacemos click en Clústeres:



ctusteres

Grupos de subredes

Grupos de parámetros

Eventos

Crear un clúster

Amazon DynamoDB Accelerator (DAX) es un servicio de almacenamiento de caché en memoria totalmente administrado y altamente disponible de DynamoDB.

Crear un clúster

Recursos

Actualmente está utilizando los siguientes componentes de DAX:

Clústeres (1)

Grupos de subredes del clúster (1)

Grupos de parámetros del clúster (1)

Eventos del clúster (7)

Paso 3. En la creación del clúster DAX podemos poner el nombre cluster-artículos



Paso 4. Tenemos que indicar el tipo de nodo procesador que va a usar, por ejemplo dax.t2.small. También el número de nodos, por ejemplo 2, y habilitamos el cifrado



Paso 5. Tenemos que indicar un rol para poder acceder a esa tabla en concreto. Creamos un nuevo rol y le ponemos un nombre, así como a la política de acceso de lectura y escritura. Por ultimo indicamos que la tabla destino será la de artículos. Aquí es donde queremos crear la caché para obtener tiempos de microsegundos.

Rol de servicio de IAM para el acceso a DynamoDB	Crear nueva Este rol de servicio de IAM y la política adjunta especifican las tablas de DynamoDB a las que el clúster de DAX tiene acceso y las API de DynamoD que el clúster de DAX puede ejecutar. Para obtener más información, consulte Control de acceso de DAX. Asimismo, para crear el rol de servicio de IAM, consulte Creación de un clúster de DAX.	
Nombre de rol de IAM Nombre de la política de IAM		0
Política del rol de IAM		0
Tabla de DynamoDB de destino	articulos ▼ Mediante la consola de IAM, añada o elimine tablas de DynamoDB adicionales a esta política.	0

Paso 6. El resto lo dejamos por defecto. Y le damos al botón lanzar el clúster

Cancelar

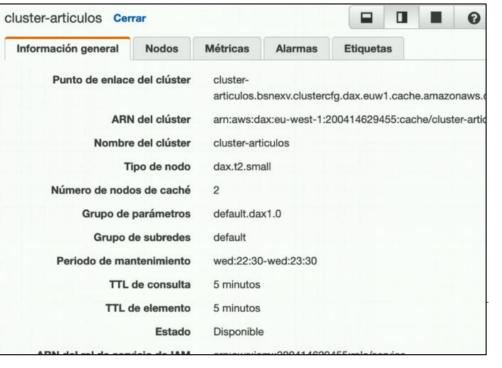
Lanzar el clúster

Esto va a crear una caché por delante de nuestra tabla de artículos donde los usuarios podrán hacer consultas con tiempos de microsegundos en lugar de milisegundos.

Paso 7. Después de esperar un cierto tiempo, ha conseguido finalizar nuestro clúster de artículos de la tabla de artículos



- Paso 8. Si pulsamos en el enlace vemos las características generales del clúster de artículos:
- la puerta de enlace del cluster a donde tenemos que apuntar
- el nombre, el tipo de procesador, el numero de nodos de la cache En la pestaña nodos observamos los nodos del cluster





Paso 9. También podemos ver las métricas, las alarmas y las etiquetas del clúster DAX de la caché de la tabla artículos.

Tendrá tiempos de acceso del orden de microsegundos en lugar de los tiempos de acceso del orden de milisegundos que se tendrán en la propia tabla.





Paso 1. Vamos a hacer ahora un laboratorio para capturar la actividad de la tabla artículos con flujos de DynamoDB, y poder ver los logs de las operaciones que se hace en nuestra tabla en CloudWatch.

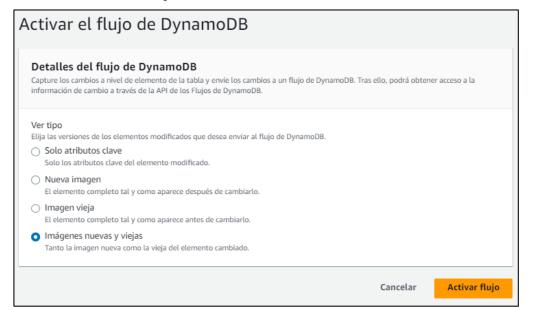
Pulsamos en la tabla artículos. En la pestaña Información general podemos ver que el Flujo de DynamoDB esta desactivado

articulos		C Acciones	Explore los elementos de la tabla
Información general Índices M	onitorear Tablas globales Copias d	le seguridad Exportaciones y flujos (Configuración adicional
Información general			
Clave de partición id_usuario (Number) Alarmas No hay alarmas activas	Clave de ordenación id_articulo (Number) Recuperación a un momento dado Información O Desactivada	Modo de capacidad Bajo demanda	Estado de la tabla → Activo
▼ Información adicional	L		
Clase de tabla Estándar de DynamoDB	Índices O globales, O locales	Flujo de DynamoDB O Desactivada	Tiempo de vida (TTL) Información

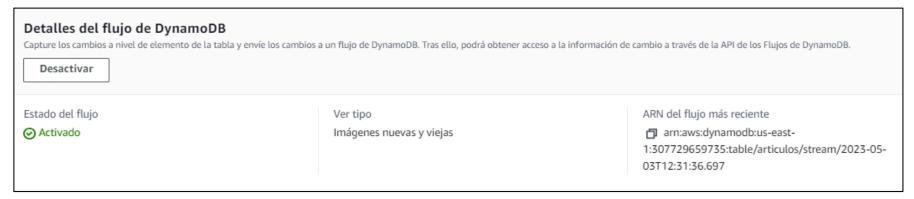
Paso 2. Vamos a la pestaña Exportaciones y Flujos y en el apartado Detalles del flujo de DynamoDB, hacemos click en el botón Activar

Detalles del flujo de DynamoDB Capture los cambios a nivel de elemento de la tabla y envíe los cambios a un flujo de DynamoDB. Tras ello, podrá obtener acceso a la información de cambio a través de la API de los Flujos de DynamoDB. Activar
Estado del flujo Desactivado

Paso 3. Seleccionamos la última opción Imágenes nuevas y viejas, y le damos al botón Activar flujo.



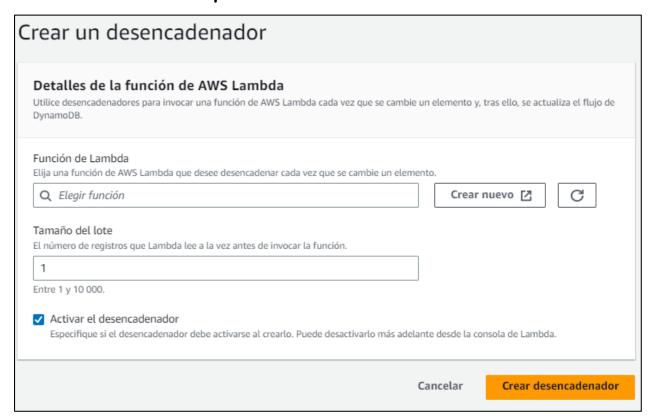
Paso 4. Una vez que hemos activado el flujo, vemos en Detalles del flujo de DynamoDB que esta activado, el tipo de Imágenes nuevas y viejas y ha puesto la url del flujo



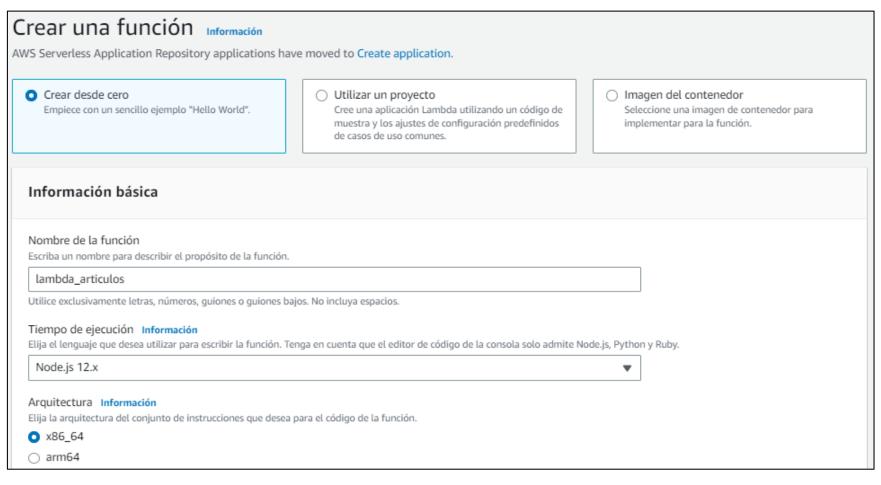
Paso 5. Ahora vamos a Desencadenadores



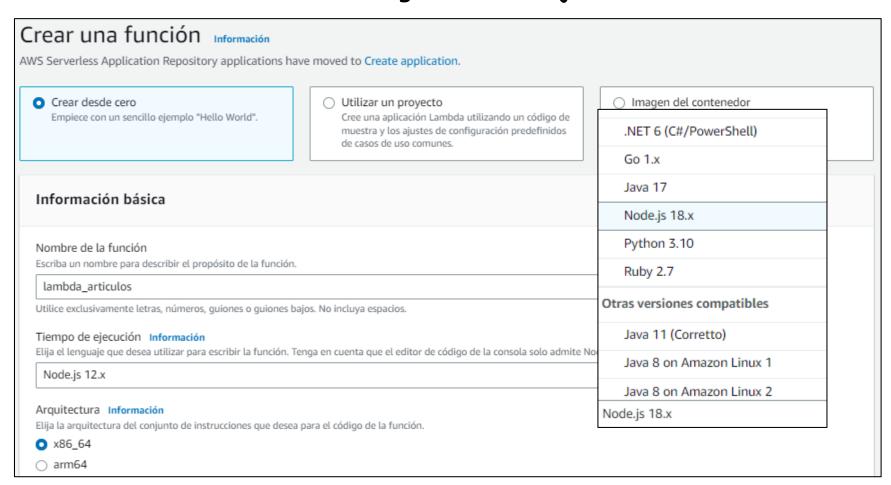
Paso 6. Un desencadenador funciona como un trigger: Cada vez que se crea, modifica, o borra un elemento de la tabla artículos, ejecutará el código de una determinado función lambda. Queremos que este código grabe la información del registro modificado en los logs para que puedan ser consultados por CloudWatch



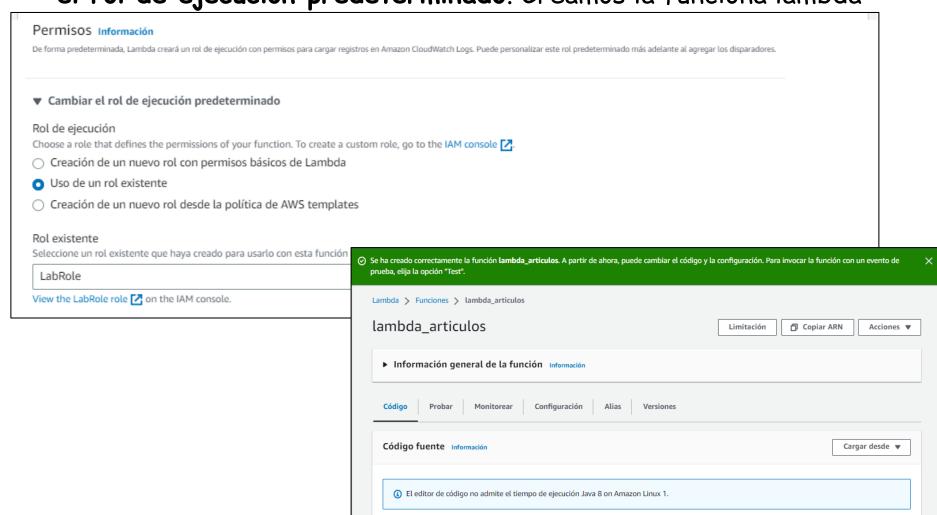
Paso 7. Debemos crear una nueva función lambda que asociaremos al desencadenador de la tabla articulos. Pondremos el nombre de lambda_artículos.



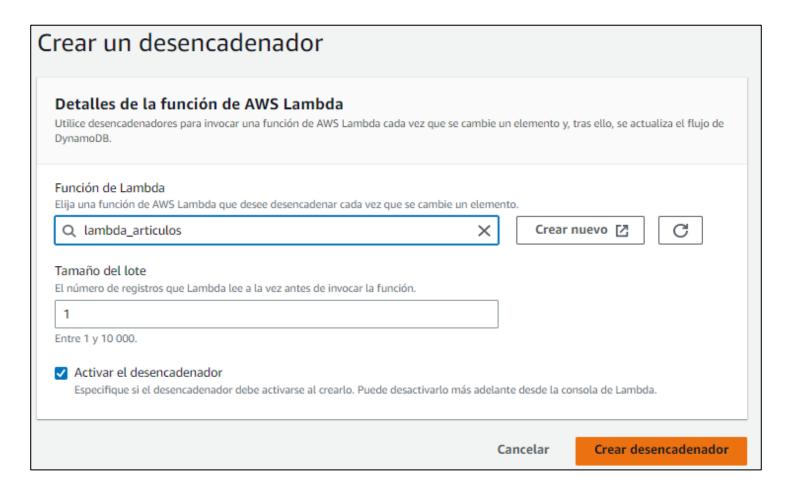
Paso 8. Las funciones lambda admiten diferentes lenguajes de programación: Java, Node.js, Python y Ruby. En tiempo de ejecución indicaremos la versión más antigua de Node.js 12x



Paso 9. Usaremos los permisos del rol existente LabRole en Cambiar el rol de ejecución predeterminado. Creamos la funciona lambda



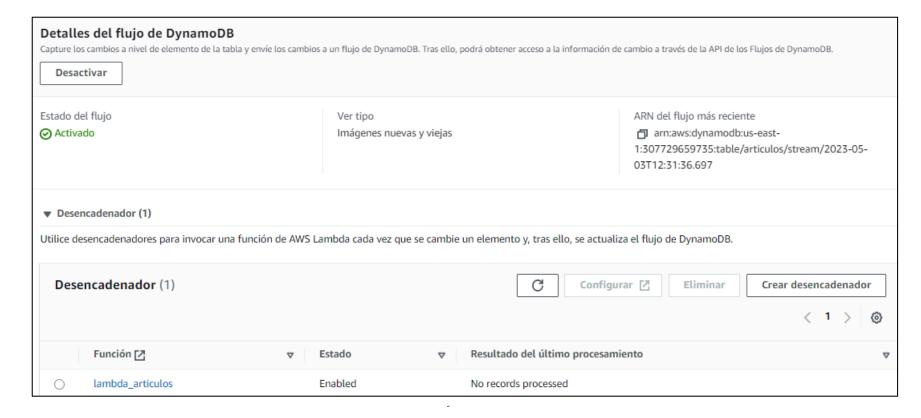
Paso 10. En la pagina de creación del Desencadenador de la tabla artículos, buscamos la función lambda_articulos recién creada, y hacemos click en el botón Crear Desencadenador



Paso 11. Vemos que esta función Lambda ya tiene un desencadenador que va a ser el dynamoDB. Cada vez que se actualice una información en la tabla artículos se ejecutará la función Lambda, que lo que hace es grabar logs con la información de lo que ha ocurrido y esta información la podemos ver automáticamente en CloudWatch

lambda_articulos						L
▼ Información general de	la fui	nción _{Infor}	mación			
	<u></u>	lambda_a	articulo	OS (0)		
DynamoDB					+ Agregar d	lestino
+ Agregar desencadenador						

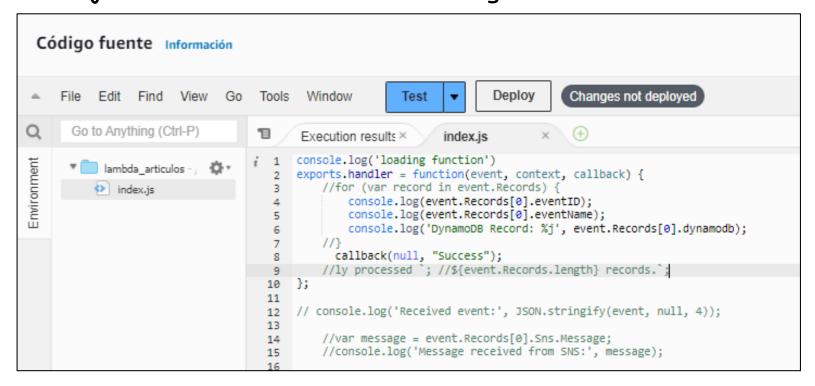
Paso 12. Si vamos a la tabla artículos en la pestaña Exportaciones y flujos, en el apartado Detalles del flujo DynamoDB vemos que esta definido el desencadenador con la función lambda creada.



Paso 13. Debemos poner el siguiente código nodejs dentro de la función lambda que se ejecutara cuando el desencadenador se lo indique. Simplemente hace que para cada uno de los eventos que se producen en la tabla articulos, va a grabar en el log de información, el tipo de evento y el registro que se ha modificado, para que pueda ser revisada en CloudWatch. El problema es que esta versión de nodejs no esta soportada actualmente

Tiempo de ejecución			Controlador Inf		
	Node.js 8.10	•	index.handler		
Window					
index.js	× ⊕				
console.l	og('Loading function');				
//con for (nandler = async (event, contensole.log('Received event:', (const record of event.Record phsole.log(record.eventID); console.log(record.eventName); console.log('DynamoDB Record en 'Successfully processed \$-	JSON.stringify(evds) {); ; %j', record.dyna	amodb);		

Paso 14. Se debe de encontrar el código correcto para la versión de nodejs 12.x o superior. Para ello, se pueden hacer pruebas para ver el correcto funcionamiento de la función lambda. El botón Deploy sirve para que guarde y actualice el nuevo código. Y con el botón Test se puede probar la función lambda y ver si no da errores. En el código de mas abajo es un intento de escribir el log con la versión 12.x



Paso 15. Vamos a Explore elementos de la tabla y vamos a introducir un nuevo elemento. Esta es la información que tenemos actualmente.

Elem	entos devu	eltos (3)					
	id_usuario	▼ id_articulo	▼ contenido	▽	descripcion	▼	titulo
	2	1			este es el primer articulo de	e	articulo1
	1	1	este es mi primer art	ticulo			
	1	2	este es mi segundo a	articulo			

81

Paso 16. Añadiremos un nuevo elemento: el usuario 4 va a crear su primer artículo y le añadimos un campo de texto de cadena contenido

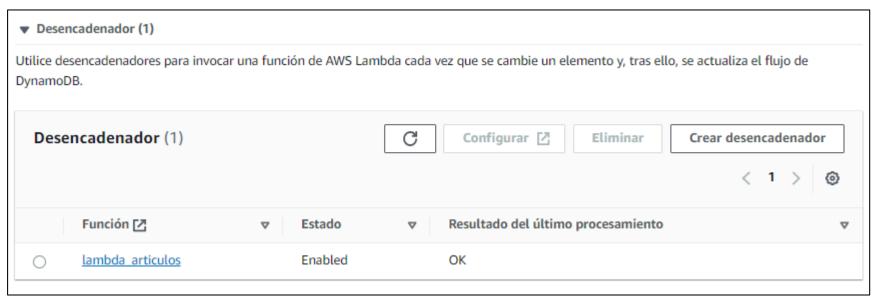


Paso 17. Guardamos y comprobamos el resultado: El usuario 4 acaba de crear un artículo que se llama "Articulo 1 del usuario 4,

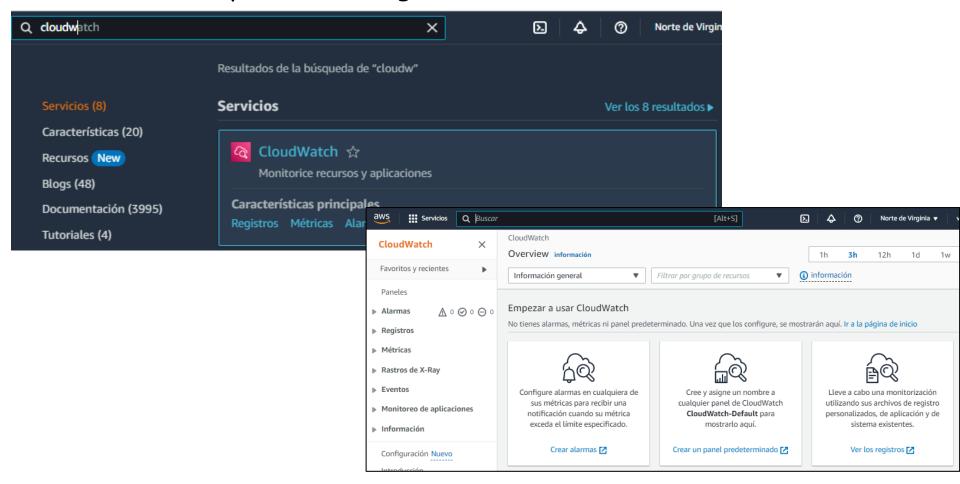


Paso 18. Si vamos a desencadenadores y le damos a actualizar la página veremos que el desencadenador lambda_artículos, la función lambda se ha activado pero ha fallado.

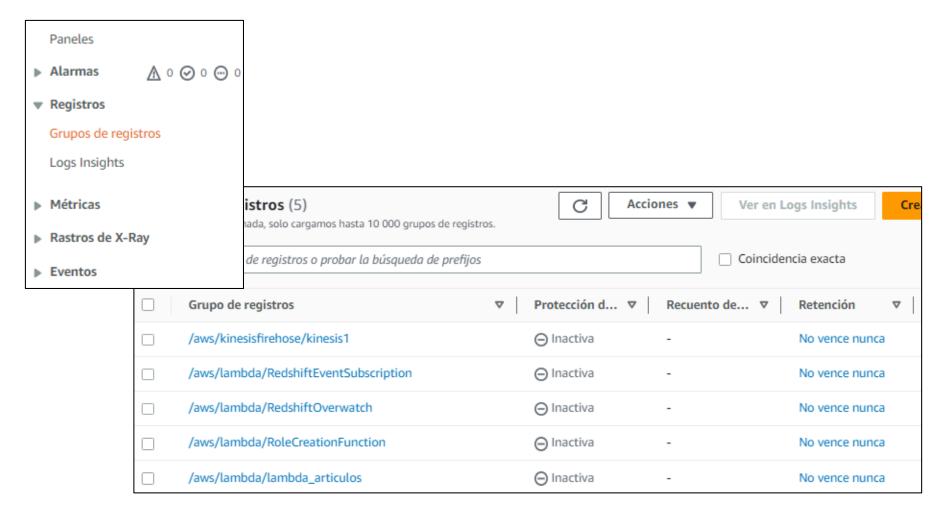
Como hemos añadido un elemento nuevo se ha activado el desencadenador, y se ha ejecutado la función Lambda y la función Lambda graba la información de log



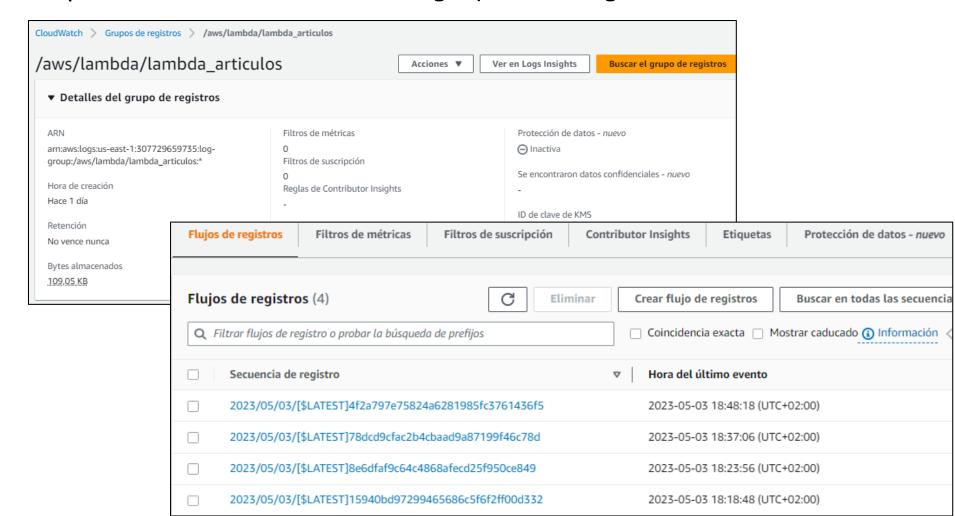
Paso 19. Podemos ir CloudWatch. Hacmeos click botón derecho sobre el logo de AWS y abrimos una nueva pestaña. Buscamos y vamos a CloudWatch para ver el registro de la actividad de la tabla.



Paso 20. Vamos al menú Registros/Grupos de registros y aquí encontramos nuestra función AWS lambda_artículos



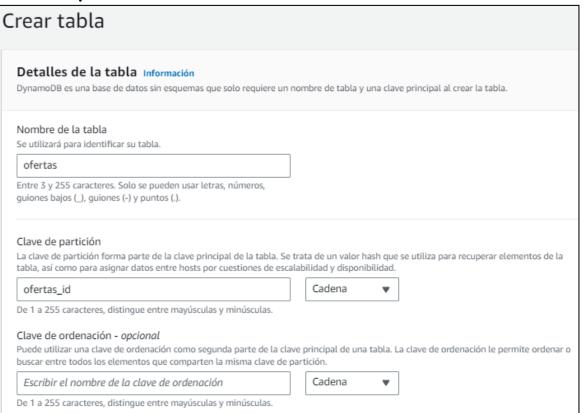
Paso 21. Si pulsamos sobre el enlace de nuestra función lambda, podemos tener acceso a los logs que hemos generado.



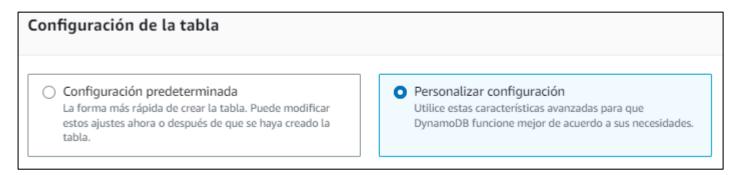
Paso 22. Si pulsamos en el ultimo log, podemos ver el insert realizado: Usuario 4, articulo 1, descripción: Articulo 1 del usuario 4. Hemos capturado la información de una acción realizada con la tabla artículos, lo hemos grabado y lo hemos podido ver con CloudWatch

```
2023-05-03T19:55:55.325+02:00
                                        2023-05-03T17:55:55.325Z ba219a14-7f09-4447-920a-b01d4451485d INFO INSERT
2023-05-03T19:55:55.325+02:00
                                        2023-05-03T17:55:55.325Z ba219a14-7f09-4447-920a-b01d4451485d INFO DynamoDB Record: {"ApproximateCreationDateTime":168.
2023-05-03T17:55:55.325Z
                                ba219a14-7f09-4447-920a-b01d4451485d
                                                                                 DynamoDB Record:
                                                                                                                                                          Copiar
    "ApproximateCreationDateTime": 1683136554,
    "Keys": {
        "id usuario": {
            "N": "4"
        "id articulo": {
            "N": "1"
    "NewImage": {
        "contenido": {
            "S": "Articulo 1 del usuario 4"
        "id usuario": {
         'id articulo": {
            "N": "1"
                                                                                                                                                     Volver arriba
    "SequenceNumber": "29284200000000047291691237",
```

Paso 1. Vamos a hacer ahora un laboratorio para ver el TTL el tiempo de vida que nos permite definir cuándo vencen los elementos de una tabla para eliminarlos automáticamente de la base de datos. Para ellos crearemos una nueva tabla, llamada ofertas donde cada oferta tendrá un tiempo de expiración. El identificador de la tabla es ofertas_id.



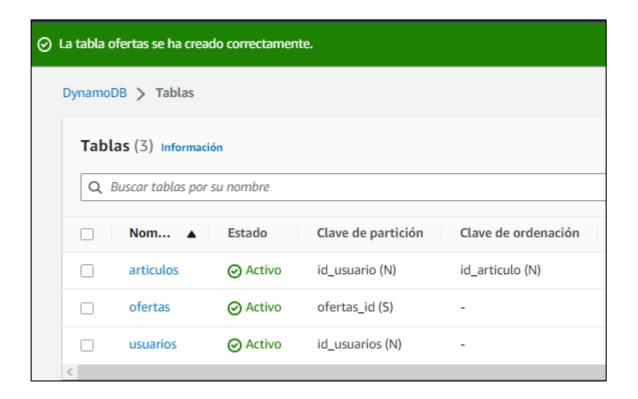
Paso 2. Deseleccionamos configuración predeterminada



Paso 3. Ponemos la opción Bajo demanda, indicando que la gestión de la facturación de la tabla la va a llevar de forma flexible Amazon

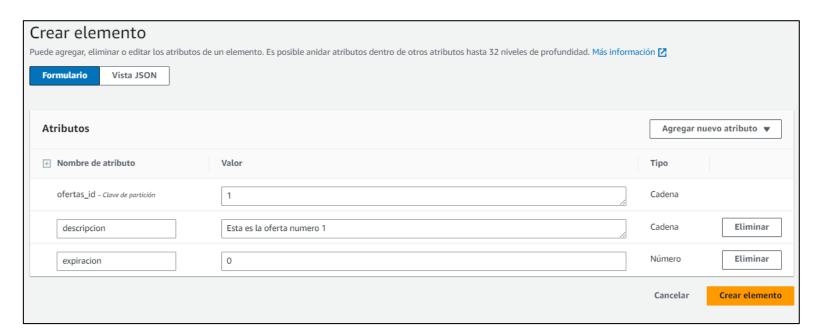


Paso 4. Dejamos los demás parámetros por defecto y le damos a crear tabla. Esto nos creará nuestra tabla de ofertas.



Paso 5. Añadimos varios elementos, a los que pondremos un tiempo de expiración, para una vez pasado este tiempo, desaparezca de la tabla. Creamos el primer elemento:

- □ ofertas_id=1
- ☐ descripción (cadena)=esta es la oferta número 1
- □ expiracion (number)= {{hay que ponerle un valor}}



Paso 6. Este valor de expiración lo sacaremos por ejemplo de la pagina web https://www.epochconverter.com. Esta pagina indica el tiempo en Unix o epoch, es decir, la cantidad de segundos transcurridos desde la medianoche del 1 de enero de 1970 hasta la fecha actual en la que estamos ahora



A partir de la fecha y hora actuales, nos permite consultar cual es el número de segundos o epoch desde el 1 de enero de 1970

Paso 7. Si queremos que algo caduque dentro de una hora, incrementamos en una hora la hora actual y hacemos click en el botón Human date to timestamp, para obtener el valor en segundos o epoch de esa hora (por ejemplo, una hora más que la hora actual).

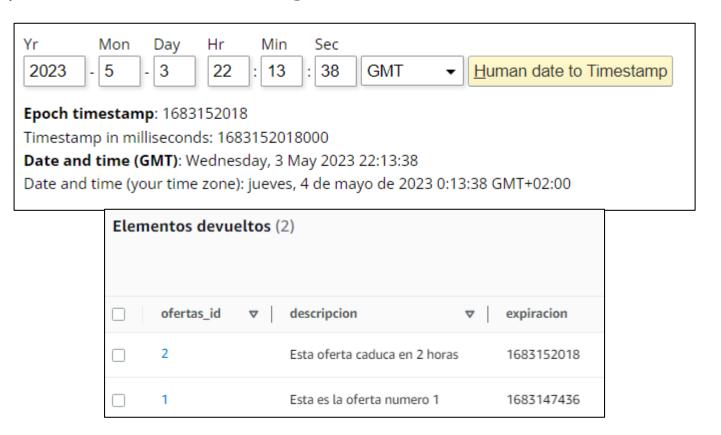


Paso 8. Copiamos este valor y lo ponemos en el primer registro que estamos introduciendo en la tabla. Estamos indicando que esta primera fila tendrá una expiración dentro de una hora.

Atributos	
Nombre de atributo	Valor
ofertas_id - Clave de partición	1
descripcion	Esta es la oferta numero 1
expiracion	1683147436

Elementos devueltos (1)					
	ofertas_id	▽	descripcion	expiracion	
	1		Esta es la ofer	1683147436	

Paso 9. Añadiremos otro elemento. Indicaremos en la descripción que esta oferta caduca en dos horas. Calcularemos ese valor de epoch (de dos horas mas respecto la hora actual) y lo insertamos en su campo de expiración. Finalmente lo guardamos



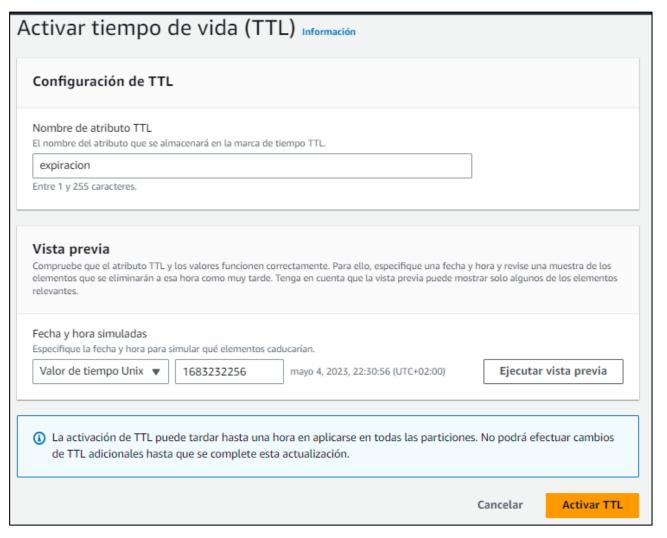
Paso 10. Ahora tenemos dos ofertas, cada una con un tiempo de expiración diferente. Tenemos que indicar a la tabla ofertas, cuál es la columna que va a hacer de TTL o de tiempo de expiración, en este caso la columna expiración

Vamos a Configuración Adicional y activamos el tiempo de vida (TTL) haciendo click en botón Activar

Tiempo de vida (TTL) Información Eliminar automáticamente los elementos caducados de una tabla.	Ejecutar vista previa Activar
Estado TTL O Desactivada	

Paso 12. Indicamos en el Nombre de atributo TTL, la columna

expiración.

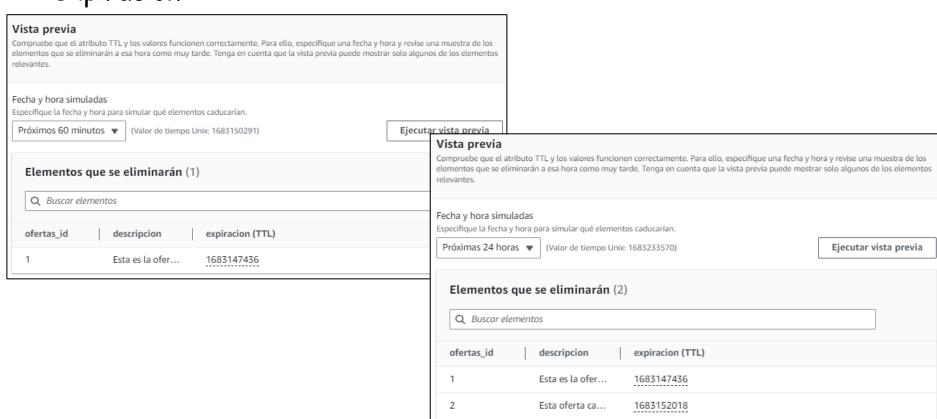


Paso 13. En Ejecutar vista prueba podemos hacer pruebas o simulaciones acerca de como quedaría la tabla en un momento determinado debido a la acción de la expiración.

Si miramos cuantos han caducado ahora mismo, veremos que ninguno (para que se elimine alguno debe de pasar como minimo una hora)



Paso 14. Si simulamos que pasará dentro de una hora y 10 minutos, debería salir la primera fila introducida como elemento eliminado, ya que pusimos un tiempo de una hora. Si ponemos 24 horas, saldrían los dos elementos puesto que tienen una hora y dos horas como tiempo de expiracion



Paso 15. Finalmente activamos el tiempo de vida de la tabla.



De esta manera ya está identificado la columna expiración como atributo TTL de tiempo de vida. Cuando el tiempo actual llegue a los valores de expiración, estos elementos se borrarán automáticamente de la tabla.

Esto es una manera de hacer limpieza en una tabla de forma automática. Por ejemplo, en una tabla de ofertas podemos programar que una oferta se elimine cuando dicha oferta haya caducado