

Servicio Products - testing de escalabilidad

Propósito:

El objetivo de este testing es probar la capacidad de escalado horizontal del servicio Products a la hora de recibir una cantidad de peticiones para así saber si el mismo puede manejar cargas aumentadas de información.

Contenido:

- **Objetivos de la prueba:** El propósito de la prueba de escalabilidad es validar que el servicio sea capaz de escalar horizontalmente al recibir una mayor cantidad de peticiones de lo normal.
- **Resultado Clave:** El servicio es capaz de escalar horizontalmente acorde a la demanda de peticiones por los usuarios según la configuración aplicada, sin verse afectado el servicio. Posteriormente a la caída de la demanda, el servicio des escala correctamente.

Entorno de Pruebas

Se realiza las pruebas en instancias con CPU de 256 unidades y 512 MegaBytes de memoria RAM.

Contenido:

- **Descripción del Entorno:** Los cluster donde trabajan las instancias son de tipo "FRAGATE", las configuraciones de red están completamente abiertas tanto la entrada como salida. Cuenta con un balanceador de carga.
- La configuración del auto escalado sea realiza teniendo 1 tarea como mínimo y un máximo de 3 tareas escaladas. La política de escalado es de tipo seguimiento de destino, con una métrica de peticiones con un valor de 10000 peticiones y perdidos de recuperación de escalado y desescalada horizontal de 30 segundos.
- Se utiliza la herramienta JMeter (V 5.6.3) para realizar las pruebas de carga.

▼ Escalado automático de servicios - *opcional*

Ajuste automáticamente el recuento deseado del servicio hacia arriba y hacia abajo dentro de un rango especificado en respuesta a las alarmas de CloudWatch. Puede modificar la configuración del escalado automático de servicios en cualquier momento para satisfacer las necesidades de la aplicación.

☒ Utilizar el escalado automático del servicio

Configurar el escalado automático de servicios para ajustar el recuento deseado del servicio

Cantidad mínima de tareas

El límite inferior al que el escalado automático de servicios puede ajustar el recuento deseado del servicio.

Cantidad máxima de tareas

El límite superior al que el escalado automático de servicios puede ajustar el recuento deseado del servicio.

Política de escalado

Eliminar

Tipo de política de escalado

Seguimiento de destino

Nombre de política

Métrica de servicio de ECS

Valor de destino

Periodo de recuperación de escalado horizontal

Periodo de recuperación de desescalado horizontal

☐ Desactivar la acción de desescalar horizontalmente

Plan de Pruebas

- Escenario de prueba: 3000 usuarios simultáneos en un tiempo de Ramp up de 30 segundos durante 10 minutos.

Thread Group

Name: Products

Comments:

Action to be taken after a Sampler error

☒ Continue ☐ Start Next Thread Loop ☐ Stop Thread ☐ Stop Test ☐ Stop Test Now

Thread Properties

Number of Threads (users): 3000

Ramp-up period (seconds): 30

Loop Count: ☒ Infinite

☒ Same user on each iteration

☐ Delay Thread creation until needed

☒ Specify Thread lifetime

Duration (seconds): 600

Startup delay (seconds):

HTTP Request

Name: HTTP Request

Comments:

Basic

Web Server

Protocol (http): http

Server Name or IP: dev-products-lb-35790505.us-east-1.elb.amazonaws.com

Port Number: 80

HTTP Request

GET

Path: products

Content encoding:

☐ Redirect Automatically ☒ Follow Redirects ☒ Use KeepAlive ☐ Use multipart/form-data ☐ Browser-compatible headers

Parameters

Body Data

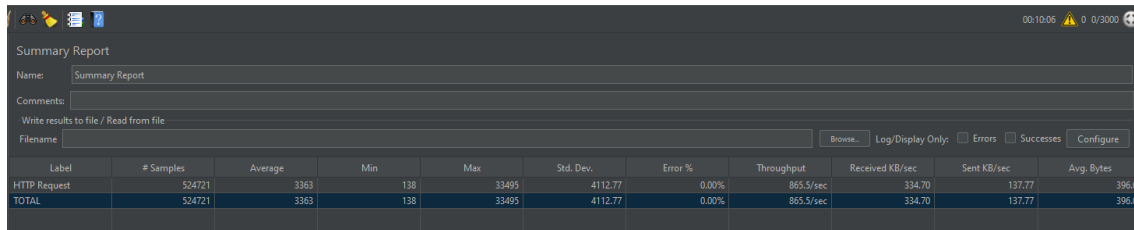
Files Upload

Send Parameters With the Request:

Name	Value	URL Encode?	Content-Type	Include Equals?
------	-------	-------------	--------------	-----------------

Resultado de la prueba

- Se observa en el reporte de Jmeter el tiempo transcurrido de 10 minutos y un envío total de 524 mil peticiones aproximadamente con un índice de éxito del 100%.



Summary Report

Name: Summary Report

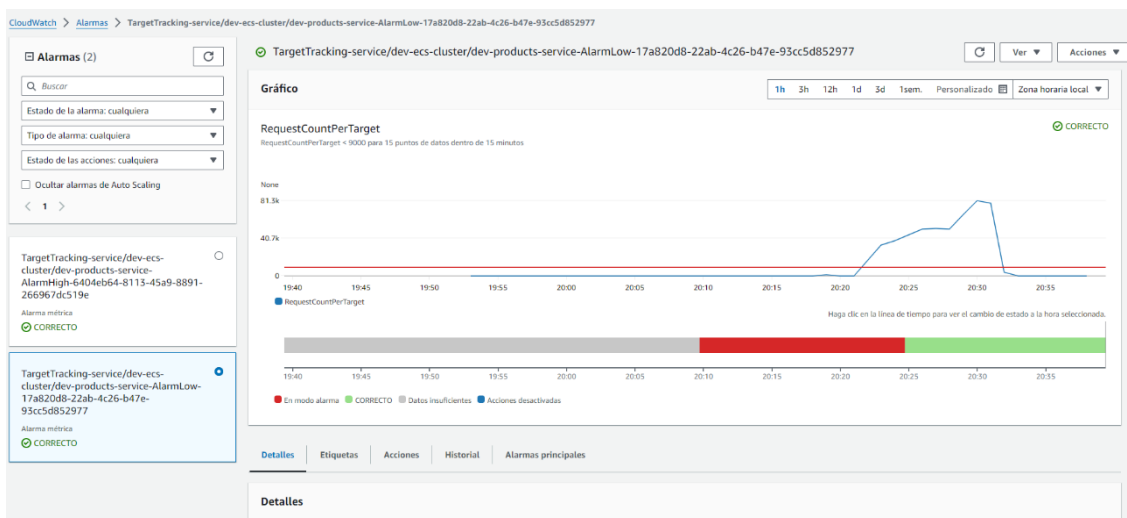
Comments:

Write results to file / Read from file

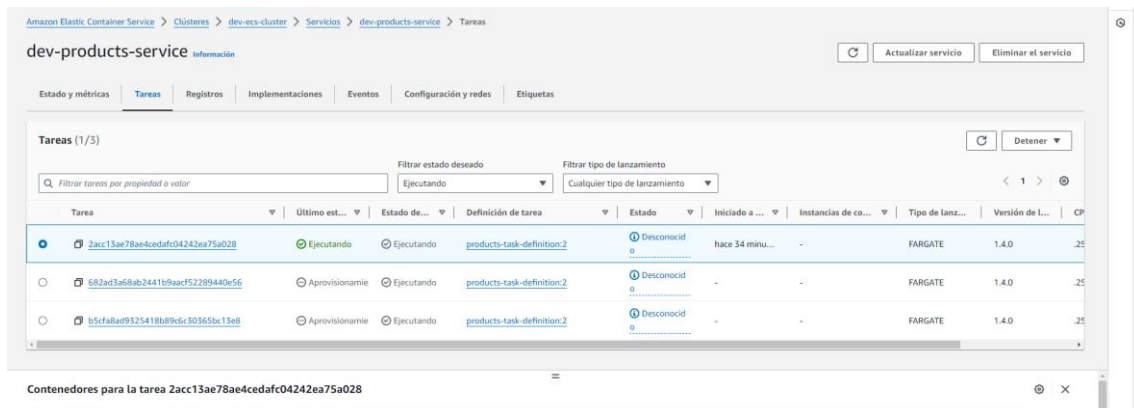
Filename: Browse... Log/Display Only: ☐ Errors ☐ Successes ☐ Configure

Label	# Samples	Average	Min	Max	Std. Dev.	Error %	Throughput	Received KB/sec	Sent KB/sec	Avg. Bytes
HTTP Request	524721	3363	138	33495	4112.77	0.00%	865.5/sec	334.70	137.77	396.0
TOTAL	524721	3363	138	33495	4112.77	0.00%	865.5/sec	334.70	137.77	396.0

- En cloud watch se observan picos de 81.3 mil peticiones posterior al escalado.



- Se puede observar como de una instancia se pasa a tres instancias en ejecución paralela, indicando el correcto escalamiento según la demanda que se tenía y la configuración.



Amazon Elastic Container Service > Clusters > dev-ecs-cluster > Servicios > dev-products-service > Tareas

dev-products-service Información

Estado y métricas Tareas Registros Implementaciones Eventos Configuración y redes Etiquetas

Tareas (1/3)

Filtrar estado deseado Ejecutando Filtrar tipo de lanzamiento Cualquier tipo de lanzamiento

Tarea	Último est...	Estado de...	Definición de tarea	Estado	Iniciado a ...	Instancias de co...	Tipo de lanz...	Versión de l...	CP
2acc13ae78ae4cedafc04242ea75a028	Ejecutando	Ejecutando	products-task-definition:2	Desconocido	hace 34 minu...	-	FARGATE	1.4.0	25
682ad3a68ab2441b9aacf52289440e56	Aprovisionamie	Ejecutando	products-task-definition:2	Desconocido	-	-	FARGATE	1.4.0	25
b5cf8a8e9325418b896c30365bc15e8	Aprovisionamie	Ejecutando	products-task-definition:2	Desconocido	-	-	FARGATE	1.4.0	25

Contenedores para la tarea 2acc13ae78ae4cedafc04242ea75a028

Amazon Elastic Container Service > Clústeres > dev-ecs-cluster > Servicios > dev-products-service > Tareas

dev-products-service **Información** Actualizar servicio Eliminar el servicio

Estado y métricas | **Tareas** | Registros | Implementaciones | Eventos | Configuración y redes | Etiquetas

Tareas (1/3)

Filtrar estado deseado: Ejecutando | Filtrar tipo de lanzamiento: Cualquier tipo de lanzamiento

Filtrar tareas por propiedad o valor

Tarea	Último est...	Estado de...	Definición de tarea	Estado	Iniciado a ...	Instancias de co...	Tipo de lanz...	Versión de L...	CPU
2acc13ae78ae4cedaf04242ea75a028	Ejecutando	Ejecutando	products-task-definition:2	Desconocid	hace 46 minu...	-	FARGATE	1.4.0	.25 v
682ad3a68ab2441b9aacf52289440e56	Ejecutando	Ejecutando	products-task-definition:2	Desconocid	hace 12 minu...	-	FARGATE	1.4.0	.25 v
b5cfa8a99125418b89c6c30365bc13e8	Ejecutando	Ejecutando	products-task-definition:2	Desconocid	hace 12 minu...	-	FARGATE	1.4.0	.25 v

Desescalado posterior a las pruebas.

Amazon Elastic Container Service > Clústeres > dev-ecs-cluster > Servicios > dev-products-service > Tareas

dev-products-service **Información** Actualizar servicio Eliminar el servicio

Estado y métricas | **Tareas** | Registros | Implementaciones | Eventos | Configuración y redes | Etiquetas

Tareas (1/1)

Filtrar estado deseado: Ejecutando | Filtrar tipo de lanzamiento: Cualquier tipo de lanzamiento

Filtrar tareas por propiedad o valor

Tarea	Último est...	Estado de...	Definición de tarea	Estado	Iniciado a ...	Instancias de co...	Tipo de lanz...	Versión de L...	CPU
2acc13ae78ae4cedaf04242ea75a028	Ejecutando	Ejecutando	products-task-definition:2	Desconocid	hace 58 minu...	-	FARGATE	1.4.0	.25 v

Análisis y conclusión de los resultados.

Según se muestra en el análisis proporcionado el servicio se comporto acorde a lo esperado y realizo el debido escalamiento de este, logrando asi picos de 81.3 mil peticiones a la vez, posterior al escalado de los servicios y aun asi mantener un 100% de éxito en estas y sin degradación en el servicio proporcionado.