

Servicio Payment - testing de escalabilidad

Propósito:

El objetivo de este testing es probar la capacidad de escalado horizontal del servicio Payment 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 2000 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.

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.

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

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

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

3

3

Eliminar

Seguimiento de destino

Auto-Scaling-Policy-dev-payments-service

ALBRequestCountPerTarget



2000

30

30

☐ Desactivar la acción de desescalar horizontalmente

- Escenario de prueba: 3000 usuarios simultáneos en un tiempo de rump up de 30 segundos durante 10 minutos.
- Se realiza la prueba a través del servicio de orders debido a que el servicio Payments no tiene un acceso público, y al mismo solo se accede a través de ese servicio.

00:10:29  0 0/3000 

Thread Group

Name:

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):

Ramp-up period (seconds):

Loop Count: ☒ Infinite

☒ Same user on each iteration


☐ Delay Thread creation until needed


☒ Specify Thread lifetime


Duration (seconds):


Startup delay (seconds):


Pruebas de Escalabilidad - Orders

 Grupo de Hilos

 Selected HTTP POST

 HTTP Header Manager

 View Results Tree

 Summary Report

HTTP Request

Name:

Comments:

Basic Advanced

Web Server

Protocol (Req): Server Name or IP: Port Number:

HTTP Request

Path: Content encoding:

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

Parameters **Body Data** Files Upload

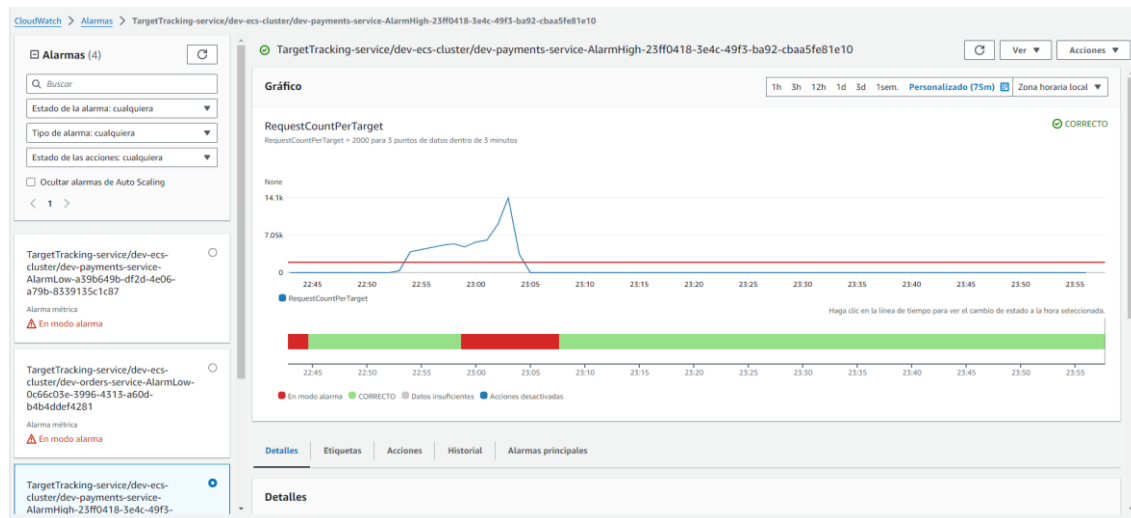
1

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	Aug. Bytes	
Selección HTTP POST	86168	20805	138	46249	13577.98	0.08%	137.0/sec	32.83	34.63	239.4	
TOTAL	86168	20805	138	46249	13577.98	0.08%	137.0/sec	32.83	34.63	239.4	

- En cloud watch se observan picos de 14.1 mil peticiones posteriores 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 > Clústeres > dev-ecs-cluster > Servicios > dev-payments-service > Tareas

dev-payments-service

Información

Actualizar servicio

Eliminar el servicio

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

Tareas (1/1)

Filtrar estado deseado

Filtrar tipo de lanzamiento

Filtrar tareas por propiedad o valor

Ejecutando

Cualquier tipo de lanzamiento

< 1 >

Tarea	Último est...	Estado de...	Definición de tarea	Estado	Iniciado a ...	Instancias de co...	Tipo de lanz...	Versión de L...	CPU
<div><div></div><div>4b0512c042ed44c8b41ec68b5a370b57</div></div>	<div><div></div><div>Ejecutando</div></div>	<div><div></div><div>Ejecutando</div></div>	<div><div></div><div>dev-payments-task-definition:1</div></div>	<div><div></div><div>Desconocid</div></div>	<div><div></div><div>hace 3 horas</div></div>	<div><div></div><div>-</div></div>	<div><div></div><div>FARGATE</div></div>	<div><div></div><div>1.4.0</div></div>	<div><div></div><div>.25 vt</div></div>

dev-payments-service

Información

Actualizar servicio

Eliminar el servicio

Estado y métricas

Tareas

Registros

Implementaciones

Eventos

Configuración y redes

Etiquetas

Tareas (1/3)

Filtrar estado deseado

Filtrar tipo de lanzamiento

Filtrar tareas por propiedad o valor

Ejecutando

Cualquier tipo de lanzamiento

<

1

>

🔍

Tarea	Último est...	Estado de...	Definición de tarea	Estado	Iniciado a ...	Instancias de co...	Tipo de lanz...	Versión de L...	CPU
<div>🔄</div> <div>36d047bf301e475487affd514e27376d</div>	<div>🟢</div> <div>Ejecutando</div>	<div>🔄</div> <div>Ejecutando</div>	dev-payments-task-definition:1	<div>🔍</div> <div>Desconocid</div>	hace 44 seg...	-	FARGATE	1.4.0	.25 v
<div>🔄</div> <div>5b9ae0c2b5b34515ae75cf01909e340b</div>	<div>🟢</div> <div>Ejecutando</div>	<div>🔄</div> <div>Ejecutando</div>	dev-payments-task-definition:1	<div>🔍</div> <div>Desconocid</div>	hace 45 seg...	-	FARGATE	1.4.0	.25 v
<div>🔵</div> <div>c75cc96044d94035b73dd1d808aaae0</div>	<div>🟢</div> <div>Ejecutando</div>	<div>🔄</div> <div>Ejecutando</div>	dev-payments-task-definition:1	<div>🔍</div> <div>Desconocid</div>	hace 1 hora	-	FARGATE	1.4.0	.25 v

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

dev-payments-service Información

Actualizar servicio
Eliminar el servicio

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

Tareas (1/1) Detener

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...	CPU
c75cc9604494035b73dd1d808aaae0	Ejecutando	Ejecutando	dev-payments-task-definition:1	Desconocido	hace 1 hora	-	FARGATE	1.4.0	25 v...

Según se muestra en el análisis proporcionado el servicio se comportó acorde a lo esperado y realizó el debido escalamiento de este, logrando así picos de 14.1 mil peticiones a la vez, posterior al escalado de los servicios y aun así mantener un 100% de éxito en estas y sin degradación en el servicio proporcionado.