

Pruebas de carga y rendimiento con Docker y Kubernetes:

Análisis del performance de un microservicio, basado en el despliegue

Muñoz Aguilar Andrés Felipe - 2210087

Osma Llanes Marisol - 2211466

Introducción

En la era digital actual, el desarrollo de aplicaciones de software distribuido que sean eficientes, escalables y robustas se ha convertido en una necesidad primordial para muchas organizaciones. En este sentido, se pretende abordar un análisis analítico de estos desafíos mediante el estudio de microservicios en, un enfoque arquitectónico que facilita la creación de aplicaciones con alta disponibilidad, rendimiento y calidad de servicio.

El presente trabajo se centra en el despliegue de la aplicación "Chaphe-Bookstore-app" en distintas etapas, utilizando Docker Desktop y Kubernetes. Se exploran varios escenarios de despliegue para evaluar y comparar el rendimiento y la calidad del servicio proporcionado por la aplicación bajo diferentes configuraciones.

Metodología

Despliegue en Docker Desktop

Implementación de la aplicación "Chaphe-Bookstore-app" utilizando Docker Desktop, lo que implica la contenerización de los componentes del backend y frontend para evaluar su rendimiento en un entorno de contenedores.

Despliegue en Kubernetes (Una réplica)

Utilización del componente integrado de Kubernetes dentro de Docker Desktop para desplegar la aplicación con una réplica para cada componente del backend y frontend. Este escenario permitirá observar el comportamiento de la aplicación en un entorno orquestado con Kubernetes.

Despliegue en Kubernetes (Tres y nueve réplicas)

Extensión del despliegue anterior mediante la creación de tres y nueve réplicas para cada componente del backend y frontend. Este escenario está diseñado para analizar cómo la replicación afecta el rendimiento y la capacidad de manejo de carga de la aplicación.

Análisis de rendimiento

Para cada uno de estos escenarios, se realizaron pruebas de carga y rendimiento utilizando dos herramientas de interfaz gráfica para testing de microservicios: Postman y JMeter. Estas pruebas permitirán evaluar la capacidad de respuesta; el rendimiento bajo carga, simulando múltiples usuarios concurrentes y midiendo el rendimiento del sistema bajo condiciones de estrés.

Desarrollo

Previo al análisis del rendimiento, se configuran pruebas de carga con los siguientes parámetros:

- 1) Número de hilos, que hace referencia al número de usuarios, fijado en 3000
- 2) Periodo de subida de datos, establecido en 1 segundo
- 3) Número de iteraciones, configurado en 5 iteraciones

Estos parámetros se utilizarán de manera consistente en todas las pruebas, desde aquellas que involucran el despliegue utilizando únicamente Docker, hasta aquellas que requieren un despliegue gestionado mediante Kubernetes. Esta uniformidad en la configuración de las pruebas de carga garantiza la comparabilidad de los resultados y facilita un análisis riguroso del rendimiento de los microservicios bajo diferentes entornos de despliegue.

Por otro lado, se configuran pruebas de carga que proporcionarán una visión integral del desempeño de la aplicación de microservicios. Los parámetros establecidos son los siguientes:

- 1) La media de bytes subidos, que permite evaluar la cantidad de datos enviados al servicio y determinar si el servicio maneja eficazmente grandes volúmenes de datos;
- 2) La latencia mínima, que indica el tiempo más rápido (en milisegundos) que tarda un paquete de datos en ser transmitido y recibido, proporcionando una medida de la eficiencia de la conexión;

- 3) La latencia máxima, que muestra el tiempo de respuesta más largo observado durante la prueba, permitiendo identificar posibles cuellos de botella
- 4) La latencia media, que ofrece una perspectiva general sobre los tiempos de respuesta predominantes, facilitando la evaluación de la consistencia del servicio
- 5) La desviación estándar de la latencia, que mide la dispersión de los tiempos de respuesta, indicando la variabilidad en el desempeño del servicio.
- 6) El porcentaje de error, que se utiliza para cuantificar la cantidad de solicitudes fallidas, proporcionando una medida de la robustez del servicio ante altas cargas.

Estas métricas son esenciales para comprender la capacidad del servicio para manejar solicitudes y su fiabilidad en diferentes condiciones de carga.

Despliegue y análisis de rendimiento con docker

En primera instancia, se realiza el despliegue de la aplicación utilizando Docker, configurando los servicios con los siguientes puertos: frontend-store (80:80), backend-reviews (3000:3000), backend-store (8082:8082) y backend-catalog (8081:8081).

Al someter la aplicación a pruebas de carga con JMeter, configuradas para simular tres mil usuarios en cinco iteraciones a lo largo de un segundo, se obtuvieron los siguientes resultados para:

- El servicio de frontend-store con protocolo "Http", una IP del servidor igual a "localhost", puerto 80, petición GET http y ruta "/"#/"
 - La media de datos subidos fue de 4011.04 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 16520 ms, reflejando posibles cuellos de botella en el servicio.
 - La latencia media se situó en 1666 ms, proporcionando una visión general del tiempo de respuesta.
 - La desviación estándar de 4363.61 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando inconsistencia en el desempeño del servicio bajo carga.

- Se observó un porcentaje de error del 44.81%, lo cual sugiere que casi la mitad de las solicitudes no fueron procesadas correctamente, destacando áreas críticas para mejorar la robustez y fiabilidad del servicio bajo condiciones de alta carga.

Nombre: Reporte resumen for frontend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo: ☐ Escribir en Log ☐ Sólo Errores ☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Bytes
Petición_for_...	15000	1666	0	16520	4363.61	44.81%	105,1/sec	360,80	10,46	4011.04
Total	15000	1666	0	16520	4363.61	44.81%	105,1/sec	360,80	10,46	4011.04

- El servicio de backend-reviews con protocolo “Http”, una IP del servidor igual a “localhost”, puerto 3000, petición GET http y ruta “/reviews”
 - La media de datos subidos fue de 3515,6 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 35981 ms, reflejando posibles cuellos de botella en el servicio.
 - La latencia media se situó en 18980 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 18 segundos en responder a las peticiones. Esto es un tiempo bastante elevado.
 - La desviación estándar de 8094,05 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 18 segundos añadiendo o restando 8 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 17,11%, lo cual sugiere que, de cada 100 peticiones, alrededor de 83 peticiones son respondidas satisfactoriamente y, alrededor de 17 son respondidas erróneamente.

Nombre:

Reporte resumen for bckend-reviews

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log

☐ Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Bytes
Petición_for_...	15000	18980	0	35981	8094,05	17,11%	105,1/sec	360,80	10,46	3515,6
Total	15000	18980	0	35981	8094,05	17,11%	105,1/sec	360,80	10,46	3515,6

- El servicio de backend-store con protocolo “Http”, una IP del servidor igual a “localhost”, puerto 8082, petición GET http y ruta “/”
 - La media de datos subidos fue de 2912,3 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 35339 ms, reflejando posibles cuellos de botella en el servicio de hasta 35 segundos.
 - La latencia media se situó en 1988 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 1 segundo en responder a las peticiones. Esto es un tiempo bastante bajo.
 - La desviación estándar de 6817.88 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 19 segundos añadiendo o restando hasta 6 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 54,95%, lo cual sugiere que, de cada 100 peticiones, alrededor de 46 peticiones son respondidas satisfactoriamente y, alrededor de 54 son respondidas erróneamente.

Reporte resumen

Nombre:

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo Log/Mostrar sólo: ☐ Escribir en Log Sólo Errores ☐ Éxitos

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Bytes
Petición_for_...	30000	1988	0	35339	6817,88	54,95%	24,6/sec	69,91	2,23	2912,3
Total	30000	1988	0	35339	6817,88	54,95%	24,6/sec	69,91	2,23	2912,3

- El servicio de backend-catalog con protocolo “Http”, una IP del servidor igual a “localhost”, puerto 8081, petición GET http y ruta “/api/getlibros”
 - La media de datos subidos fue de 4280,3 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 65045 ms, reflejando posibles cuellos de botella en el servicio de hasta 65 segundos.

- La latencia media se situó en 6498 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 6 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
- La desviación estándar de 12361,61 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 6 segundos añadiendo o restando hasta 12 segundos, dependiendo de la velocidad del servicio.
- Se observó un porcentaje de error del 14,36%, lo cual sugiere que, de cada 100 peticiones, alrededor de 86 peticiones son respondidas satisfactoriamente y, alrededor de 14 son respondidas erróneamente.

Reporte resumen

Nombre:

Comentarios:

Escribir todos los datos a Archivo

Nombre de archivo: Log/Mostrar sólo: ☐ Escribir en Log Sólo Errores ☐ Éxitos

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Bytes
Petición_for_...	45000	6498	0	65045	12361,61	14,36%	10,6/sec	44,31	1,04	4280,3
Total	45000	6498	0	65045	12361,61	14,36%	10,6/sec	44,31	1,04	4280,3

Despliegue y análisis de rendimiento con kubernetes (1 réplica)

- El servicio de frontend-store con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30102, petición GET http y ruta “/#”
 - La media de datos subidos fue de 3422,6 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 9492 ms, reflejando posibles cuellos de botella en el servicio de hasta 9 segundos.
 - La latencia media se situó en 15000 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 15 segundos en responder a las peticiones. Este es un tiempo con un rendimiento costoso.
 - La desviación estándar de 1257,84 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 15 segundos añadiendo o restando hasta 1 segundo, dependiendo de la velocidad del servicio.

- Se observó un porcentaje de error del 67,42%, lo cual sugiere que, de cada 100 peticiones, alrededor de 33 peticiones son respondidas satisfactoriamente y, alrededor de 67 son respondidas erróneamente.

Nombre:

Reporte resumen for frontend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	277	0	9492	1257,84	67,42%	357,9/sec	1196,23	13,32	3422,6
Total	15000	277	0	9492	1257,84	67,42%	357,9/sec	1196,23	13,32	3422,6

- El servicio de backend-reviews con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30001, petición GET http y ruta “/reviews”
 - La media de datos subidos fue de 2964,7 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 15021 ms, reflejando posibles cuellos de botella en el servicio de hasta 15 segundos.
 - La latencia media se situó en 2482 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 2 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
 - La desviación estándar de 3643.98 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 2 segundos añadiendo o restando hasta 3 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 66,88%, lo cual sugiere que, de cada 100 peticiones, alrededor de 34 peticiones son respondidas satisfactoriamente y, alrededor de 66 son respondidas erróneamente.

Nombre:

Reporte resumen for bakcend-reviews

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	2482	0	15021	3643,98	66,88%	338,4/sec	979,87	13,57	2964,7
Total	15000	2482	0	15021	3643,98	66,88%	338,4/sec	979,87	13,57	2964,7

- El servicio de backend-store con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30002, petición GET http y ruta “/”
 - La media de datos subidos fue de 2042,4 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 21240 ms, reflejando posibles cuellos de botella en el servicio de hasta 9 segundos.
 - La latencia media se situó en 300 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 0,3 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
 - La desviación estándar de 2279,42 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 0,2 segundos añadiendo o restando hasta 2 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 100%, lo cual sugiere que, de cada 100 peticiones, alrededor de 0 peticiones son respondidas satisfactoriamente y, alrededor de 100 son respondidas con el código de error.

Nombre:

Reporte resumen backend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	300	0	21240	2279,42	100,00%	224,1/sec	446,90	6,31	2042,4
Total	15000	300	0	21240	2279,42	100,00%	224,1/sec	446,90	6,31	2042,4

- El servicio de backend-catalog con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30000, petición GET http y ruta “/api/getlibros”
 - La media de datos subidos fue de 2816,2 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 23599 ms, reflejando posibles cuellos de botella en el servicio de hasta 23 segundos.

- La latencia media se situó en 341 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 0,3 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
- La desviación estándar de 2390,70 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 0,3 segundos añadiendo o restando hasta 2 segundos, dependiendo de la velocidad del servicio.
- Se observó un porcentaje de error del 74,10%, lo cual sugiere que, de cada 100 peticiones, alrededor de 26 peticiones son respondidas satisfactoriamente y, alrededor de 74 son respondidas con el código de error.

Nombre:

Reporte resumen backend catalog

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	341	0	23599	2390,70	74,10%	227,0/sec	624,15	7,46	2816,2
Total	15000	341	0	23599	2390,70	74,10%	227,0/sec	624,15	7,46	2816,2

Despliegue y análisis de rendimiento con kubernetes (3 réplicas)

- El servicio de frontend-store con protocolo "Http", una IP del servidor igual a "127.0.0.1", puerto 30102, petición GET http y ruta "/"#
 - La media de datos subidos fue de 3425,4 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 12377 ms, reflejando posibles cuellos de botella en el servicio de hasta 12 segundos.
 - La latencia media se situó en 285 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 0,2 segundos en responder a las peticiones. Este es un tiempo con un rendimiento propicio.
 - La desviación estándar de 1642,14 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede

demorar 0,2 segundos añadiendo o restando hasta 1 segundo, dependiendo de la velocidad del servicio.

- Se observó un porcentaje de error del 67,31%, lo cual sugiere que, de cada 100 peticiones, alrededor de 33 peticiones son respondidas satisfactoriamente y, alrededor de 67 son respondidas erróneamente.

Nombre:

Reporte resumen for frontend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	285	0	12377	1642,14	67,31%	144,2/sec	482,43	5,39	3425,4
Total	15000	285	0	12377	1642,14	67,31%	144,2/sec	482,43	5,39	3425,4

- El servicio de backend-reviews con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30001, petición GET http y ruta “/reviews”
 - La media de datos subidos fue de 2958,8 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 31594 ms, reflejando posibles cuellos de botella en el servicio de hasta 31 segundos.
 - La latencia media se situó en 5832 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 5 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
 - La desviación estándar de 9117,43 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 5 segundos añadiendo o restando hasta 9 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 67,33%, lo cual sugiere que, de cada 100 peticiones, alrededor de 33 peticiones son respondidas satisfactoriamente y, alrededor de 67 son respondidas erróneamente.

Nombre:

Reporte resumen for bakcend-reviews

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	5832	0	31549	9117,43	67,33%	142,7/sec	412,30	5,65	2958,8
Total	15000	5832	0	31549	9117,43	67,33%	142,7/sec	412,30	5,65	2958,8

- El servicio de backend-store con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30002, petición GET http y ruta “/”
 - La media de datos subidos fue de 1814,0 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 16296 ms, reflejando posibles cuellos de botella en el servicio de hasta 16 segundos.
 - La latencia media se situó en 249 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 0,2 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
 - La desviación estándar de 1546,65 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 0,2 segundos añadiendo o restando hasta 1 segundo, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 100%, lo cual sugiere que, de cada 100 peticiones, alrededor de 0 peticiones son respondidas satisfactoriamente y, alrededor de 100 son respondidas con el código de error.

Nombre:

Reporte resumen backend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	249	0	16296	1546,65	100,00%	144,2/sec	255,51	5,74	1814,0
Total	15000	249	0	16296	1546,65	100,00%	144,2/sec	255,51	5,74	1814,0

- El servicio de backend-catalog con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30000, petición GET http y ruta “/api/getlibros”

- La media de datos subidos fue de 2911,5 bytes, indicando el volumen de información transmitida durante las pruebas.
- La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
- La latencia máxima alcanzó los 26885 ms, reflejando posibles cuellos de botella en el servicio de hasta 26 segundos.
- La latencia media se situó en 643 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 0,3 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
- La desviación estándar de 2335,75 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 0,3 segundos añadiendo o restando hasta 2 segundos, dependiendo de la velocidad del servicio.
- Se observó un porcentaje de error del 62,88%, lo cual sugiere que, de cada 100 peticiones, alrededor de 38 peticiones son respondidas satisfactoriamente y, alrededor de 62 son respondidas con el código de error.

Nombre:

Reporte resumen backend catalog

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	643	0	26885	2335,75	62,88%	144,7/sec	411,49	6,82	2911,5
Total	15000	643	0	26885	2335,75	62,88%	144,7/sec	411,49	6,82	2911,5

Despliegue y análisis de rendimiento con kubernetes (9 réplicas)

- El servicio de frontend-store con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30102, petición GET http y ruta “/#”
 - La media de datos subidos fue de 3779,1 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 13429 ms, reflejando posibles cuellos de botella en el servicio de hasta 13 segundos.

- La latencia media se situó en 283 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 0,2 segundos en responder a las peticiones. Este es un tiempo con un rendimiento propicio.
- La desviación estándar de 1582,01 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 0,2 segundos añadiendo o restando hasta 1 segundo, dependiendo de la velocidad del servicio.
- Se observó un porcentaje de error del 53,37%, lo cual sugiere que, de cada 100 peticiones, alrededor de 43 peticiones son respondidas satisfactoriamente y, alrededor de 57 son respondidas erróneamente.

Nombre:

Reporte resumen for frontend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐

Escribir en Log Sólo Errores

☐

Éxitos

Configurar

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	283	0	13429	1582,01	53,37%	60,3/sec	222,40	3,21	3779,1
Total	15000	283	0	13429	1582,01	53,37%	60,3/sec	222,40	3,21	3779,1

- El servicio de backend-reviews con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30001, petición GET http y ruta “/reviews”
 - La media de datos subidos fue de 3129,3 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 39533 ms, reflejando posibles cuellos de botella en el servicio de hasta 39 segundos.
 - La latencia media se situó en 4168 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 4 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
 - La desviación estándar de 6738,36 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 4 segundos añadiendo o restando hasta 6 segundos, dependiendo de la velocidad del servicio.

- Se observó un porcentaje de error del 52,51%, lo cual sugiere que, de cada 100 peticiones, alrededor de 48 peticiones son respondidas satisfactoriamente y, alrededor de 52 son respondidas erróneamente.

Nombre:

Reporte resumen for bakcend-reviews

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Min	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	4168	0	39533	6738,36	52,51%	60,3/sec	184,22	3,47	3129,3
Total	15000	4168	0	39533	6738,36	52,51%	60,3/sec	184,22	3,47	3129,3

- El servicio de backend-store con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30002, petición GET http y ruta “/”
 - La media de datos subidos fue de 1468,4 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 56711 ms, reflejando posibles cuellos de botella en el servicio de hasta 56 segundos.
 - La latencia media se situó en 2349 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 2 segundos en responder a las peticiones. Este es un tiempo con buen rendimiento.
 - La desviación estándar de 8432,02 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 2 segundos añadiendo o restando hasta 8 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 100%, lo cual sugiere que, de cada 100 peticiones, alrededor de 0 peticiones son respondidas satisfactoriamente y, alrededor de 100 son respondidas con el código de error.

Nombre:

Reporte resumen backend-store

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Petición_for_...	15000	2349	0	56711	8432,02	100,00%	60,5/sec	86,80	3,47	1468,4
Total	15000	2349	0	56711	8432,02	100,00%	60,5/sec	86,80	3,47	1468,4

- El servicio de backend-catalog con protocolo “Http”, una IP del servidor igual a “127.0.0.1”, puerto 30000, petición GET http y ruta “/api/getlibros”
 - La media de datos subidos fue de 2599,0 bytes, indicando el volumen de información transmitida durante las pruebas.
 - La latencia mínima registrada fue de 0 ms, lo que sugiere que algunas solicitudes fueron procesadas instantáneamente.
 - La latencia máxima alcanzó los 120742 ms, reflejando posibles cuellos de botella en el servicio de hasta 120 segundos.
 - La latencia media se situó en 16569 ms, proporcionando una visión general del tiempo de respuesta que tarda, en promedio, 16 segundos en responder a las peticiones. Este es un tiempo con un rendimiento costoso.
 - La desviación estándar de 19735,47 milisegundos revela una significativa variabilidad en los tiempos de respuesta, indicando que el servicio puede demorar 16 segundos añadiendo o restando hasta 19 segundos, dependiendo de la velocidad del servicio.
 - Se observó un porcentaje de error del 61,44%, lo cual sugiere que, de cada 100 peticiones, alrededor de 38 peticiones son respondidas satisfactoriamente y, alrededor de 62 son respondidas con el código de error.

Nombre:

Reporte resumen backend catalog

Comentarios

Escribir todos los datos a Archivo

Nombre de archivo

Navegar...

Log/Mostrar sólo:

☐ Escribir en Log Sólo Errores

☐ Éxitos

Configurar

Etiqueta	# Muestras	Media	Mín	Máx	Desv. Estándar	% Error	Rendimiento	Kb/sec	Sent KB/sec	Media de Byt...
Peticion_for_...	15000	16569	0	120742	19735,47	61,44%	60,5/sec	153,63	4,07	2599,0
Total	15000	16569	0	120742	19735,47	61,44%	60,5/sec	153,63	4,07	2599,0

Discusión de Resultados

Tras el análisis del rendimiento de los microservicios anteriormente mencionados, bajo las distintas condiciones de despliegue previas, se han revelado resultados significativos. En primer lugar, el porcentaje de error fue considerablemente menor en el despliegue con Docker. Los servicios lograron alcanzar valores de error tan bajos como 14%, 55%, 17% y 45%, en contraste con otros despliegues donde el porcentaje de error no descendió del 50%. Esto sugiere que los contenedores de Docker proporcionan un entorno más estable para la ejecución de microservicios.

Por otro lado, el servicio de backend store presentó un porcentaje de error del 100% en todas las pruebas. Este resultado no implica un fallo completo del servicio, sino más bien la ausencia de una API implementada para recibir peticiones desde la dirección IP con el puerto especificado. Como resultado, se genera un error 404 proporcionado por White Label. Este comportamiento indica que la aplicación está respondiendo adecuadamente, ya que la generación del error 404 por parte de White Label confirma que el servidor está en funcionamiento y es capaz de gestionar solicitudes; de otra forma, se mostraría un error 404 genérico, generado por el navegador respectivo que realice la solicitud.

Finalmente, se observó que el despliegue con Kubernetes tuvo un impacto positivo notable en la reducción de la latencia de los microservicios. Mientras que el promedio de latencia en el despliegue con Docker oscilaba entre 1200 ms y 18000 ms, el despliegue con Kubernetes ofrecía latencias medias de entre 200 y 500 milisegundos. Este rendimiento es excepcionalmente bajo, ya que representa un tiempo de respuesta de apenas 0.2 segundos. En conjunto, estos hallazgos subrayan las ventajas de utilizar Docker para minimizar errores y Kubernetes para optimizar la latencia en el despliegue de microservicios.

Conclusiones

En conclusión, los resultados obtenidos de las pruebas de carga y estrés sobre los microservicios han permitido identificar diferencias clave en el rendimiento entre Docker y Kubernetes. Docker, al utilizar contenedores, ofrece ambientes más seguros y con menor probabilidad de error, destacándose por su estabilidad operativa. En contraste, Kubernetes proporciona ambientes con una latencia significativamente menor, aunque con una probabilidad de error más elevada. Además, se ha observado que la latencia en Kubernetes disminuye con el aumento de réplicas de los servicios, lo que lo convierte en una solución idónea para el escalado horizontal de servicios. Ambas

tecnologías facilitan la gestión y administración automática de los componentes de una aplicación de microservicios. Por tanto, la decisión informada de utilizar Docker en ciertos contextos y Kubernetes en otros, según las necesidades específicas de estabilidad y rendimiento, es crucial para asegurar un rendimiento óptimo de una aplicación de microservicios en su entorno de desarrollo.