

Maestría en Ingeniería de Software

Universidad de los Andes

Desarrollo de software en la nube

Plan de pruebas

Despliegue básico en la nube - entrega Semana 3

Entregado por:

- Daniel Andrés Chala Burbano: d.chala@uniandes.edu.co
- Jorge Enrique Suárez González: je.suarezg1@uniandes.edu.co
- Carlos Andrés Mora Rocha: ca.morar1@uniandes.edu.co
- John Alexander Gutiérrez Lizarazo: j.gutierrezl@uniandes.edu.co

Objetivo:

- Evaluar el nivel de desempeño de la aplicación Web desplegada en un entorno cloud para identificar la capacidad de carga que puede soportar.
- Verificar la capacidad de procesamiento de archivos de la aplicación hasta un nivel aceptable de tiempos de espera.
- Contrastar los resultados obtenidos frente a las pruebas realizadas en un entorno de pruebas local, para identificar aspectos favorables y desfavorables de cada tipo de implementación.

Objetivos específicos:

- Evaluar la cantidad de archivos que se pueden procesar por minuto, bajo unos tiempos de espera de procesamiento aceptables y tamaño mínimo de archivos.
- Medir la cantidad de requests que la aplicación es capaz de soportar cuando existen 30 archivos disponibles.

Descripción general:

En términos generales, se desea evaluar el desempeño de la aplicación cuando se reciben solicitudes de conversión de audio de manera concurrente por parte de los usuarios de la aplicación. En principio, se efectuarán pruebas de carga con 30 archivos de audio para medir la máxima capacidad de procesamiento de la aplicación, bajo ciertas restricciones en cuanto a tiempos de respuesta y tasa de errores aceptada.

Por otro lado, se evaluará el throughput que soporta la aplicación al procesar archivos de audio superiores a un tamaño determinado.

Tipos de pruebas a realizar:

- Pruebas de estrés:
 - Se realizarán pruebas con diferente cantidad de usuario concurrentes, de manera ascendente, con el fin de determinar la carga máxima que soporta la aplicación, cumpliendo las restricciones de tiempo de respuesta y tasa de errores.
 - Se ejecutarán pruebas para diferente número de usuarios concurrentes para identificar la cantidad máxima de archivos procesados durante un minuto, teniendo en cuenta que el tiempo para iniciar el procesamiento de un archivo desde que es

enviado no puede superar los 10 minutos, y que el tamaño de los archivos debe ser de al menos 5MB.

Criterios de aceptación:

- Para calcular la máxima carga soportada, se debe obtener un tiempo de respuesta promedio menor a 1500ms, así como una tasa de errores menor al 1%.
- Para calcular la cantidad de archivos procesados por minuto, el tamaño mínimo de estos archivos será de 5MB.

Datos de prueba:

Para las pruebas tanto del escenario 1 como del escenario 2 se utilizarán archivos de audio con formato .mp3 con un tamaño superior a 5MB. Se ejecutarán peticiones para realizar la conversión de los archivos a los formatos .wav y .ogg, que fueron los 3 formatos escogidos para desarrollar la aplicación de conversión.

Iteraciones:

Con el fin de generar confiabilidad y precisión en los resultados obtenidos, cada escenario será ejecutado al menos 3 veces.

- Prueba de carga exploratoria: Al no conocer de antemano como va a responder la aplicación a escenarios de prueba con usuarios concurrentes, esta prueba nos permitirá determinar cuál es el valor aproximado de carga donde se comienza a saturar la aplicación, ya sea por tiempos de respuesta promedia como por tasa de error.
- Basados en los resultados obtenidos en la primera entrega, se han definido pruebas base para cada escenario planteado.
- Para el primer escenario la configuración es: virtual users: 150, ramp-up: 100 seg, loop count=1. En el caso del segundo escenario, la configuración de la prueba base es: virtual users: 500, ramp-up: 1000, loop count: 4.

A partir del resultado obtenido en la prueba exploratoria podemos ejecutar una prueba de stress con variaciones incrementales de usuarios, iniciando desde una cantidad baja y aumentando de forma gradual hasta alcanzar en el último escalón la cantidad máxima de usuarios soportados según lo estimado en la prueba exploratoria. La distribución de carga a lo largo del tiempo de prueba se observa a continuación:

Iteraciones	Cantidad de VU (Usuarios concurrentes)	Tiempo de escalonado ramp-up (segundos)
1	50% carga máx estimada	100
2	75% carga máx estimada	100
3	100% carga máx estimada	100

Si de llegar a presentarse en una de las iteraciones que la aplicación se satura antes de alcanzar la carga esperada, se volverá a ejecutar. Si el resultado de saturación persiste, se buscará reducir la carga para determinar si se ha llegado al umbral de capacidad.

Si al ejecutar una iteración con la misma carga se presenta una alta saturación de la aplicación se buscará reiniciar los servicios o componentes involucrados, con el fin de mantener las mismas condiciones de operación en cada prueba.

Configuración del sistema:

La prueba será ejecutada en una máquina virtual alojada en la nube Google Cloud Platform, la cual se encuentra dentro de la misma subred de la aplicación bajo pruebas. Esta contiene las siguientes especificaciones:

- Tipo de máquina: e2-medium
- Plataforma de CPU: Intel Broadwell (2VCPU)
- Arquitectura: x86-64
- Sistema operativo: Windows server 2022
- Almacenamiento: 50GB
- Memoria RAM: 4GB

Herramientas de la prueba:

Para ejecutar las pruebas de rendimiento en esta ocasión se utilizará Apache Jmeter, herramienta recomendada para la generación de peticiones a la aplicación, así como creación de reportes de resultados.

Métricas:

Se identificaron dos escenarios principales de pruebas, cada uno con sus mediciones y métricas correspondientes:

- **Escenario 1:** Máxima cantidad de requests HTTP por minuto con usuarios que tengan al menos 30 archivos disponibles para conversión:
 - Respuestas HTTP por minuto
 - Tiempo de respuesta promedio máximo: 1500 ms
 - Tasa de errores: menor al 1%
- **Escenario 2:** Máxima cantidad de archivos procesados por minuto
 - Tiempo de inicio de procesamiento máximo: 10 minutos
 - Tamaño máximo de archivos: 5MB

Riesgos:

Para la ejecución de este plan de pruebas se tienen en cuenta las siguientes limitaciones y consideraciones:

- Al realizar la ejecución de las pruebas de carga durante varias iteraciones, en la máquina se puede presentar una utilización excesiva de recursos. Por lo cual es recomendable, verificar antes de la ejecución que el uso de los recursos se encuentre estables.
- La conexión a internet, así como la conectividad entre la máquina de pruebas y la máquina bajo pruebas son claves para un resultado confiable de la prueba.
- Al realizar la ejecución dentro de un ambiente cloud, la disposición geográfica de las máquinas virtuales, así como de la máquina de pruebas puede influir en los tiempos de respuesta (latencia) entre las mismas.