

OBJETIVOS

- Diseñar escenarios de prueba que permitan medir la capacidad máxima que pueden soportar algunos componentes del sistema. El resultado debe expresarse en el documento del plan de análisis de capacidad con el que se simula el acceso, la carga, el estrés y la utilización de la aplicación.
- Ejecutar las pruebas de estrés que permitan dimensionar la capacidad de una aplicación y su infraestructura de soporte con base al plan de pruebas.

LUGAR Y FORMATO DE ENTREGA

Para cada una de las entregas del proyecto se debe entregar los siguientes recursos:

- Documento del plan de pruebas
- Documento de análisis de capacidad – ejecución de las pruebas de estrés, análisis de resultados y conclusiones.
- Entregar toda la documentación vía GitHub/GitLab (Utilice el mismo repositorio del proyecto).

INFRAESTRUCTURA REQUERIDA PARA LA EJECUCIÓN DE LAS PRUEBAS DE ESTRÉS

En este proyecto se planea utilizar la siguiente infraestructura y servicios:

- Se requiere 1 máquina virtual en su ambiente local, con 2 vCPUs, 2 GiB en memoria RAM. Para realizar las pruebas de estrés. La máquina virtual de pruebas debe estar localizada en el mismo segmento de red de la aplicación, no es necesario que este sobre el mismo host anfitrión.

Dado esto es necesario diseñar una carga efectiva (estrés) e interpretar los resultados correctamente, para esto responda las siguientes preguntas orientadoras:

1. (2%) ¿Cuál es su entorno de prueba? Identifique las características y limitaciones de la infraestructura donde se despliegue su aplicación en producción, así como las características de las herramientas que soportan su aplicación. Es necesario identificar estos aspectos para el equipo de prueba. El entorno físico incluye configuraciones de hardware, software y red. Tener un conocimiento profundo de todo el entorno de pruebas desde el principio permite un diseño y una planificación de pruebas más eficientes y le ayuda a identificar los desafíos de las pruebas al principio del proyecto.

El entorno de prueba consta de la siguiente infraestructura:

- Instancia de EC2 t2.micro de Amazon Web Services con 1 vCPU, 1 GiB RAM, 20 GiB en almacenamiento.
- El servidor funciona con la configuración Ubuntu Server 22.04 LTS (HVM), SSD Volume Type.

- El servidor es desplegado dentro de un contenedor de Docker.
- El servidor debe conectarse a una base de datos S3 de AWS. Esa base de datos es construida en PostgreSQL.

2. (2%) ¿Cuáles son los criterios de aceptación? Identifique los objetivos y limitaciones de tiempo de respuesta, rendimiento y utilización de recursos. El tiempo de respuesta es una preocupación del usuario, el rendimiento es una preocupación comercial y la utilización de recursos es una preocupación del sistema. Además, identifique los criterios de éxito del proyecto que pueden no ser capturados por esos objetivos y limitaciones; por ejemplo, utilizando pruebas de rendimiento para evaluar qué combinación de ajustes de configuración dará como resultado las características de rendimiento más deseables. Valide la información que hemos definido para los escenarios de prueba.

- La aplicación debe seguir funcionando al menos para la mitad de los usuarios totales estimados en las pruebas de estrés.
- El tiempo de respuesta máximo esperado no debe exceder los 1500 ms
- Los archivos manejados por la aplicación deben pesar entre 15 mb y 20 mb sin superar ninguno de los valores.
- El porcentaje de errores no debe sobrepasar el 1%

3. (4%) ¿Cuáles son los escenarios de prueba? Identificar escenarios clave, determinar la variabilidad entre servicios representativos y cómo simular esa variabilidad, definir datos de prueba y establecer qué métricas se deben recopilar. Consolide esta información en uno o más modelos de uso del sistema para implementar, ejecutar y analizar.

- Se quiere probar que la aplicación comprima correctamente los archivos. Se usarán distintos archivos como imágenes o audio y se espera que al menos el 99% de estos archivos se comprima de forma exitosa
- Se desea probar la capacidad de la aplicación de tolerar múltiples usuarios al mismo tiempo. Para ello se harán pruebas con 100 usuarios simultáneos y se espera que los tiempos de respuesta no excedan los 1500 ms.

4. (2%) ¿Cuáles son los parámetros de configuración? Prepare el entorno de prueba, las herramientas y los recursos necesarios para ejecutar cada estrategia a medida que las características y los componentes estén disponibles para la prueba. Asegúrese de entender los requerimientos, las limitaciones y las restricciones de Apache Bench (ab) o JMeter y su APM.

- Se usará la herramienta Jmeter para evaluar las solicitudes Http a la aplicación en cuestión.
- Sus parámetros se configurarán de manera global con la ip donde está corriendo la aplicación y su respectivo puerto.
- Se verificarán solicitudes del tipo POST para los escenarios pedidos
- Se agregará un token de autorización para no violar la seguridad de la aplicación

5. (5%) Escenario 1. Se deberá definir un escenario donde se pueda probar cuál es la máxima cantidad de requests HTTP por minuto que soporta la aplicación web con usuarios. Para hacer pruebas de estrés se debe utilizar la herramienta Apache Bench (ab) o JMeter. Las pruebas de estrés deberán realizarse desde otros equipos diferentes a los utilizados para ejecutar el servidor web y el servidor de base de datos. El escenario y los resultados de las pruebas de estrés deberán ser documentados con gráficas que ilustran cómo se comporta el sistema a medida que Universidad de los Andes el número de clientes accediendo a la aplicación se incrementa hasta llegar al punto de degradar completamente el rendimiento de esta. Restricciones del escenario. En las pruebas de estrés el tiempo de respuesta promedio de la aplicación debe ser de máximo 1.500 ms, si este tiempo no se cumple, se concluye que el sistema NO soporta la cantidad de requests de la prueba. En caso de que durante una prueba se generen más de un 1% de errores en los requests de la prueba, se concluye que la aplicación NO soporta la cantidad de requests de la prueba.

- Se va aprobar el servicio de solicitudes de HTTP para verificar cuantas solicitudes es capaz de soportar antes de llegar a los parametros solicitados. Para este escenario usaremos como tiempo maximo para todos los usuarios 60 seg y aumentaremos la cantidad de usuarios dependiendo del aumento de tiempo.

Archivos procesador por minuto		
# Usuarios	Tiempo medio de respuesta(ms)	% de error
100	410	0
200	470	0
400	492	0
1000	516	0.10%
10000	945	0.45%
47852	1512	0.89%





6. (5%) Escenario 2. Se deberá definir un escenario donde se pueda probar cuál es la máxima cantidad de archivos que pueden ser procesados por minuto en la aplicación local. Para hacer pruebas de estrés se recomienda utilizar la herramienta Apache Bench (ab) o JMeter. Las pruebas de estrés deberán realizarse desde otro equipo diferente a los utilizados para ejecutar el servidor web y el servidor de base de datos. El escenario y los resultados de las pruebas de estrés deberán ser documentados con gráficas que ilustran cómo se comporta la aplicación a medida que el número de usuarios procesando archivos se incrementa, hasta llegar al punto en que el tiempo para iniciar el procesamiento de un archivo enviado por un usuario supere los 10 minutos (600 segundos). Restricciones del escenario. El archivo enviado a convertir durante las pruebas debe ser de un tamaño mínimo de 15 MiB y un máximo de 20 MiB. Nota Importante: Las pruebas de estrés de los escenarios 1 y 2 deben ejecutarse de forma aislada.

- Se va a probar el servicio de solicitudes de HTTP para verificar cuantos archivos es capaz de comprimir antes de llegar a los parametros solicitados. Para este escenario usaremos como tiempo maximo para todos los usuarios 60 seg y aumentaremos la cantidad de usuarios dependiendo del aumento de tiempo.

Archivos procesador por minuto			
# Usuarios	Tiempo medio de respuesta(ms)	Maximo tiempo de espera(ms)	Minimo tiempo de espera(ms)
100	534	732	479
200	542	1826	475
400	551	2355	476
1000	616	623994	477
10000	1086	623494	500
100000	9586	652411	1052
548630	61230	702356	1786

