

## **Análisis de capacidad**

### **Entrega 2 – Desarrollo de Software en la Nube**

**Diego Alejandro Camelo Giraldo, Juan Sebastián Alegría Zúñiga, María Camila Gómez Hernández, Andrés Felipe Lugo Saavedra**

#### **Grupo 6**

Nota: Antes de ejecutar las pruebas de estrés, ajuste el documento del plan de análisis de capacidad con base a la retroalimentación realizada por los tutores del curso. 1. Pruebas de estrés sobre el Modelo de Despliegue.

Escenario 1. Se deberá definir un escenario donde se pueda probar cuál es la máxima cantidad de peticiones HTTP por minuto que soporta el API con 30 archivos publicados. Para hacer pruebas de estrés se sugiere utilizar la herramienta Apache Bench (ab) o JMeter que debe instalarse en el ambiente de nube (para no sesgar los resultados de la prueba con la latencia que introduce la red, la máquina de pruebas debe ubicarse en el mismo segmento de red de la aplicación). 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 el número de usuarios virtuales accediendo a la aplicación se incrementa hasta llegar al punto de degradar completamente el rendimiento del sistema.

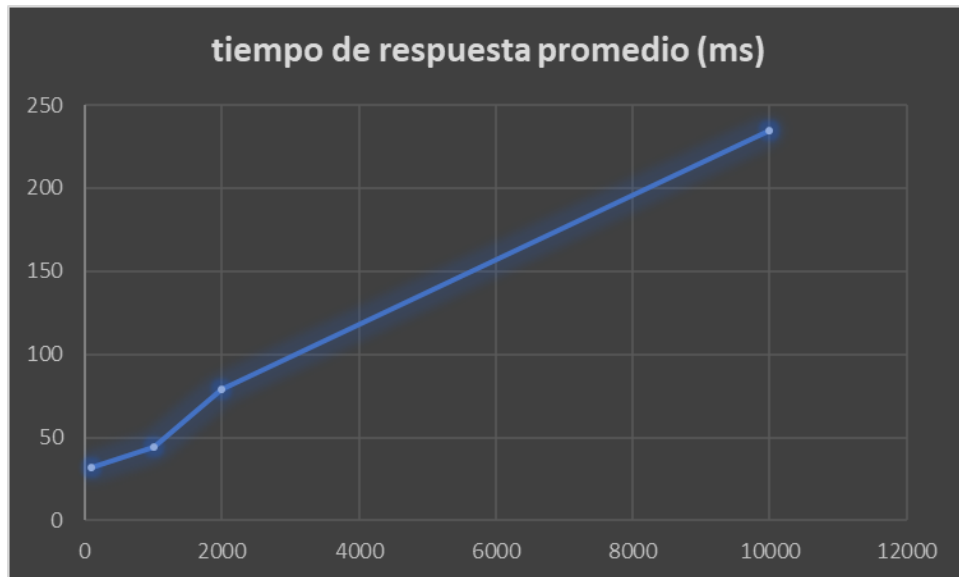
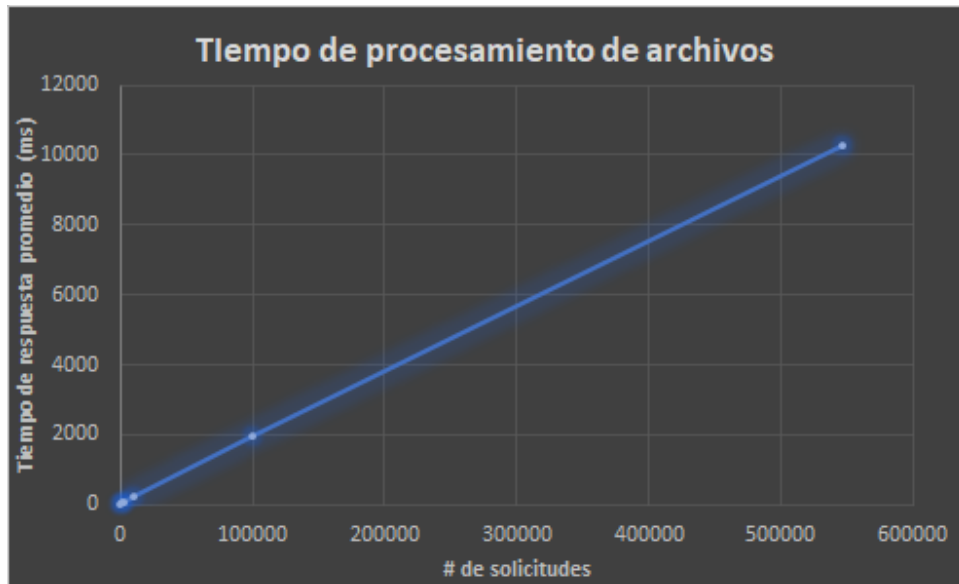
Restricciones del escenario. En las pruebas de stress el tiempo de respuesta promedio de la página debe ser de máximo 1.500 ms, si este tiempo no se cumple, se concluye que el sistema NO soporta la cantidad de peticiones de la prueba. En caso de que durante una prueba se generen más de un 1% de errores en las peticiones de la prueba, se concluye que el sistema NO soporta la cantidad de peticiones.



En este escenario, llevamos a cabo una prueba de carga en la que fuimos aumentando gradualmente el número de usuarios que enviaban solicitudes, mientras registramos los tiempos de respuesta esperados y el porcentaje de error en cada solicitud. Durante la primera etapa de la prueba, en la que se enviaron menos de 15000 solicitudes, se observó que el porcentaje de error era casi inexistente y los tiempos de respuesta se mantuvieron estables y por debajo de las restricciones del escenario. Sin embargo, al superar las 15000 solicitudes, tanto el porcentaje de error como los tiempos de espera aumentaron de manera

significativa, mostrando un crecimiento exponencial como se puede ilustrar en la gráfica mostrada anteriormente. Finalmente, antes de alcanzar las 16000 solicitudes, se superaron los límites establecidos para los errores y los tiempos de espera. En conclusión, se determinó que el sistema es capaz de soportar con éxito poco más de 15000 solicitudes. Es importante resaltar que las pruebas pudieron verse afectadas por la cantidad de recursos ofrecidos por la máquina virtual, ya que los recursos para realizar una gran cantidad de solicitudes podrían sobresaturar los componentes básicos como la memoria RAM o el CPU.

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. Para hacer pruebas de estrés se debe utilizar la herramienta Apache Bench (ab) o JMeter que debe instalarse en el ambiente de nube (para no sesgar los resultados de la prueba con la latencia que introduce la red, la máquina de pruebas debe ubicarse en el mismo segmento de red de la aplicación). 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 el número de usuarios convirtiendo 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).



En el presente escenario, se ha observado que la aplicación es capaz de gestionar una cantidad aproximada de 2000 archivos sin ninguna dificultad. Siguiendo con el aumento de archivos, vemos que la respuesta de los servicios no cambia, pues se comporta de manera lineal. No obstante, al momento de elevar la cantidad de solicitudes a 100000, se puede evidenciar la aparición de un cuello de botella en el proceso de archivos, el cual se traduce en un aumento en los tiempos de respuesta y en la ocurrencia de fallos en algunas de las

solicitudes. Esto notado en el programa, ya que los tiempos promedio empiezan a exceder los 2 minutos y algunos archivos esperan tanto tiempo por ser atendidos, que empiezan a cancelarse y generar un error. Asimismo, se ha constatado un incremento considerable en el consumo de recursos durante la prueba. Es a partir de esta cifra que se hace evidente un incremento considerable en los tiempos de procesamiento. Es importante destacar que la cantidad de solicitudes pudo haberse visto afectada por la disponibilidad de recursos de la máquina virtual. Sin embargo, es preciso mencionar que las solicitudes pudieron ser completadas de manera satisfactoria sin superar un porcentaje de error del 2%. Con base en estos hallazgos, se puede concluir que la aplicación presenta un buen rendimiento en cuanto a la gestión de solicitudes menores a 100000 archivos de procesamiento simultáneo, pero a partir de esta cifra, se podrían experimentar fallos en algunas de las solicitudes.

### **Cuellos de botella:**

Al probar los servicios de la aplicación con la herramienta jmeter, logramos descubrir la cantidad de solicitudes que estas son capaces de soportar antes de empezar a presentar fallas y que el rendimiento de la aplicación comience a fallar.

Servicio de Sign-Up: Soporta un aproximado de 17000 solicitudes en un minuto, con un rendimiento aproximado de 345 ms por solicitud. El cuello de botella comienza cerca de las 19000 solicitudes subiendo los tiempos de respuesta a 1647 ms y creciendo los porcentajes de error a 1,03 %

Servicio de Login: Soporta un aproximado de 15000 solicitudes en un minuto, con un rendimiento aproximado de 345 ms por solicitud. El cuello de botella comienza cerca de las 16000 solicitudes subiendo los tiempos de respuesta a 1521 ms y creciendo los porcentajes de error a 0,94 %

Servicio de carga de archivos: Soporta un aproximado de 10000 solicitudes en un minuto, con un rendimiento aproximado de 225 ms por solicitud. El cuello de botella es un poco más implícito en esta situación, pero se puede notar que el rendimiento empieza a bajar desde las 100000 solicitudes, permitiendo que se excedan los 10 minutos entre las 500000 y las 600000 solicitudes. El porcentaje de error oscila entre 2% y el 4.7%

Servicio de acceso a archivos: Soporta un aproximado de 15000 solicitudes en un minuto, con un rendimiento aproximado de 521 ms por solicitud. El cuello de botella comienza cerca de las 16000 solicitudes subiendo los tiempos de respuesta a 1748 ms y creciendo los porcentajes de error a 1,35 %