

Análisis de Capacidad. Grupo 5, Desarrollo de Soluciones Cloud

Jairo Céspedes Plata, María Catalina Ibáñez, José Manuel Moreno, David Saavedra Martínez

Contexto del proyecto

El proyecto sobre el cual se adelantarán los planes descritos a continuación, tiene como principal propósito la conversión de distintos formatos de archivos mediante un servicio web. En consecuencia, resulta necesaria la ejecución de un análisis de capacidad.

Herramientas a utilizar

Para llevar a cabo el análisis de capacidad exhaustivo se utilizará una instancia de cómputo de Amazon EC2 (t2.micro, ver Imagen 1) dentro del ambiente AWS Academy tal que ejecute la herramienta de carga JMeter. La información obtenida se visualizará en Tableau, PowerBi u herramientas similares.

La elección de esta instancia obedece a que el servicio de AWS permite de manera gratuita, realizar pruebas. Además, se buscará que la instancia esté ubicada en una región que permita minimizar la latencia frente a la ubicación del API que estará en una zona de GCP (Norte De Virginia (us-east4)), por tanto, se seleccionará en AWS la Región Este de EE. UU. (Norte de Virginia) y también se busca maximizar la disponibilidad de recursos durante las pruebas, ya que muchas veces se tiene la máquina local ocupada [2,3].

Nombre	vCPU	RAM (GiB)	Créditos de CPU/h
t2.nano	1	0,5	3
t2.micro	1	1,0	6

Imagen 1. Capacidad instancia t2.micro AWS [1].

Adicionalmente, es necesario hacer uso de herramientas de prueba de carga que permitan una evaluación del rendimiento de un servicio o aplicación. En el presente contexto se ha decidido utilizar JMeter ya que es un framework ampliamente conocido, así como por la gran cantidad de documentación disponible. Se configurarán detalladamente los parámetros de JMeter (ver Imagen 2) ajustando el número de hilos, el tiempo de rampa y otras variables relevantes para simular cargas realistas en el sistema.

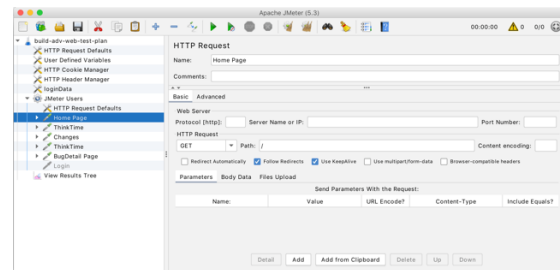


Imagen 2. Interfaz de JMeter.

Adicionalmente, se diseñarán escenarios de prueba detallados y realistas para abordar diferentes aspectos críticos del sistema.

Criterios de medición de desempeño

Como marco general para la construcción de las pruebas se tuvo la metodología GSM (Goal, Signal, Metric) desarrollada por Google [4]. Dicho enfoque permite construir mediciones basadas en las necesidades del proyecto que podrán variar a lo largo del tiempo y no realizar el proceso inverso de fijar métricas arbitrarias para luego ajustar las pruebas.

Como principales criterios están la carga, la conversión y la respuesta, explicados a detalle en la tabla 1.

	GOAL	SIGNAL	METRIC
Carga	Usuarios podrán acceder a la carga de archivos de manera eficiente.	Cuando un usuario intenta cargar un archivo, la plataforma está disponible y funciona de manera rápida.	Respuesta de las peticiones en menos de 1000ms.
Conversión	Los documentos convertidos deberán ser correctos y procesados de manera rápida.	Cuando un usuario realiza una conversión recibe el documento correcto y de manera rápida.	100% de los archivos deben ser correctos y procesados en menos de 1 minuto.
Respuesta	La plataforma es rápida y muestra la información de manera fluida.	La información es accesible de manera rápida y la navegación es fluida.	LCP 500 milisegundos FID 1 segundo CLS 800 milisegundos

Tabla 1. Fijación de métricas estimadas.

Adicionalmente y buscando brindar mayor robustez se presentan criterios de aceptación para ejecutar los escenarios de carga con características y técnicas mejor definidas.

Escenarios de prueba

1. Carga de Archivos:

- Se simulará la carga de archivos en el sistema mediante solicitudes HTTP utilizando datos de prueba variados que reflejen la diversidad de archivos que pueden ser cargados por los usuarios en la aplicación (archivos con extensión .docx, .xlsx y .pptx). Se evaluará la capacidad del sistema para manejar archivos de 5 Mb, así como la concurrencia de múltiples usuarios cargando archivos simultáneamente.
- Estos *requests* deberían responderse en promedio, en menos de 1000 ms. En el caso que el tiempo establecido sea superado, se considerará que el sistema perdió rendimiento y la aplicación llegó a su límite. La cantidad máxima de usuarios será aquella en la que no degrade ese tiempo de respuesta. Por lo tanto, se

espera que, en promedio, se logre subir 1 archivo por segundo, aunque esto también podrá depender, para efectos de la prueba y la simulación del número de usuarios concurrentes. Finalmente, se espera que la carga de archivos no supere una utilización de RAM de alrededor del 30%; esto dado que se puede cargar un archivo, mientras se está convirtiendo otro a PDF, por lo que se necesita capacidad disponible para esta operación.

2. Procesamiento de Archivos a PDF y métricas de desempeño:

- Se simulará el procesamiento de archivos para convertirlos a formato PDF, una operación crítica en el flujo de trabajo de la aplicación. Se medirá el tiempo necesario para procesar cada archivo y se evaluará la capacidad del sistema para manejar cargas de trabajo intensivas considerando factores como el tamaño y la complejidad de los archivos, así como la

disponibilidad de recursos del sistema.

- Se estimará la cantidad de archivos con extensión .docx, .pptx y .xlsx que fue capaz de convertir a pdf, después de que se lance la tarea en 1 minuto, mediante una cola de mensajes, cada documento cargado debe pesar 5 Mb.
- Se espera que una conversión no demore más de 1 minuto, que se logren procesar 2 archivos por minuto y que la utilización de la CPU y RAM no sea mayor al 60%. Si alguno de estos SLOs es superado se considerará que el sistema perdió rendimiento y la aplicación llegó a su límite

Cada prueba se va a ejecutar al menos 5 veces para evitar el sesgo estadístico y se anotarán los resultados de cada una, estos se mostrarán en la herramienta de visualización escogida (ver Imagen 3) mediante histogramas, gráficas de dispersión y de línea que permitirán sacar conclusiones más específicas como:

- Cantidad de usuarios concurrentes que suben archivos al API sin afectar

el tiempo de respuesta de menos de 1 segundo.

- Cuántos archivos es capaz de convertir el sistema en menos de 1 minuto siendo de diferentes tipos (.docx, .pptx y .xlsx) y de un tamaño de 5Mb.
- Comportamiento de las métricas en un escenario de estrés (utilización CPU y RAM, capacidad de procesamiento y tiempo de respuesta).

Bibliografía

- [1] <https://aws.amazon.com/es/ec2/instance-types/t2/#:~:text=Las%20instancias%20T2%20se%20pueden,750%20horas%20de%20instancias%20t2.>
- [2] <https://cloud.google.com/about/locations?hl=es#lightbox-regions-map>
- [3] https://aws.amazon.com/es/about-aws/global-infrastructure/regions_az/
- [4] [Software Engineering at Google \(abseil.io\)](https://abseil.io)



Imagen 3. Interfaz PowerBI.