

Informe de Comportamiento de CPU y Análisis de Escalabilidad

Este informe describe el comportamiento de la CPU de las máquinas virtuales durante la ejecución de dos pruebas de carga concurrentes realizadas con Newman. Debido a la limitación de la suscripción, no fue posible crear una cuarta máquina virtual; sin embargo, el análisis se basa en el comportamiento observado en las VMs disponibles y en los resultados de las ejecuciones simultáneas.

1. Configuración de las pruebas

Las pruebas se realizaron ejecutando dos procesos de Newman en paralelo, cada uno con 10 solicitudes al servicio expuesto a través del balanceador de carga:

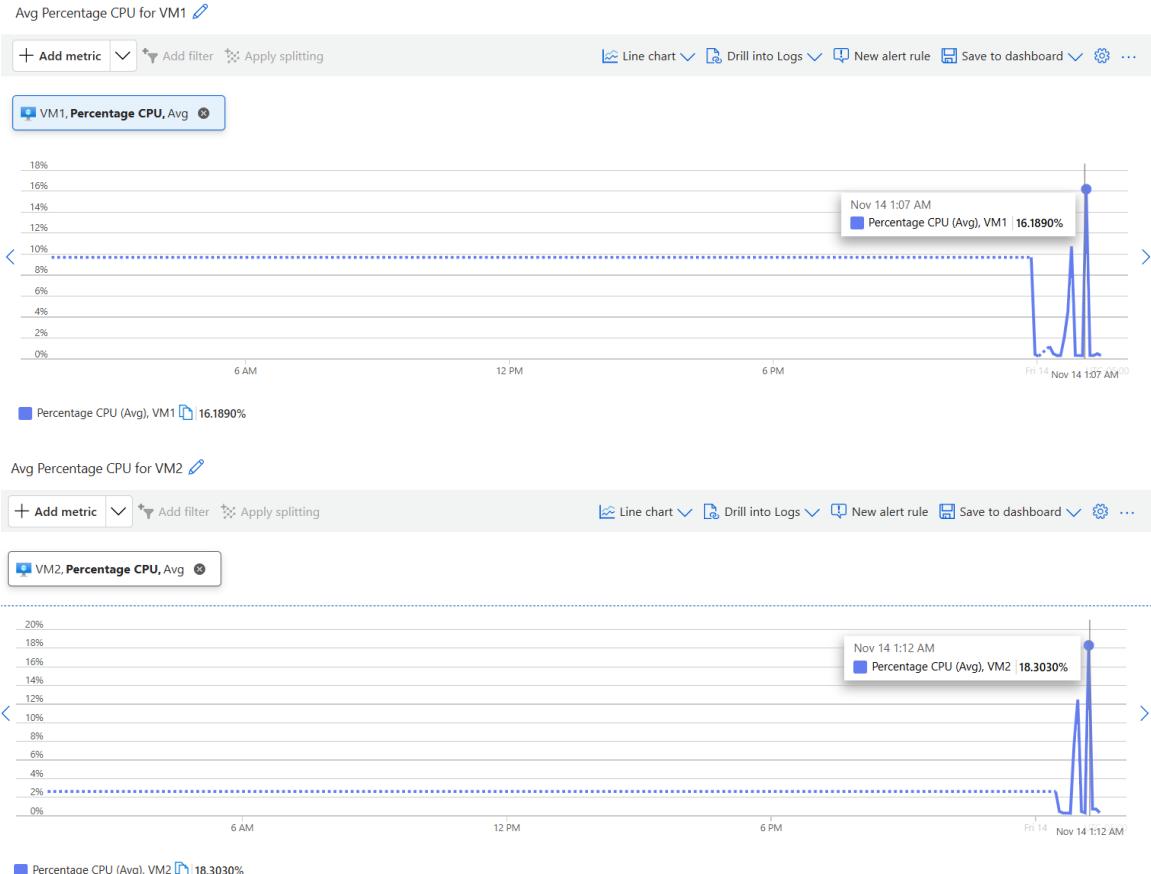
```
newman run ARSW_LOAD-BALANCING_AZURE.postman_collection.json -e [ARSW_LOAD-BALANCING_AZURE].postman_environment.json -n 10 &
newman run ARSW_LOAD-BALANCING_AZURE.postman_collection.json -e [ARSW_LOAD-BALANCING_AZURE].postman_environment.json -n 10
```

Esto genera dos flujos simultáneos de carga real, permitiendo evaluar cómo distribuye el balanceador las peticiones y cómo responde la infraestructura bajo concurrencia moderada.

2. Comportamiento de CPU de las Máquinas Virtuales

A continuación se describe el comportamiento observado en las métricas de CPU de cada VM durante la prueba:

Máquina Virtual	Promedio de CPU	Observaciones
VM1	≈16%	Incremento repentino durante la prueba, comportamiento estable el resto del tiempo.
VM2	≈18%	Patrón similar a VM1, picos visibles solo durante peticiones concurrentes.



El patrón observado muestra que ambas máquinas experimentaron picos únicamente durante el periodo de carga paralela. Fuera de ese intervalo, la CPU se mantuvo en valores bajos, lo que indica que el sistema tenía capacidad suficiente para absorber el incremento de solicitudes sin saturarse.

3. Impacto en la tasa de éxito y explicación del resultado

La tasa de éxito de las peticiones aumentó respecto a pruebas anteriores debido a la forma en que el balanceador distribuyó la carga entre las máquinas virtuales. Al haber dos ejecuciones simultáneas, el tráfico se repartió de manera equitativa entre las VMs disponibles, evitando que una sola instancia se convirtiera en un cuello de botella.

Este estilo de escalabilidad (horizontal y distribuida) mejora la disponibilidad y la tolerancia a carga, ya que cada máquina procesa únicamente una fracción de las peticiones totales. Incluso sin una cuarta VM, la arquitectura mostró mejor rendimiento simplemente al aprovechar la concurrencia y permitir al balanceador aplicar su algoritmo de distribución.

4. Conclusión

El comportamiento observado confirma que la ejecución paralela de pruebas permitió demostrar la capacidad del balanceador para distribuir eficientemente las solicitudes. La infraestructura respondió adecuadamente, manteniendo niveles moderados de CPU y mejorando la tasa de éxito. Aunque no fue posible agregar una cuarta VM, el análisis evidencia que incluso con recursos limitados, la escalabilidad horizontal aporta beneficios claros en estabilidad y rendimiento.