**Simulación de desempeño de un servicio de Cloud Computing para la página web de E-Commerce**

**Rodriguez Echaniz, Mora; Piloni, Ignacio; Fedorenko, Martin Fernando**

***Universidad Tecnológica Nacional, Facultad Regional Buenos Aires***

**Abstract**

*Este estudio examina el rendimiento de un servicio de computación en la nube para un sitio web de E-commerce* *bajo condiciones operativas dinámicas mediante el uso de eventos a intervalos constantes de tiempo. La empresa enfrenta una demanda fluctuante, que incluye un aumento significativo de peticiones durante las horas pico y picos ocasionales de ataques DDoS. Para gestionar estos desafíos, los períodos de mantenimiento reducen temporalmente la capacidad del servidor al detectar una brecha. Los eventos estacionales y festivos amplifican aún más el tráfico, poniendo a prueba los recursos del servidor. Las capacidades del servidor, influenciadas por las especificaciones de RAM y CPU, determinan la máxima tasa sostenible de procesamiento de solicitudes. Los criterios de evaluación para la selección de servicios en la nube incluyen la capacidad de manejo, la efectividad en la gestión de recursos y la mitigación de solicitudes rechazadas.*

**Palabras Clave**

Cloud Computing, E-commerce, Ataques DDoS, Δt constante, periodos de mantenimiento

**Introducción**

Es innegable que con el correr de los años los servicios de Cloud Computing han cobrado una gran importancia en el mundo, afianzándose como una alternativa atractiva a la hora de disponer recursos de almacenamiento, procesamiento e infraestructura. Diversas empresas, tanto grandes como pequeñas han optado por utilizar estos servicios debido a su facilidad de acceso, flexibilidad y escalabilidad, brindadas por el régimen bajo demanda que ofrecen estos servicios, donde solo es necesario pagar por la cantidad de recursos contratados por el usuario.

Sin embargo, las ventajas de este tipo de servicios pueden convertirse en un inconveniente de no contratar un servicio web acorde a las necesidades que se tengan.

Un ejemplo de esto puede verse en los sitios web de las empresas de e-commerce. En estos sitios uno de los principales protagonistas son las peticiones o transacciones realizadas por los usuarios con el fin de adquirir productos o servicios en la página web. Sin embargo, la cantidad de peticiones recibidas por un sitio web no puede ser sencillamente predecida, debido a que esta está influenciada por una gran diversidad de factores como los horarios, las estaciones, las fechas especiales o incluso los ataques que se realicen sobre el sitio. Debido a esto, no conocer exactamente la cantidad de peticiones que recibe un sitio web podría provocar una mala asignación o contratación de los recursos de Cloud Computing, conllevando consecuencias económicas o funcionales para la empresa.

Por lo tanto, tomando en consideración este contexto, el presente trabajo tendrá como objetivo analizar y determinar mediante técnicas de simulación una óptima asignación de recursos de Cloud Computing sobre un sitio web de e-commerce con el fin de garantizar un mejor servicio para sus clientes.

**Elementos del Trabajo y metodología**

La empresa de e-commerce Solo Deportes ha proporcionado datos detallados sobre el flujo de peticiones diarias y sus variaciones debido a diversos eventos. Analizando dichos eventos se observa que los mismos vienen dados en densidades, por lo que la metodología usada es por eventos dados a intervalos constantes de tiempo con un ΔT de 1 hora.

Con respecto a la clasificación de las variables extraídas de los datos, se identifica la variable exógena peticiones por hora **(PH)** que representa la cantidad de peticiones recibidas en horarios normales dada por una función de densidad de probabilidad (FDP) con valores de alrededor de 2000. La función de probabilidad fue obtenida mediante el uso de un programa implementado con el lenguaje Python y un archivo de tipo CSV otorgado por el cliente.

Además, también se definen como variables exógenas a las variables de control de cantidad de núcleos de CPU **(N)**, cantidad de hilos de CPU **(H)**, cantidad de instrucciones por ciclo **(IPC)**, la velocidad del clock **(VC)** y la cantidad de memoria RAM **(M)** ya que representan aquellas variables susceptibles de manipulación por los responsables de contratar el servicio de Cloud Computing. Para esta tarea se debe considerar que cada petición ocupa determinada cantidad de memoria RAM y posee cierta cantidad de instrucciones para su ejecución.

Continuando con la clasificación de variables, se identifica la variable endógena de estado **(P)** como la cantidad de pedidos en cada hora. Mientras que las variables endógenas de resultado son la cantidad de veces que la capacidad del servidor fue superada por la cantidad de peticiones (**CVCS**), el porcentaje de RAM faltante (**PRF**), el porcentaje de CPU faltante (**PCF**), el porcentaje de peticiones rechazadas por hora (**PPR**) y, finalmente, la cantidad de veces que hubo ataques DDoS (**CVDDoS**).

En base al análisis llevado a cabo junto al cliente se determinaron los distintos eventos que deben ocurrir en nuestra simulación.

En primer lugar se consideran, como eventos propios la llegada de peticiones al sitio web, las cuales pueden ser influenciadas por: horarios pico de 18 a 22 horas, aumentando las peticiones en un 300%. Eventos estacionales, que incrementan la demanda del servicio en un 150% cada tres meses, relacionados con cambios de estación y festividades. Así como eventos fijos con fechas específicas, como Día del Niño, Día del Padre, Día de la Madre, Navidad, Hot Week, Black Friday y Cyber Monday. En la semana previa a estos eventos, el número de peticiones aumenta en un 15% diario a medida que se acerca la fecha.

Por otro lado, se considera como evento propio el procesamiento de peticiones por el servidor donde la cantidad máxima de peticiones que podrá procesar el servidor está dada por la siguiente fórmula:

Mientras que la cantidad máxima de peticiones que podrá almacenar el servidor está dada por la siguiente fórmula:

Por último, el estudio considera otro evento propio, que son los ataques DDos. Existe una probabilidad del 0.08% de enfrentar un ataque DDoS por hora, lo cual incrementa las peticiones recibidas en un 1000%. Para identificar un ataque, se revisa si el número de peticiones en la última hora excede en un 100% la capacidad máxima del servidor. Si se detecta un ataque, el servidor debe entrar en mantenimiento, limitando la capacidad de procesamiento al 80% durante 2 horas. Adicionalmente, dicho mantenimiento debe ser una tarea programada mensualmente más allá de los mantenimientos realizados por ataques DDoS.

A partir de la clasificación de variables y eventos se define la siguiente Tabla de Eventos:

| **Eventos Propios** | **Eventos Comprometidos ΔT Anteriores** | **Eventos Comprometidos ΔT Futuros** |
| --- | --- | --- |
| Llegada de peticiones al  sitio web | - | - |
| Procesamiento de  peticiones por el servidor | - | - |
| Ataque DDoS | - | - |

*Tabla 1: Tabla de Eventos*

**Resultados**

Para la simulación se utilizó un tiempo final de **10 años,**  que representa un total de **8760 horas**, donde se tuvieron en cuenta los siguientes escenarios de acuerdo a distintos valores asignados a las variables de control.

| **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** |
| --- | --- | --- | --- |
| M=48Gb | M=64Gb | M=32Gb | M=32Gb |
| N=20 | N=20 | N=16 | N=8 |
| H=16 | H=32 | H=32 | H=8 |
| VC=3 GHz | VC=3 GHz | VC=2 GHz | VC=1 GHz |
| IPC=6 | IPC=10 | IPC=4 | IPC=4 |

*Tabla 1: Tabla de Escenarios*

| **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** |
| --- | --- | --- | --- |
| **CVCS**=1013 | **CVCS**=979 | **CVCS**=1045 | **CVCS**=1036 |
| **PRF**=1,42% | **PRF**=0,79% | **PRF**=2,75% | **PRF**=2,72% |
| **PCF**=0,67% | **PCF**=0,16% | **PCF**=1,31% | **PCF**=3,12% |
| **PPR**=8,1% | **PPR**=6,03% | **PPR**=10,42% | **PPR**=10,34% |
| **CVDDoS**=58 | **CVDDoS**=78 | **CVDDoS**=86 | **CVDDoS**=77 |

*Tabla 1: Tabla de Resultados*

**Discusión**

Al analizar los resultados de los diferentes casos de simulación, se puede observar que las configuraciones de hardware más robustas tienden a manejar mejor la carga y los ataques DDoS. El Escenario 2, con 64 GB de RAM, 20 núcleos, 32 hilos, 3 GHz y 10 IPC, mostró el mejor rendimiento general, con un porcentaje de capacidad de RAM faltante del 0.79% y de CPU del 0.16%, y el menor porcentaje de peticiones rechazadas (6.03%). Esto sugiere que aumentar tanto la capacidad de RAM como la eficiencia de procesamiento (número de hilos y IPC) mejora significativamente la capacidad del servidor para manejar cargas elevadas y ataques. Por otro lado, el Escenario 4, con la configuración más modesta (32 GB de RAM, 8 núcleos, 8 hilos, 1 GHz y 4 IPC), tuvo el peor desempeño, con un porcentaje de capacidad de CPU faltante del 3.12% y un porcentaje de peticiones rechazadas del 10.34%, indicando que esta configuración es la más deficiente para las demandas simuladas.

**Conclusión**

El análisis derivado de esta simulación muestra la importancia de evaluar adecuadamente las distintas capacidades de los servicios a contratar. De estas capacidades dependerá la eficiencia y eficacia del sitio web a la hora de procesar peticiones de los clientes, sentenciado de cierta forma que costos y beneficios tendrá la empresa, ya que ante una contratación excesiva de estos recursos se podría generar un gran aumento en los costos de la empresa, mientras que una adquisición escasa de recursos podría provocar una disminución drástica de los beneficios.

Finalmente, como futuras líneas de investigación se plantea estudiar la adición de nuevas variables de control y la identificación de nuevos eventos relevantes que enriquezcan el modelo.

**Agradecimientos**

Al cliente por proveer colaboración e información necesaria para el desarrollo del proyecto.

También agradecer a la cátedra de Simulación de la Universidad Tecnológica Nacional Facultad Regional de Buenos Aires (UTN FRBA) por brindar todos los conocimientos teóricos y metodologías necesarias para llevar adelante este trabajo.