**Informe experimento 2 Solumovil**

Pre experimentación

Problemática:

La ubicación y el seguimiento de un gran número de vehículos es un aspecto de vital importancia para la empresa de transporte Tbc. Por esto Tbc ha decidido contratar a SoluMovil para diseñar y construir una aplicación que sea capaz de responder ante un gran número de solicitudes simultáneas de sus vehículos (Mobibus, vcub y tranvía eléctrico) así como una respuesta rápida ante estas. Adicionalmente la aplicación construida debe atender todas las solicitudes que se le hagan.

Objetivo del experimento:

Aceptar, procesar y responder a las solicitudes de múltiples usuarios simultáneos en la aplicación SoluMovil con el fin de satisfacer las necesidades de la empresa Tbc. Específicamente, responder al 100% de las solicitudes enviadas por los vehículos manteniendo los atributos de desempeño y escalabilidad.

Descripción del experimento:

Para cumplir con los requerimientos del cliente se siguieron los siguientes pasos:

1. Terminación, corrección y pruebas de los requerimientos funcionales de la entrega del experimento 1. Esto fue completar los servicios REST así como implementar estos en las aplicaciones móviles.
2. Investigación de métodos de balanceo de carga así como de tecnologías que permiten su implementación de forma sencilla con Django y Python. Selección de una tecnología.
3. Implementación de la tecnología escogida en el paso anterior.
4. Pruebas de carga teniendo como métrica la disponibilidad. Está siendo del 100%.
5. Al no obtener los resultados deseados se repitieron los pasos 3 y 4 hasta que se cumplieron los objetivos.

Artefacto a construir.

Módulo de balanceo de carga en la aplicación. Aplicaciones móviles completamente terminadas. Servicio REST.

Recursos de Experimentación

Seleccionar los recursos no fue una tarea sencilla. Esto ya que se debía respetar los elementos de la arquitectura que proporcionaban la escalabilidad y el desempeño y a la vez implementar nuevos que satisficieran la disponibilidad.

Hardware.

Resultados esperados.

Se espera que cualquier petición al servicio REST sea resuelta en menos de 1 s además de que el 100% de las peticiones tengan respuesta.

Duración y etapas:

**Post experimentación**

Para satisfacer la disponibilidad del 100% se decidió utilizar balance de carga como se muestra en la imagen 1

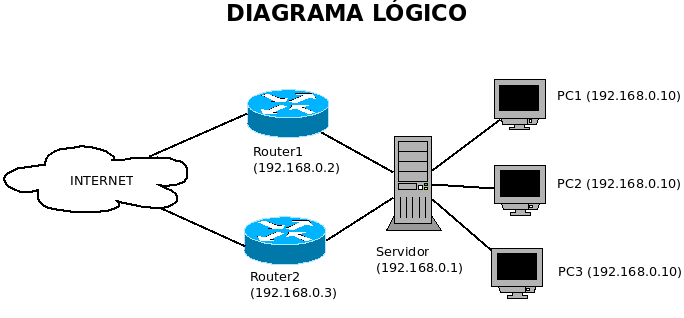


Imagen Balanceo de crga

Balanceo de carga.

El balance de carga ha sido uno de los problemas principales de los sistemas de computación distribuidos. Esto, ya que se deben combinar especificaciones del cliente así como las restricciones del sistema para obtener la mejor respuesta posible. Adicionalmente el poder distribuir los recursos durante la ejecución de una solicitud para así asegurar que se responda lo antes posible sin consumir recursos extra. A este problema se le llama balanceo de carga dinámico.

Existen distintas técnicas de balanceo d carga que se aplican mejor a diferentes problemas. A continuación se presentan una lista de técnicas consideradas para la solución deseada. [1]

Round Robin: Distribuye cada petición en una línea de servidores para que así eventualmente cada servidor tenga la misma cantidad de conexiones que los demás. Se considera esta técnica puesto que funciona bien en la mayoría de los casos. Sin embargo existen peticiones que tiene una mayor prioridad que otras como por ejemplo la emergencia de un Mobibus.

Ratio (member) Ratio (node): Distribuye las peticiones en los servideros de acuerdo a una ponderación definida por el usuario. Esta técnica se considera ya que se tiene un equipo con más capacidad de procesamiento entonces este debería tener una mayor ponderación.

Dynamic Ratio (member) Dynamic Ratio (node): Ratio (member) Ratio (node) solo que las ponderaciones las hace el sistema automáticamente basándose en el comportamiento de los servidores. Esta técnica posee la ventaja de detectar posibles fallos en los servidores como pueden ser que un servidor se bloquee o que su conexión sea más lenta que la de los demás.

Least Connections (member) Least Connections (node): Distribuye las peticiones dependiendo de sus conexiones activas. Siendo a los servidores con menores conexiones a quienes se les pasa la nueva petición. Este método es más apropiado de utilizar cuando los servidores tienen capacidades de procesamiento similares, además se debe tener cuidado con la saturación de servidores ya que puede que un servidor tenga más conexiones tenga una capacidad mayor de procesamiento.

Weighted Least Connections (member) Weighted Least Connections (node): Distribuye las peticiones no solo dependiendo de las conexiones sino que además de la capacidad de procesamiento del servidor. Se utiliza como criterio de selección el % de capacidad que se calcula como las conexiones del servidor sobre su capacidad. La nueva petición se le asigna al servidor con menor % de capacidad. Este método sirve mejor para servidores con diferentes capacidades.

Predictive (member) Predictive (node): Observa el comportamiento de cada servidor y así saber cual tiene una mayor o menor rendimiento. Las nuevas peticiones se asignan a los servidores que tengan un mayor rendimiento. Esta técnica no es recomendada para grandes números de peticiones.