¿Buscando alta disponibilidad en sus servidores?

¿Buscando alta disponibilidad en sus servidores? En HostDime, siempre pendientes de estar a la zaga en los desarrollos tecnológicos del sector siempre estamos buscando romper nuestros paradigmas y apuntar cada vez más alto.

Los clientes cada vez tiene sitios webs más exigentes, las cargas de tráfico se hacen cada vez más y más grandes y siempre quieren mayor velocidad, tiempo al aire de sus servidores y exigencias de este calibre.

¿Qué son los servidores De Alta Disponibilidad?

Se trata de computadoras poderosas (gran procesador o procesadores, memoria RAM mayor que el promedio etc etc) que está conectada a una conexión a Internet de alta velocidad y alojada en un centro de datos remoto de última generación o en una instalación de datos optimizada.

Un servidor dedicado de alta disponibilidad es un sistema avanzado equipado con fuentes de alimentación redundantes, una red totalmente redundante (por decirlo de alguna forma, doblemente respaldada), torres de discos RAID y copias de seguridad, lo que garantiza el mayor tiempo de actividad y la máxima fiabilidad sin un solo punto de falla.

Otra forma de verlo es decir que un servidor web de alta disponibilidad es un concepto informático relativamente reciente, que incorpora mecanismos de conmutación por error, que disminuye la posibilidad que un fallo de alguno de sus componentes. Se trata de una suerte de seguro, de maximizar la eficiencia como tal del hardware.

Los sitios web de alto tráfico deben admitir cientos de miles, si no millones, de usuarios de manera rápida y confiable. Las organizaciones de TI admiten estos grandes volúmenes al agrupar servidores en un grupo (a veces denominado granja de servidores). Un equilibrador de carga se encuentra frente al grupo de servidores y enruta el tráfico a servidores individuales de una manera que maximiza la eficiencia. Con múltiples servidores que actúan como una sola unidad, las compañías pueden asegurarse de que si un solo servidor deja de funcionar, la solicitud del usuario se enruta al siguiente servidor disponible. De esta manera, los usuarios no reciben un error o falla absoluta. Esto se conoce como una configuración de alta disponibilidad .

La alta disponibilidad a menudo se mide en términos del porcentaje de tiempo que el servicio está disponible para los clientes. Una calificación de tiempo de actividad del 99.999%, es una medida ampliamente conocida.

¿Qué no es la alta disponibilidad?

La alta disponibilidad no significa que su sitio o aplicación nunca se caiga. Aunque proporciona una serie de cajas de seguridad y apunta a un tiempo de actividad del 99.999%, ningún sistema es perfecto. La disponibilidad seguirá dependiendo de la cantidad de componentes, sus ajustes de configuración y los recursos asignados a cada componente. Si bien las soluciones de alta disponibilidad ofrecen a su sitio o aplicación un mayor tiempo de actividad que un solo host, recuerde que el sistema solo es igual a la suma de sus partes.

Sin embargo, debido a que los sistemas de alta disponibilidad están formados por tantos componentes, pueden escalarse horizontalmente cuando sea necesario, mejorando así su capacidad para servir contenido.

¿Cómo Funciona El Equilibrio De Carga?

Un equilibrador de carga se encuentra frente a sus servidores y enruta las solicitudes de los visitantes a través de los servidores. Asegura una distribución uniforme, es decir, todas las solicitudes deben cumplirse de manera que maximice la utilización de la velocidad y la capacidad de todos los servidores y ninguno de ellos esté sobreutilizado o subutilizado.

Cuando sus clientes visitan su sitio web, primero se conectan al equilibrador de carga y el equilibrador de carga los dirige a uno de los servidores web de su infraestructura. Si algún servidor se cae, el equilibrador de carga redirige instantáneamente el tráfico a los servidores en línea restantes.

A medida que aumenta el tráfico web, puede agregar nuevos servidores de forma rápida y sencilla al grupo existente de servidores con equilibrio de carga. Cuando se agrega un nuevo servidor, el equilibrador de carga comenzará a enviar solicitudes al nuevo servidor automáticamente. Así es, no se requiere intervención del usuario.

Tipos De Equilibrio De Carga

El equilibrio de carga se puede realizar con uno de los siguientes métodos.

Equilibrio de carga a través de DNS

Equilibrio de carga a través del hardware

Equilibrio de carga a través del software

Equilibrio De Carga Con DNS

El servicio DNS equilibra el tráfico web en los múltiples

servidores. Tenga en cuenta que cuando realiza el equilibrio de carga de tráfico a través de este método, no puede elegir qué algoritmo equilibrará la carga.

Equilibrio De Carga A Través Del Hardware

Esta es la forma más costosa de equilibrio de carga. Utiliza un dispositivo de hardware dedicado que maneja el equilibrio de carga de tráfico.

La mayoría de los sistemas de equilibrador de carga basados en hardware ejecutan una distribución Linux integrada con una herramienta de administración de equilibrio de carga que permite un fácil acceso y una descripción general de la configuración.

Equilibrio De Carga A Través Del Software

El equilibrio de carga basado en software es uno de los métodos más confiables para distribuir la carga entre servidores. En este método, el software equilibra las solicitudes entrantes a través de una variedad de algoritmos.

Algoritmos De Equilibrio De Carga

Hay varios algoritmos que se pueden usar para lograr el equilibrio de carga en las solicitudes entrantes. La elección del método de equilibrio de carga depende del tipo de servicio, el tipo de equilibrio de carga, el estado de la red y sus propios requisitos comerciales.

Típicamente, para sistemas de baja carga, bastarán los métodos simples de equilibrio de carga (es decir, Round Robin), mientras que, para sistemas de alta carga, se deben usar métodos más complejos. Consulte este enlace para obtener más información sobre algunos Algoritmos de equilibrio de carga estándar de la industria utilizados por los equilibradores de carga.

¿Cuándo es importante la alta disponibilidad?

Al configurar sistemas de producción robustos, minimizar el tiempo de inactividad y las interrupciones del servicio a menudo es una alta prioridad. Independientemente de cómo sus sistemas y software confiables son, pueden ocurrir problemas que pueden derribar sus aplicaciones o sus servidores. La implementación de alta disponibilidad de la infraestructura es una estrategia útil para reducir el impacto de este tipo de eventos. Los sistemas de alta disponibilidad pueden recuperarse de la falla del servidor o componente automáticamente.

¿Qué hace que un sistema esté altamente disponible?

Uno de los objetivos de alta disponibilidad es eliminar puntos únicos de falla en su infraestructura. Un único punto de falla es un componente de su pila de tecnología que causaría una interrupción del servicio si no estuviera disponible. Como tal, cualquier componente que es un requisito para el funcionamiento correcto de la aplicación que no tiene la redundancia es considerado como un único punto de fallo.

Para eliminar puntos únicos de falla, cada capa de su pila debe estar preparada para la redundancia. Por ejemplo, imagine que tiene una infraestructura que consta de dos servidores web idénticos y redundantes detrás de un equilibrador de carga. El tráfico proveniente de los clientes se distribuirá por igual entre los servidores web, pero si uno de los servidores se cae, el equilibrador de carga redirigirá todo el tráfico al servidor en línea restante.

La capa del servidor web en este escenario no es un único punto de falla porque los componentes redundantes para la

misma tarea están en su lugar

El mecanismo en la parte superior de esta capa (el equilibrador de carga) puede detectar fallas en los componentes y adaptar su comportamiento para una recuperación oportuna

Pero, ¿qué sucede si el equilibrador de carga se desconecta?

Con el escenario descrito, que no es raro en la vida real, la capa de equilibrio de carga en sí misma sigue siendo un punto único de falla. Sin embargo, eliminar este único punto de falla restante puede ser un desafío; Aunque puede configurar fácilmente un equilibrador de carga adicional para lograr la redundancia, no hay un punto obvio por encima de los equilibradores de carga para implementar la detección y recuperación de fallas.

La redundancia por sí sola no puede garantizar una alta disponibilidad. Debe existir un mecanismo para detectar fallas y tomar medidas cuando uno de los componentes de su pila no esté disponible.

La detección y recuperación de fallas para sistemas redundantes se puede implementar utilizando un enfoque de arriba a abajo: la capa superior se hace responsable de monitorear la capa inmediatamente debajo de ella para detectar fallas. En nuestro escenario de ejemplo anterior, el equilibrador de carga es la capa superior. Si uno de los servidores web (capa inferior) no está disponible, el equilibrador de carga dejará de redirigir las solicitudes para ese servidor específico.

Si esta es su necesidad puntual, agende hoy mismo una cita con nuestro personal técnico especializado para encontrar soluciones a sus requerimientos.

Ver otros recursos valiosos del blog al respecto: [¿Qué es la](#)

tolerancia a fallas y por qué difiere de la alta disponibilidad? ; ¿Porqué se cae un servidor web o una página web? ; Cómo la disponibilidad de datos está cambiando las necesidades comerciales