

TUTORIAL DE CONTENT DELIVERY NETWORK (CDN)

Néstor Mata Cuthbert

nestor.mata@gmail.com

<http://nestor.profesional.co.cr/es/2009/12/03/tutorial-content-delivery-network-cdn-1469>

INTRODUCCIÓN

Hace años las computadoras no solían tener la cantidad de recursos que tienen actualmente, por lo que se debía ser muy cuidadoso con el uso responsable y eficiente de los servidores.

Con el tiempo y el aumento de los recursos en los sistemas muchos programadores fueron creciendo en un ambiente en el que no requerían ser tan eficientes, pero actualmente los sistemas se han vuelto más móviles y por tanto más pequeños y no tan potentes como lo que estábamos acostumbrados, y también los sistemas han migrado mucho más enfocados a web por lo que la cantidad de usuarios accediendo a los servidores es mucha.

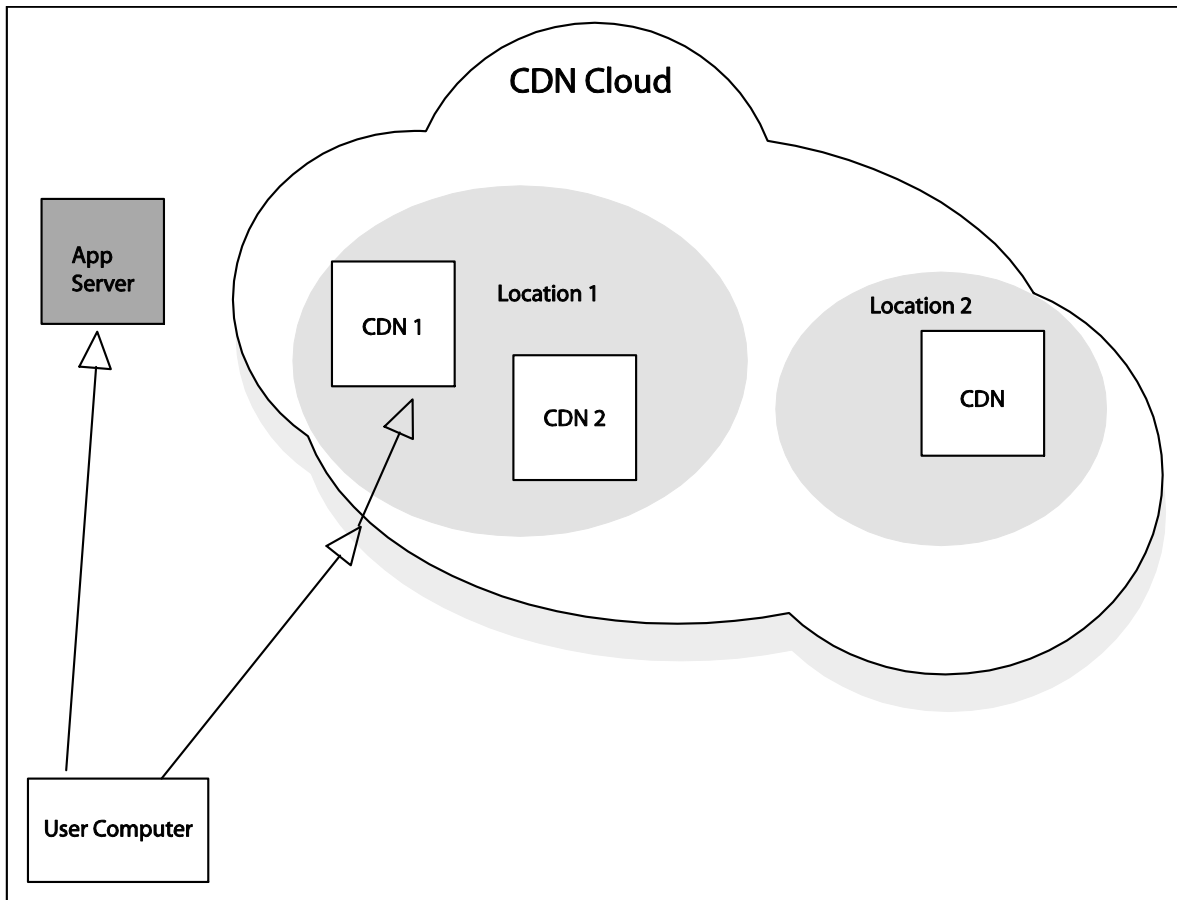
Esto nos trae a que en la actualidad estemos forzados otra vez (por dicha) a tener la necesidad de ser eficientes en cuanto a la utilización de los recursos.

En un sistema web de alta demanda la gran cantidad de recursos de un servidor se ve dividida entre la cantidad de usuarios accediendo simultáneamente, lo cual crea una cantidad limitada de recursos por cada usuario.

QUÉ ES CDN?

Una red de distribución de contenido es una colección de servidores web distribuidos alrededor de múltiples localizaciones para distribuir contenido más eficientemente a los usuarios. El servidor seleccionado para distribuir el contenido a un específico usuario se basa normalmente en la medida de proximidad de la red.

Los CDN's son servidores utilizados específicamente para distribuir contenido estático, siendo este el más sencillo de distribuir independiente de la plataforma, el conjunto de servidores se visualiza como una nube, los datos se encuentran distribuidos y replicados dentro de esta nube de servidores de manera que cualquiera de los servidores pueda entregar el mismo contenido cuando se le solicite.



POR QUÉ DEBERÍA USAR CDN?

Al utilizar CDN's se le entrega al usuario el contenido estático de manera más eficientemente, en beneficio de la velocidad con que el usuario final percibe la respuesta total de la página.

Además de mejorar la experiencia del usuario, al utilizar un CDN para distribuir el contenido estático se está a su vez liberando a él o los servidores de aplicaciones de la carga de procesar una gran parte de estas solicitudes que normalmente son más del 80% o 90% de la cantidad de solicitudes que se hacen al servidor, permitiendo así al servidor de aplicaciones utilizar toda su capacidad para procesar las solicitudes dinámicas.

FUNCIONAMIENTO DE LOS SERVIDORES WEB

Independientemente de la plataforma, los servidores web trabajan en memoria y debido a que los recursos de los servidores son limitados el servidor solo puede procesar un número máximo de solicitudes eficientemente y un número máximo de solicitudes antes de colapsar o de dejar de procesarlas.

Cada solicitud que ingresa al servidor es atendida en una cola la cual puede crecer exponencialmente si el servidor no es capaz de ir liberando la cola con la misma velocidad con la que ingresan solicitudes a esta cola.

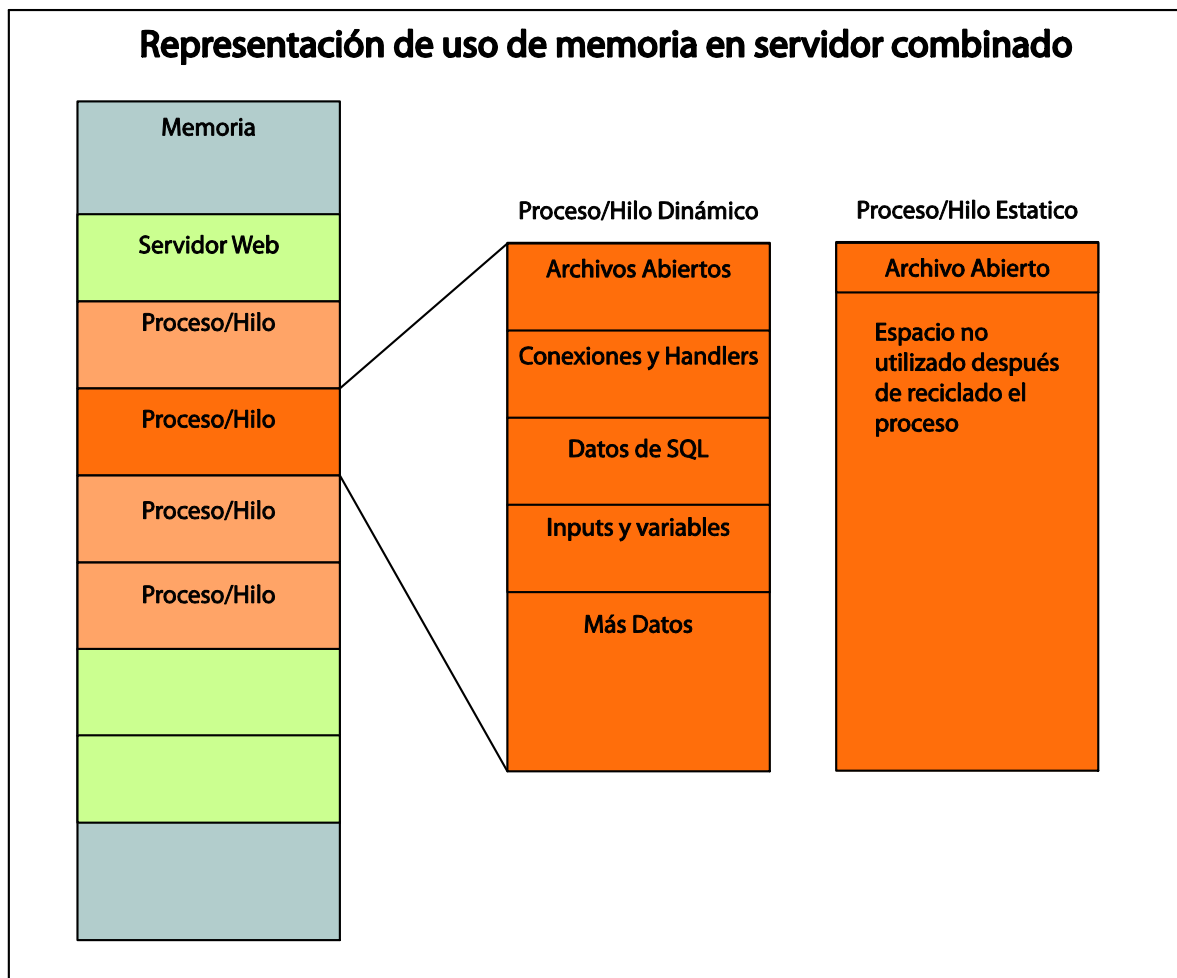
Hay solicitudes que son más fáciles de resolver para el servidor que otras, esto normalmente depende de la cantidad de archivos y handlers que deba abrir para esa solicitud y si lo que requiere esta o no en un cache. Dentro de los handlers están también las conexiones a bases de datos y/u otros sistemas, como sockets o solicitudes a servidores externos.

Dependiendo de la cantidad y tipo de recursos que se están accediendo además de recursos toma tiempo, lo que hace que las solicitudes en cola tengan que esperar más acumulándose aún más el tiempo que estas y las que siguen deberán esperar.

Además de esto, existen filtros o procesos previos o posteriores a cada solicitud que agregan tiempo y consumo de recursos a cada uno de las solicitudes.

Por el costo que tiene crear e inicializar cada recurso nuevo, los servidores además suelen crear pilas de recursos, como pilas de conexiones a bases de datos o pilas de procesos pre-inicializados para atender a las solicitudes.

Cada uno de estos procesos tomará la cantidad de memoria que requiera para efectuar lo que la solicitud requiere y si este proceso se recicla es probable que mantenga su uso de memoria aún cuando no todos los procesos que atiende requieran esa misma cantidad de memoria, esto conlleva que cuando se sirven diferentes tipos de solicitudes mezcladas la cantidad de memoria que utilicen los procesos será el del proceso que más requiera memoria, lo cual limita aún más la capacidad de procesar solicitudes más sencillas en sacrificio de poder atender las solicitudes más complejas.



De aquí que cuando se separan los tipos de solicitudes en servidores diferentes queda cada servidor optimizado para ese tipo de solicitudes, permitiéndole trabajar de manera más eficiente y darle un uso mucho más eficaz a los recursos del servidor.

FUNCIONAMIENTO DEL ESQUEMA DE CDN

Para trabajar con un CDN en el concepto básico lo que se requiere es que las referencias hacia los archivos estáticos que se encuentran en la aplicación web estén apuntando hacia el CDN y que este contenga estos archivos.

Un CDN normalmente recibe los archivos y los replica dentro de los diferentes servidores y se balancea la carga entre los diferentes servidores.

Existen algunos servicios de CDN que simplifican un poco más el concepto extrayendo ellos los archivos del servidor dinámico ya sea la primera vez que se intenta acceder al archivo o posterior a que se vence el archivo.

CONCLUSIONES

Los CDN's nos ayudan a eliminar la carga de los servidores de aplicaciones dándoles la oportunidad de trabajar más eficientemente y para lo que fueron destinados, permite además que los exploradores de los usuarios obtenga los archivos estáticos de una manera más eficiente.

Debido a que los archivos estáticos son por mucho la mayoría de los archivos que suelen servirse en cada aplicación web, el uso de CDN's suelen dar un incremento en la eficiencia bastante dramático.