

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS



Redes de distribución de contenidos (CDN)

Autor:
Miguel Torres Eric Miguel

Profesor:
Salvador López Mendoza

Introducción

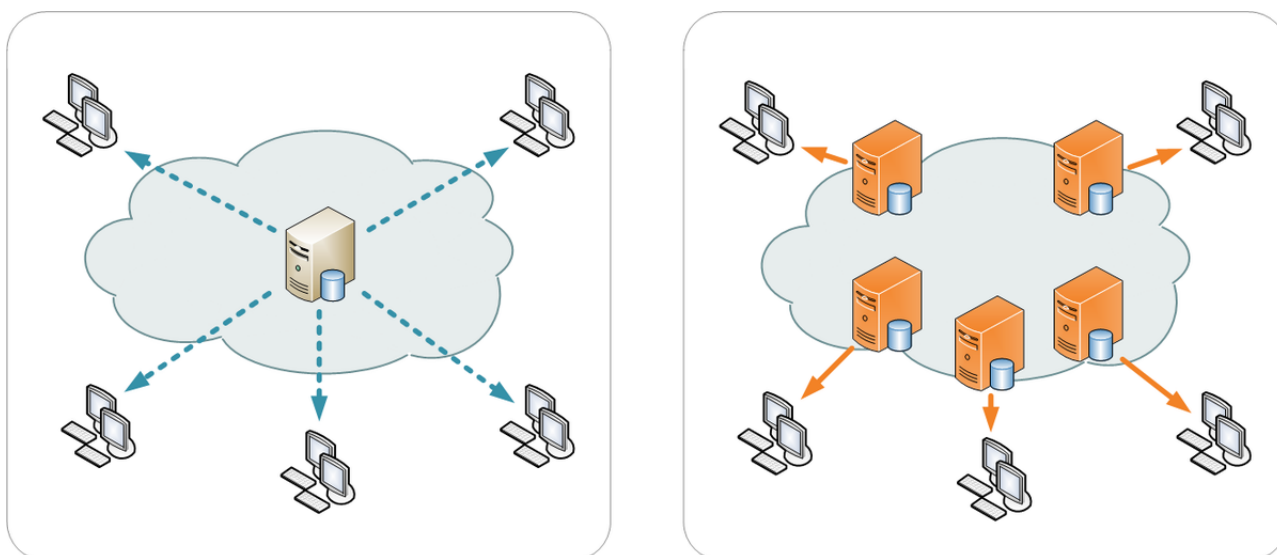
Las Redes de Distribución de Contenido comenzaron a ser relevantes a medida que la World Wide Web fue creciendo en popularidad durante la década de los 90. Los líderes técnicos notaron que el Internet no podía manejar el creciente nivel de tráfico de la red sin métodos más inteligentes para gestionar el flujo de datos.

En 1998, Akami Technologies fue la primera empresa en construir un negocio a gran escala alrededor de las CDNs. Otros siguieron con diferentes grados de éxito. Más tarde, algunas compañías como AT&T, Deutsche Telekom y Telstra también construyeron sus propias CDNs. Estas redes hoy en día llevan una porción significativa del contenido de la web, especialmente archivos grandes como vídeos, descarga de aplicaciones, etc.

Tras este crecimiento, notamos como la distribución de contenido multimedia tradicional basada en un servidor central no escala apropiadamente para un elevado número de usuarios, las grandes redes no disponen de forma nativa de mecanismos de distribución a nivel de red ni reserva de recursos anticipada, por lo que el desarrollo más veloz y al alcance de la comunidad científica y empresarial sugiere el empleo de redes de nivel de aplicación.

Con la pandemia, esto incrementó aún más, pues todos los procesos se comenzaron a digitalizar. Desde la entrega de tareas escolares hasta la entrega de papeles oficiales por parte de alguna dependencia de gobierno, pedir préstamos sin ir a sucursal. Y no solo eso, sino que día a día utilizamos medios digitales como ver vídeos, películas, escuchar música, ver historias de Facebook, Instagram, Twitter. Y todas estas, están controladas por CDNs.

Las CDNs son muy importantes en el día a día, y a continuación presentaremos cómo funcionan, qué hacen, los retos a los que se enfrentan y cómo los solucionan a través de Protocolos de Red y sus respectivos Servicios.

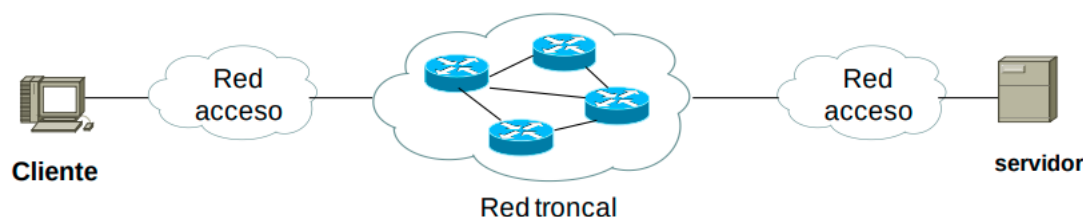


¿Por qué una CDN?

Antes de pasar a explicar lo qué es una CDN, empecemos viendo las motivaciones que llevaron a la industria a crearlas.

El servicio Web ha evolucionado proporcionando una infraestructura y plataforma de desarrollo para un gran número de aplicaciones distribuidas, que se pueden agrupar en cuatro tipos fundamentales:

1. Intercambio de contenido estático: se trata de aplicaciones donde el contenido no varía con el tiempo, o lo hace poco frecuentemente (por ejemplo páginas web personales o artículos académicos). No es más que la extensión natural de las aplicaciones anteriores con un poco de interactividad.
2. Intercambio de contenido dinámico: consiste en ofrecer contenido personalizado, es decir, dicho contenido se crea en el momento de la solicitud dependiendo de varios criterios. Casos de uso de contenido dinámico incluyen portales que proporcionan noticias, cotizaciones en bolsa, previsión del tiempo, todas estas basadas los intereses y la localización del usuario. Cabe destacar en la definición de contenido dinámico que un servidor web puede entregar un contenido distinto a usuarios individuales solicitando el mismo recurso web. Esto lo hace diferente de un contenido estático que varíe frecuentemente.
3. Intercambio de contenido Streaming: considerado, en ocasiones, como la reproducción de un flujo continuo de audio o vídeo. Una descripción más acertada distingue entre un Streaming real y la simple reproducción de ficheros de audio o video previamente descargados. De hecho, el concepto de streaming puede aplicarse a cualquier contenido, como texto e imágenes, aunque típicamente se asocia a sonido y vídeo. Ejemplos de este tipo de aplicaciones incluyen vídeo bajo demanda y radio sobre Internet.
4. Colaboración interactiva: Esta consiste en ofrecer un entorno de colaboración con un mayor grado de interactividad a través del web en tiempo real. Típicamente los entornos previamente existentes estaban limitados a actividades asíncronas, sin capacidad de interacción en tiempo real. Aplicaciones web como videoconferencia, juegos en red y mensajería instantánea permiten a los usuarios reaccionar ante acciones de otros usuarios en tiempo real. Tanto las empresas como los usuarios finales perciben el Web como una infraestructura fiable y de calidad capaz de entregar contenido multimedia enriquecido. Sin embargo, recientes desarrollos en los campos de contenido multimedia, aplicaciones interactivas y contenido dinámico ponen de manifiesto una falta de rendimiento del modelo tradicional web y ha conducido al sector académico e industrial a incorporar mejoras en las tecnologías de red sobre Internet, denominadas content networks. Este último concepto da origen a los CDN.



Redes de Distribución de Contenidos (CDN)

Una Red de Distribución de Contenidos, abreviado como CDN por sus siglas en inglés Content Delivery Network, es una red de servidores en diferentes centros de datos de todo el mundo que a grandes rasgos, almacenan de manera temporal una copia de los sitios web. A diferencia de los métodos tradicionales, para guardar tu sitio web en un lugar (aquel lugar en el que se encuentre el centro de datos de la empresa del alojamiento web), una CDN se comporta como si el sitio web estuviera alojada en diferentes lugares sin que sea necesario sincronizar manualmente el contenido del sitio entre todos los hosts. Las CDNs son sistemas distribuidos que permiten un control y monitorización total de los recursos existentes, y cuya tendencia actual es la distribución de streaming media.

El control de recursos ofrecido por una CDN permite no sólo una escalabilidad del servicio sino también una elevada eficiencia energética. En los últimos años, las técnicas de virtualización y el concepto de nube permiten gestionar recursos de red bajo demanda de tal forma que se puede minimizar el consumo energético necesario. Si bien las CDNs pueden comprender una gran cantidad de servidores dispersos geográficamente, el consumo energético de estos no es constante sino que varía dependiendo de las exigencias del tráfico demandado en cada momento. Esto permite minimizar el consumo energético cuando el número de usuarios es reducido, maximizando la viabilidad económica de estos sistemas.

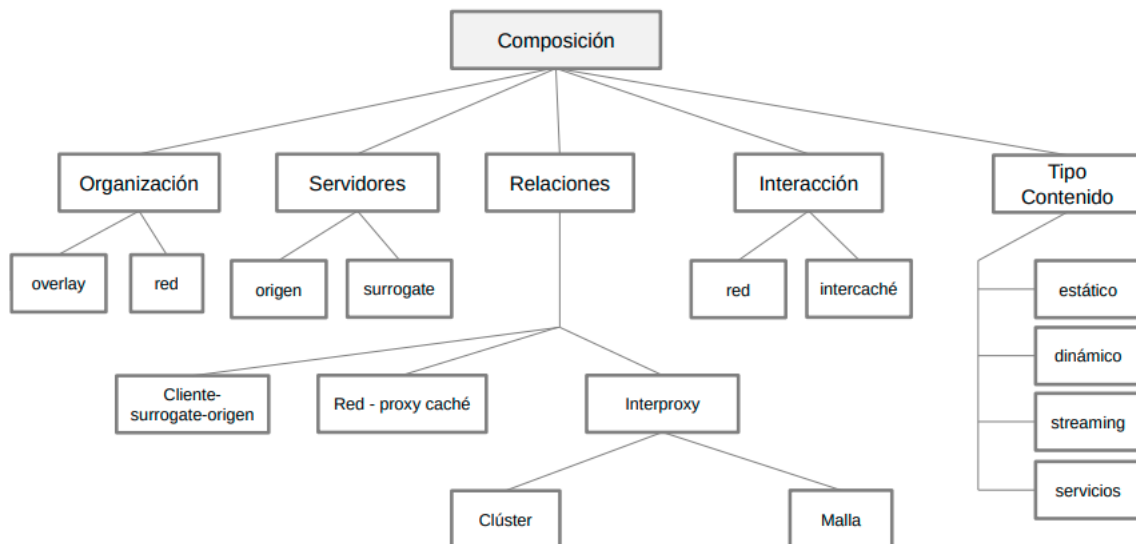
La primera ventaja de utilizar una CDN es la rapidez. Los tiempos de carga para tu sitio web se reducen considerablemente, ya que cada visitante accede a los datos del servidor que le queda más cerca. Por ejemplo, si tenemos un centro de datos en Alemania para algún dominio, y realizamos desde México una petición a dicho centro, tardará mucho en llegar la señal de México a Alemania, sin embargo, al tener un CDN, esa página tendrá una copia en algún servidor de otro país, por ejemplo Estados Unidos. Entonces la señal que nosotros enviamos nos redirigirá a el CDN de Estados Unidos, lo cuál hará que la velocidad se reduzca hasta en la mitad de tiempo en cargar la web.

Llamaremos surrogate al servidor distinto al de origen. Este recibirá la petición mapeada del usuario y será el responsable de hacer la entrega del contenido al cliente.

Las funcionalidades básicas de una CDN son:

1. Mecanismo de redirección, para dirigir una solicitud al surrogate más cercano, utilizando mecanismos para evitar la congestión.
2. Servicios de distribución de contenido, para replicar o cachear contenido desde un servidor origen a los surrogates dispersos en Internet.

3. Servicios de negociación de contenido, para cubrir los requerimientos concretos de usuarios específicos (o grupos de usuarios).
4. Servicios de gestión, para administrar los componentes de red, gestionar la contabilidad y monitorizar y reportar el uso y acceso al contenido



Estructura de una CDN

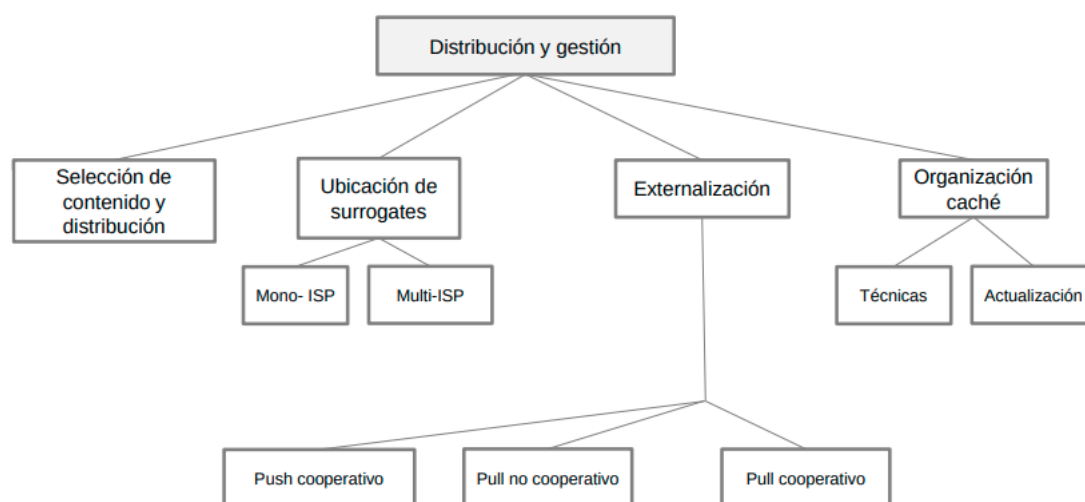
El análisis de los atributos estructurales de una CDN revela que los componentes de la infraestructura de una CDN están fuertemente relacionados. Es más, la estructura de una CDN varía dependiendo del contenido y de los servicios que proporcionan a los usuarios. Dentro de la estructura de una CDN, se emplea a un conjunto de surrogates para conformar el componente encargado de distribuir contenido, se emplea una combinación de relaciones y mecanismos para redirigir peticiones de clientes y se emplean protocolos de interacción para la comunicación entre elementos de la CDN.

En el aspecto de la organización, existen dos formas de crear una CDN. La primera forma es mediante una red overlay (o red de capa de aplicación) formada por un conjunto de servidores y proxy cachés que atienden las peticiones de ciertos tipos de contenido (web, media streaming, vídeo en tiempo real, etc.). Los componentes de la red troncal (routers y switches) no desempeñan ningún papel activo en la distribución de contenido, sino simplemente proporcionar conectividad básica de red y, en algunos casos, un cierto QoS para tráfico específico. La segunda forma de crear una CDN consiste en involucrar de manera directa a los componentes de red (routers, switches) de tal forma que mediante políticas configuradas en estos componentes se identifica ciertos tipos de aplicación y se encaminan las peticiones adecuadamente. Los ejemplos más claros de este tipo de CDNs serían aquellas que incluyen dispositivos que redirigen las peticiones a servidores caché o específicos, optimizados para servir un tipo de contenido concreto.

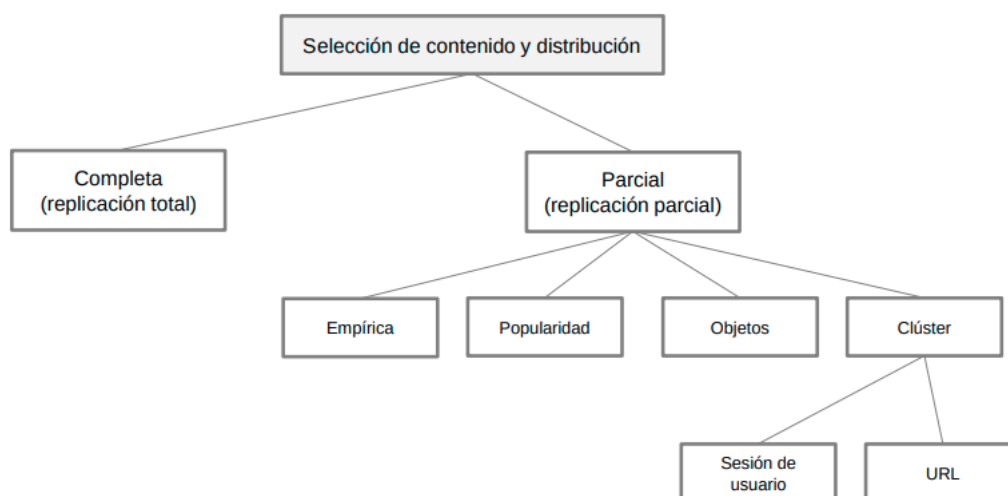
En cuanto a las relaciones, puede decirse que la arquitectura compleja y distribuida de una CDN puede llevar asociada varias relaciones entre sus componentes (clientes, surrogates, servidor origen, proxy cachés y otros elementos de red). La mayoría de estos componentes se comunican entre sí para replicar o cachear contenido dentro de la CDN. En la replicación, básicamente consiste en hacer ‘pushing’ de contenido desde el servidor origen a los surrogates. Por otro lado, caching implica almacenar respuestas cacheables para futuras peticiones. En un entorno CDN, la relación más básica en la distribución de contenido tiene lugar entre el cliente, los surrogates y el servidor origen. Cada cliente se comunica con uno o más surrogates que satisfacen (de forma transparente) las peticiones en lugar de uno o más servidores origen. Si no se emplean surrogates, el cliente se comunica directamente con el servidor origen.

Finalmente, en relación con el tipo de contenido, las CDNs realizan hosting de diferentes tipos de contenido (estático, dinámico, streaming), así como de diferentes servicios (servicio de directorio, servicio de transferencia de ficheros). Cada tipo de contenido o servicio tiene implicaciones en la arquitectura, por lo que algunas CDNs pueden centrarse sólo en algún tipo particular de contenido o servicio. El contenido estático hace referencia a aquel que no cambia frecuentemente, como pueden ser páginas HTML estáticas, imágenes, documentos PDF, ficheros, etc. Actualmente, todos los proveedores de CDNs soportan este tipo de contenido, ya que éste puede ser cacheado fácilmente y la consistencia se puede gestionar empleando mecanismos tradicionales de caching. El contenido dinámico hace referencia a aquel que está personalizado para el usuario o se genera bajo demanda, y suele cambiar frecuentemente, como pueden ser animaciones, scripts y HTML. Dada su naturaleza cambiante, este tipo de contenido no es cacheable. El contenido de streaming puede ser en vivo (live) o bajo demanda (on-demand), y algunos ejemplos son retransmisiones deportivas, películas a la carta, etc. Los servidores de streaming adoptan protocolos especializados (RTP, RTCP, RTSP, etc.) para la distribución de este tipo de contenido.

La distribución de contenido y la gestión constituyen un aspecto vital para una CDN eficiente.



Este aspecto representa uno de los parámetros más importantes que determinan la eficiencia en la distribución de contenido, ya que implica la reducción en el tiempo de respuesta percibido por los clientes. El contenido que se debe seleccionar para su replicación, presente en el servidor origen, se puede distribuir hacia los surrogates de forma total o parcial.



Organización y Gestión de una CDN

La correcta gestión del contenido es fundamental para un buen rendimiento en una CDN, que está principalmente basado en el esquema de organización de caché empleado, que abarca tanto las técnicas de caching como la frecuencia de las actualizaciones para asegurar disponibilidad y fiabilidad del contenido. Además, es posible el uso de forma integrada en la infraestructura de la CDN de caching y replicación con la finalidad de introducir mejoras potenciales en la latencia percibida, hit ratio y byte hit ratio.

Algoritmos de redirección

Los algoritmos de redirección se pueden clasificar básicamente en dos categorías:

1. Adaptativos: consideran las condiciones en tiempo real de la red y/o los servidores para decidir el surrogate más adecuado a partir de ciertas métricas (congestión en los enlaces, carga de los servidores, etc.)
2. No adaptativos: emplean ciertas heurísticas para tomar la decisión en lugar de considerar las condiciones del sistema en tiempo real.

Los algoritmos no adaptativos son más sencillos de implementar, ya que los adaptativos suelen ser más complejos al tener que adaptarse a las condiciones cambiantes de la red y/o de los servidores en tiempo real. Esto los hace más robustos que los no adaptativos, ya que sólo son eficientes cuando se cumplen las condiciones o asunciones de las heurísticas.

El algoritmo de encaminamiento no adaptativo más común y sencillo es uno de tipo round robin, que distribuye todas las peticiones a los surrogates de la CDN balanceando la carga entre ellos, asumiendo la misma capacidad para todos los servidores. Estos algoritmos sencillos resultan eficientes para clústeres, donde los servidores están ubicados físicamente en una misma localización. Sin embargo, no son adecuados en un entorno WAN donde los servidores están ubicados en lugares distintos, y un cliente podría ser redirigido a un

surrogate lejano, con una latencia percibida claramente mejorable. Por otro lado, el objetivo principal del balanceo de carga puede no quedar satisfecho ya que cada petición puede ser diferente y requerir cargas computacionales distintas.

En otro ejemplo de algoritmo no adaptativo, los surrogates se ordenan dependiendo de la carga esperada en cada uno de ellos. Dicha predicción está basada en el número de peticiones que cada servidor ha satisfecho, y considera también la distancia clienteservidor. De esta forma, se balancea la carga entre los servidores en base a estos dos parámetros. Si bien el algoritmo resulta eficiente, la latencia percibida por el cliente es mejorable.

Medidas de Rendimiento de una CDN

Las medidas de rendimiento en una CDN tienen en cuenta la capacidad de ésta para proporcionar a los clientes el contenido o servicio deseado. Generalmente se consideran cinco tipos de métricas por parte de los proveedores de contenido para evaluar el rendimiento de una CDN:

1. Cache hit ratio: se define como el ratio entre el número de objetos cacheados frente al total de objetos solicitados. Un valor elevado de esta métrica indica que la CDN está empleando una técnica efectiva de caching para gestionar sus proxy caches.
2. Ancho de banda reservado: hace referencia al ancho de banda empleado por el servidor origen.
3. Latencia: indica el tiempo de respuesta experimentado por el usuario. Un tiempo de latencia reducido indica que el servidor origen requiere menor ancho de banda.
4. Utilización de surrogates: indica la fracción de tiempo que el surrogate en cuestión está ocupado. Esta métrica permite al administrador calcular la carga de CPU, el número de peticiones servidas y el uso de almacenamiento E/S.
5. Fiabilidad: las medidas de pérdidas de paquetes se emplean para determinar la fiabilidad de una CDN. Una fiabilidad alta indica que apenas hay pérdidas de paquetes y la CDN siempre está disponible para los clientes. Las medidas de rendimiento se pueden obtener desde dos puntos de vista:
 - a) Medidas internas: los servidores de una CDN se equipan con capacidad de recoger estadísticas de uso con la finalidad de obtener una medida de rendimiento extremo a extremo. Adicionalmente, se pueden desplegar sondas (hardware o software) a través de toda la red y correlar esta información con los logs de los servidores.
 - b) Medidas externas: adicionalmente a las medidas internas, las medidas externas llevadas a cabo por terceros, como Keynote Systems permiten verificar con mayor fiabilidad para los clientes el rendimiento de una CDN. Este proceso es eficiente y fiable puesto que estos terceros soportan el benchmarking de redes dispersas, ya que ellos también disponen de servidores de prueba distribuidos por toda la red de Internet. Estos servidores miden el rendimiento de desde la perspectiva del cliente, considerando varias métricas.

6. Proximidad geográfica: implica identificar a un usuario dentro de una región. Se emplea para redirigir a todos los usuarios de una cierta región al mismo punto de presencia (PoP, Point of Presence). La medida de proximidad de red se deriva típicamente a partir de las tablas de encaminamiento del protocolo BGP (Border Gateway Protocol).
7. Latencia: la latencia percibida por el usuario es una métrica muy útil para seleccionar el surrogate más adecuado para dicho usuario.
8. Pérdida de paquetes: la contabilidad de pérdidas de paquetes a través de varias rutas de red permite seleccionar la ruta con menor tasa de pérdidas.
9. Ancho de banda medio, tiempo de arranque, tasa de tramas: se trata de métricas empleadas para seleccionar la mejor ruta para la distribución de contenido multimedia.
10. Carga del servidor: la carga de un servidor se puede estimar en base a métricas internas como carga de CPU, carga del interfaz de red, conexiones activas, y carga de almacenamiento. Esta métrica agregada permite seleccionar el surrogate con menor carga.

Conclusión

En este escrito se ha partido del crecimiento del tráfico en Internet y sobretodo en la actual pandemia para hablar de como las redes deben escalar y mejorar los mecanismos de distribución de contenidos en general, esto con la finalidad de ofrecer mejores servicios a los usuarios en Internet. Notamos como la calidad en Internet está relacionada con la capacidad de ofrecer un tiempo reducido de espera a los clientes cuando contactan con un servidor, lo que no es suficiente para algunos tipos de aplicaciones actualmente desplegados.

Para abordar este aspecto, se han propuesto numerosas tecnologías desde prácticamente el comienzo de Internet. Así, el caching, aplicado al web y, más recientemente, a objetos multimedia, es capaz de reducir el tiempo de latencia percibido por los usuarios.

Las redes de distribución de contenidos (CDNs) representan una evolución dentro de las redes de caché y los espejos, integrando ambas tecnologías de una forma eficiente dentro del concepto de content networking. El objetivo principal es reducir el tiempo de espera del usuario mediante la redirección a un servidor cercano, denominado surrogate, que implementa tanto la funcionalidad de proxy cache como de espejo. Evidentemente, en un entorno de Internet, el número de surrogates a desplegar debe ser muy alto para minimizar el efecto de la latencia.

Dado este y muchos otros problemas, algunos tipos de aplicaciones desplegadas en servidores centralizados no son capaces de ofrecer este tipo de calidad, y se han ido produciendo novedades a lo largo de los últimos años.

Por otro lado, los sistemas de réplicas, mirrors o espejos duplican la información almacenada en un servidor origen en otros servidores remotos. El objetivo principal es descargar al servidor origen de un exceso de peticiones, además de aumentar la disponibilidad de los datos.

Finalmente, hablemos de como las CDNs son una evolución muy grande que nos ha permitido solventar nuestras necesidades, pero como todo en este mundo, nada es perfecto, pues lleva consigo muchas complicaciones de duplicación de información, falta de servidores a lo largo del mundo. Aun así, el creciente avance en las nuevas tecnologías, nos permiten llegar a soluciones que si bien, no son perfectas, son suficientemente estables para ser utilizadas y que todos los usuarios tengan una respuesta de cualquier servidor en el mundo muy rápida.

0.1. Bibliografía

1. Ionos. (s.f). *¿Qué es una red de distribución de contenidos (CDN)?*. Recuperado el 23 de enero del 2022 de <https://www.ionos.mx/ayuda/hosting/content-delivery-network-cdn/que-es-una-red-de-distribucion-de-contenidos-cdn/>
2. RYTE Wiki. (s.f). *Red de distribución de contenidos (CDN)*. Recuperado el 23 de enero del 2022 de https://es.ryte.com/wiki/Red_de_Distribuci%C3%B3n_de_Contenidos
3. Kionetworks. (s.f). *Red de distribución de contenidos*. Recuperado el 24 de enero del 2022 de <https://www.kionetworks.com/blog/data-center/red-de-distribucion-de-contenido>
4. Encarna Pastor. *Sistemas de distribución y entrega de contenidos. Nuevos paradigmas.* (2005)
5. Akas, Nicholas (29 de noviembre de 2011). *How content delivery networks (CDNs) work?*. NCZOnline
6. Tecnonautas. (s.f). *La historia de las redes de distribución y entrega de contenidos (CDN)*. Recuperado el 24 de enero del 2021 de <https://tecnonautas.net/la-historia-de-las-redes-de-distribucion-y-entrega-de-contenidos-cdn/#:~:text=con%20las%20CDNs-,Historia%20de%20las%20CDNs,durante%20la%20d%C3%A9cada%20de%201990.&text=Fundada%20en%201998%2C%20Akamai%20Technologies,escala%20alrededor%20de%20las%20CDNs>
7. Lopez, A. (28 de febrero del 2021). [CDN] ¿Qué es una CDN? Red de Distribución de Contenido — Content Delivery Network — Alberto López. Recuperado el 24 de enero del 2022 de <https://www.youtube.com/watch?v=6JDpkH3fMUw>
8. BBVA. (1 de septiembre del 2021). *Redes de entrega de contenido (Content Delivery Networks)*. Recuperado el 24 de enero del 2022 de <https://www.bbvaopenmind.com/tecnologia/mundo-digital/redes-de-entrega-de-contenido-cdns/>
9. Molina, B. (2013). *Estudio, análisis y desarrollo de una red de distribución de contenido y su algoritmo de redirección de usuarios para servicios web y streaming*