UNIVERSIDAD TECNICA
FEDERICO SANTA MARIA

DEPARTAMENTO DE
ELECTRONICA

Proyecto ELO 322: Funcionamiento de YouTube

Integrantes: Felipe Córdova
Javiera Díaz
Diego San Martín

INTRODUCIÓN

En este proyecto abordaremos muy brevemente de cual es el funcionamiento de YouTube intrínsecamente, revisando los temas que más nos ponen en duda o los que nos dio más curiosidad. Empezando con un poco de su origen en la Internet y su evolución. Las plataformas que utiliza para poder funcionar, que es lo que hace para poder almacenar tantos videos y no colapse. También revisaremos que sucede con los videos que son muy solicitados por muchos usuarios al mismo tiempo y el tema sobre las imágenes miniatura de los videos. Así como también para que tener un cache en el servidor y como solucionar los cuellos de botella.

HISTORIA

YouTube fue fundada por Chad Hurley, Steve Chen y Jawed Karim en febrero de 2005 en San Bruno, California. Todos ellos se conocieron cuando trabajaban en PayPal. Hurley y Karim como ingenieros, y Chad como diseñador.

El dominio fue activado el 15 de febrero de 2005, y el 23 de abril fue cargado el primer vídeo, que se llamó “Me at the Zoo” (Yo en el zoológico).

Para octubre de 2006 las oficinas de YouTube permanecían en el condado de San Mateo, y contaba con unos 60 empleados. Un ejecutivo de Universal Music Group había anunciado hace poco que “YouTube les debía decenas de millones de dólares”, por violación de derechos de autor.

Mark Cuban, cofundador de www.Broadcast.com, en una radio por Internet comprado por Yahoo en 1999, había declarado un mes antes que “Sólo un tonto compraría YouTube por los potenciales problemas legales que enfrentaría”. Pero los propietarios del sitio ya se habían comprometido con Warner Music para mejorar el servicio, de manera tal que pudiesen detectar más rápidamente cuándo un vídeo de su propiedad era cargado al sitio. No obstante, a inicios de ese mes se publicó un reporte que indicaba que Google iba a comprar YouTube por 1600 millones de dólares. Dicha información inicialmente fue negada por YouTube y Google, que la calificaron como rumores.

Ese mismo mes Google compró YouTube por 1650 millones de dólares en acciones. En el momento de la compra, 100 millones de vídeos en YouTube eran visualizados y 65 mil nuevos vídeos era añadidos diariamente. Además, unos 72 millones de personas la visitaban por mes. Hurley y Chen mantuvieron sus cargos, al igual que los 60 empleados que en ese momento trabajaban en la empresa. En los días anteriores, YouTube había firmado dos acuerdos con Universal Music Group y la CBS; y Google había firmado acuerdos con Sony BMG y Warner Music para la distribución de vídeos musicales.

Para junio de 2008 el 38% de los vídeos visualizados en Internet provenían de YouTube; el competidor más cercano sólo llegaba a representar el 4%. Aunque Google no reveló las cifras, se estimó que el sitio generó 200 millones de dólares estadounidenses ese año.

Y así YouTube siguió hasta conseguir un crecimiento muy importante, donde en la actualidad se sirven alrededor de 100 millones de videos diariamente.

PLATAFORMAS DE YOUTUBE

1. Sistema Operativo:

Usa la versión SUSE de Linux que es una de las más conocidas distribuciones Linux existentes a nivel mundial, se basó en sus orígenes en Slackware. Entre las principales virtudes de esta distribución se encuentra el que sea una de las más sencillas de instalar y administrar, ya que cuenta con varios asistentes gráficos.

2. Sistema Web:

Usa Apache que es una programa para crear una servidor web HTTP de código abierto que se puede crear desde cualquier ordenador y que permite que otros ordenadores entren a él y se conecten. Es el software más grande usado para esto. Youtube usa la versión lighttpd de apache, que está diseñada para ser más rápido, seguro, flexible y fiel a los estándares, ya que consume menos CPU y memoria RAM que otros servidores. Es el más apropiado para servidores que tienen problemas de carga.

3. Lenguajes de programación:

El lenguaje de programación que usa es Python y C, donde Python permite un desarrollo más rápido y flexible. Usa además Psyco que es un compilador C que se puede importar desde Python y que acelera notablemente las aplicaciones Python.

4. Base de datos:

Usa MySQL de código abierto, es la más usada para este tipo de servidores aquí es donde se almacenan los nombre de usuarios, descripciones, etiquetas, tags, etc.

YouTube también ocupa Netscaler que es un dispositivo de entrega de aplicaciones web que optimiza la disponibilidad de las aplicaciones, acelera el rendimiento y se preocupa del balanceo de carga. Netscaler también actúa como cortafuergos para protegerlo de ataques.

ALMACENAMIENTO

Para ocupar bien los recursos de almacenamiento es fundamental para el correcto el funcionamiento de este medio, para esto se construyo un tipo de FAT, un FAT es una estructura lógica que permite ordenar los datos almacenados en un medio de almacenamiento: discos duros, Dikettes, DVD, CD, etc.), en este caso especial para almacenar los videos en los discos duros. En el caso de YouTube utilizan una FAT que segmenta el espacio de forma más pequeña de lo habitual. La cual permite dividir un video en miles de pedazos pequeños que se guardarán en distintos sectores de la unidad física.

Esta manera simple de organización de videos tiene las ventajas de ahorrar espacio en el disco duro, optimizar la lectura de datos y los medios de almacenamiento de forma más eficaz en la fragmentación. También para cada video que es segmentado en fragmentos es guardado en varios discos duros, de modo que, una máquina puede estar en mantenimiento o puede fallar sin afectar a los usuarios, y esto también permite mayor velocidad y tolerancia a fallos. Además los videos poseen un sistema on-line de respaldos.

VIDEOS ALTA DEMANDA

El siguiente problema fue la alta demanda de videos ya que requiere un ancho de banda y una capacidad de procesamiento grande, también generan efecto long tail con mucho volumen de tráfico y no conviene cachearlos pues sale caro, pero mayor aún debe ser la capacidad de escalar y hacer frente a los cuellos de botella. Es imprescindible mantener una estructura lo más llana posible, con ordenadores de consumo para reducir costes y la importancia del fallo de uno de los equipos, la solución del problema fue llevar contenidos más populares a un CDN, un CDN es un gran sistema distribuido de servidores desplegados en múltiples centros de datos en Internet, que replica el contenido y está más cerca (en cuanto a saltos de red) del ordenador del usuario. La ventaja de ocupar una red de distribución de contenidos es la de servir de contenido a los usuarios finales con una alta disponibilidad y alto rendimiento.

IMÁGENES DE VIDEOS

Gran parte de los problemas venían de las imágenes en miniatura (thumbnails), era muy difícil ser eficientes ya que al ser cuatro imágenes por cada video la cantidad de imágenes es grande y como son archivos pequeños están en unos pocos servidores, al final el sistema se tardaba de 6 a 10 horas en cachear un equipo y el número de lecturas y escrituras era insuficiente para buscarlos, la solución a este problema fue crear una base de datos noSQL Google Big Table que permite tener muchos elementos y su acceso es mucho más rápido, así pudieron tener un sistema redundante y con alta tolerancia a fallos.

CACHE DE YOUTUBE

Lo que nosotros vemos en YouTube funciona precacheando la mayoría de los datos. Esto quiere decir que los datos HTML, se encuentran cargados en las memorias de los servidores de YouTube y por ende son de rápida respuesta. Mientras el sistema esté activo esos datos permanecen en memoria y de esta manera es más fácil y ágil responder a los usuarios.

Otro dato interesante es que la arquitectura del sistema funciona de manera que guarda en memoria cache de alta respuesta, los datos del sistema que con más frecuencia se solicitan o se muestran. Cuando se detecta que un sector comienza a ser menos requerido por los usuarios u otras partes del sistema, los datos se descargan de la memoria, y dejan paso a otras funciones.

CUELLOS DE BOTELLA

Es un embotellamiento de paquetes de datos que circulan por una conexión causando demoras en la comunicación. Desde sus inicios, YouTube entró en una gran cantidad de cuellos de botella ante el increíble éxito que alcanzaban cada mes. Para escalar tenían que mantenerlo sencillo y barato el sistema, utilizando hardware de consumo que como la tónica general en estas plataformas y la mínima cantidad de nodos posibles. El objetivo era soportar el volumen de video a procesar y el ancho de banda de los videos de mayor éxito. Para esto YouTube hace una iteración continuamente que es que mientras funcione YouTube y si existen cuello de botella los libera creando un balanceo de la redes este tipo de seudo código que explica a grandes rasgos que ante una saturación en las líneas o el sistema que deriva o canaliza la carga o usuarios, hacia otros sistemas alternativos. Entonces gracias a esto el sistema siempre continúa funcionado correctamente.

CONCLUSIÓN

En conclusión en este proyecto pudimos sintetizar de cual es el funcionamiento de YouTube internamente y logramos sacarnos la curiosidad de los distintos temas, como su gran crecimiento de popularidad y que cada vez aumenta más. También conocimos las plataformas que utiliza para su funcionamiento, que gracias a un FAT puede almacenar los videos y con un CDN para poder solucionar el problema de los videos de alta demanda. Además, las imágenes miniatura de los videos como se creaban cuatro por video crearon la base de datos noSQL Google Big Table para almacenarlas. Finalmente aprendimos que el servidor de YouTube usa un cache para facilitar agilidad al enviar los videos y que para solucionar los cuellos de botella se necesita la iteración mencionada.