YouTube Technology Stack
By: Matt Maiville

Company Introduction

The video distribution superpower, YouTube, provides users a platform to upload and view billions of videos. Currently YouTube is surpassing one billion views per day, technological support for a demand this great requires a strong and stable tech stack that we will delve into in the rest of the article.

Web Servers

For their web servers YouTube utilizes NetScalar servers for their abilities in load balancing and caching static content. The servers run Apache with the Fast CGI protocol to improve performance and persistence, while avoiding the limitations faced by some server specific API's. The servers also route requests for handling using a Python application server, because Python allows for rapid flexible development and deployment, an important aspect because if YouTube's service was down for long periods during updates they risk losing customers to competitors. YouTube's servers also use psyco, which is a dynamic Python C compiler. Psyco uses a JIT compiler approach, which optimizes inner loops and allows for significantly faster code processing. The servers also use for CPU intensive activities like encryption, C extensions to combine the power of Python and C, to increase the speed and efficiency of the servers.

Database Servers

During the early years of YouTube, the company utilized databases that ran on MySQL to store Meta data, for users, tags and descriptions. YouTube leased all their hardware and faced significant lag so they prioritized traffic into a video watch and a general information cluster, to help better support the video playback the main function of the site.

Today YouTube uses database partitioning, split into shards with different users being assigned to different shards. This structure spreads writes and reads, and provides significantly better cache locality. This resulted in a 30% hardware reduction, and reduced replica lag to zero.

Special Software

Distributing video simply, effectively, and efficiently is the most important goal for YouTube. If their customers are unsatisfied with delays or load times, they may be inclined to go elsewhere for a video. YouTube as the top video distribution company understands this and institutes a framework to speed video serving. Servers storing video use the lighttpd web server, switching from a single process to a multiple process configuration to handle more connections, and to avoid the high overhead of Apache.

Two other useful resources used by YouTube for video serving is having every video hosted by a mini-cluster, to ensure the failure of one machine doesn't result in the failure of a video, to improve playback speed and efficiency, and to create an online

backup. The other utilized resource is the moving of popular content to a CDN or Content Delivery Network. The CDN replicates content in multiple places, not only achieving the same benefits of a mini cluster, but also increasing the odds of the viewer being closer to the database and servers storing the video, and thus decreasing load times.

Open Source Software/Hardware

YouTube utilizes the Linux operating system, as a form of open source software. Open source is code that allows anyone to view or edit the source code, and share these edits with others. YouTube uses the SuSe version of Linux using many of the tools built into Linux, and layering on top of them. This allows YouTube to utilize simple, affordable, readily available tools for their benefit.

Deployment Strategies

To ensure the site isn't down for long periods of time, or entirely, during development, and deployment of changes to the main code for the site YouTube utilizes Python because of its ability for rapid flexible development and deployment. YouTube understands how being down and unusable can not only hurt the immediate profit margin, but also their long term by losing potential customers. This is why YouTube, strives for minimal downtime and effects, and what has helped them become the dominate video distribution platform.

Security

For security YouTube uses the Hypertext Transfer Protocol Secure, which offers secure communication over a computer network, which adds the SSL/TLS protocol on top of HTTP, to help secure input and delivery of sensitive information.

Conclusion

YouTube has been able to grow as a company because of it's ability to increase the size, speed, efficiency, and effectiveness of its tech stack, relative to its growth as a company. This increase of server size and power, coupled with the design and utilization of software has helped YouTube become the largest and most used video distribution platform on the market.