

Google's Technology Stack

Web/Database Servers

Google currently has six different specialized functions that they require their servers to perform. These functions are detailed below:

- Web Servers: Responsible for executing user-submitted queries, and format the results as an HTML page.
- Data-gathering Servers: These servers are permanently dedicated to spidering the web. Spidering is what an Internet bot does to update its content index of a site.
- Index Servers: These servers contain a set of index shards, which are small partitions of a database or search engine. When queried, these servers return a list of document IDs that contain any of the keywords requested.
- Document Servers: These servers contain the documents that are searched. Each document is typically stored across dozens of servers. When queried, these servers return a summary of documents. They can also be queried to return a complete document.
- Ad Servers: These servers manage advertisements for Google's advertising services, most notably [AdWords](#) and [AdSense](#).
- Spelling Servers: These servers serve to correct any spelling errors in user queries

Software Stack

Most of the software that Google deploys on their servers is developed in-house, which allows for a high level of customization. Google has not officially commented on the matter, it is believed that C++, Java, and Python are the favored languages. The details about their more specific software stacks are detailed below:

- Operating Systems (OS):
 - Google Web Server (GWS): This is the customized version of the 64-bit Linux kernel that Google runs on their servers.
- Storage Systems:
 - Google File System (GFS): This is a file system that was developed in-house. Its purpose is to "provide efficient and reliable access to data using large clusters of commodity hardware".
 - BigTable: This is a proprietary data storage system, developed by Google, which is built on top of the Google File System. BigTable is a three-dimensional mapping system which maps row and column keys to a timestamp. This system is specifically designed to handle petabytes of data.
 - Spanner: This is the relational database management system (RDBMS) that Google uses.
 - Google F1: This is a new database system, also created in-house, which combines NoSQL and MySQL resulting in a system that scales easily, and, compared to most current RDBMS has a larger fault tolerance.
- Indexing and Search Systems:

- TeraGoogle: This is Google's very large search index.
- Caffeine (Percolator): This is Google's continuous crawler and indexer that was launched fairly recently. It is a platform proprietary to Google that was previously kept very hush-hush. It was implemented as a faster search system than what they previously had in-place.

Hardware

As of January 2012, it was estimated that Google owned around 1.8 million servers spread through at least 13 data-centers worldwide. The number of servers was projected to grow to at least 2.4 million by early 2013. Google typically uses commodity-class servers, meaning they are cheap, relatively low performance compared to most large companies servers, and only contain the necessary hardware required to perform their tasks. As of 2010, each server consisted of:

- Open-top systems with two processors, each with two cores
- Large amounts of RAM spread over 8 DIMM slots
- Two SATA hard drives connected to large power supplies
- One 12V battery for back-up power
- One 12V only power supply
- One set of PS-2 style keyboard/mouse inputs
- One PCI slot, used for video cards when servicing units
- One DB-9 serial port
- One Ethernet port
- Two USB ports

These servers run a customized "Google" version of the 64-bit Linux kernel, which only contains Ethernet and disk drivers. Ultimately, Google's goal when purchasing and designing their server setup is to reduce the cost, power consumption, and heat build-up while at the same time maximizing performance and ease of service.

Sources

- (1) <http://www.i-programmer.info/news/84-database/4286-googles-f1-scalable-alternative-to-mysql.html>
- (2) <http://michaelnielsen.org/blog/lecture-course-the-google-technology-stack/>
- (3) <http://thehyperrealists.tumblr.com/post/7533111001/what-software-technology-stack-does-google-use>
- (4) http://en.wikipedia.org/wiki/Google_platform
- (5) http://www.theregister.co.uk/2010/09/24/google_percolator/