Hardware and Software that build up the Facebook

Facebook is a mammoth using by 400 million active users who generate nearly 200 billion page views every month. Here is a peek into the technology Facebook uses to handle such enormous customer base every day.

Web front-end presentation layer is written using PHP - 3 million lines of code. HipHop compiler is used to convert the basic PHP code. A well performing Web logic execution layer is created by compiling this code again using g++. HipHop Interpreter and HipHop Virtual Machine are used to convert the code into HipHopByteCode. Static Compilation is done. Hip Hop enables Facebook API tier to serve double the traffice with 30% less CPU usage. Log Balancer is used to manage incoming web page access. Tornado non-blocking web server framework created through Python is ued to handle millions of simultaneous connections at a time.

Server: LAMP (Linux, Apache, MySqL and PHP).

Applications on Facebook: Directly rendering HTML, CSS and Java Script, iFrames, FBML for data-driven execution mark-up.

Thrift (Protocol): It is a lightweight framework used for cross language development. It is useful for automated type synchronization, binding generation and documentation. It enables transparent interaction between Java, PHP, Python and C++. Services implemented in Java use Facebook custom application server and use mostly Thrift, not Tomcat or Jetty. PHP is used because it is very simple to debug without any necessity to re-compile frequently.

Simple data storage is randomly distributed across a huge server in MySQL. It is not used as RDB.

MemCache memory caching system is responsible for the fast performance as they cache data in RAM. Hadoop's HBase is also used. Apache Zookeeper is also used to store data in hierarchical nodes. Quicker access to shared configuration services is provided through these nodes.

Page rendering is done through BigPipe. The idea behind using this is to divide huge websites into small pieces known as page lets to pipe line them easily while executing them inside web browsers and servers. Over 300 TB of data is stored in Memcached processes currently and it keeps increasing every day.

Scribe (log server): Hadoop and Hive are used for Offline processing. Online processes like logging in, clicks and feeds are passed through Scribe. They are stored in HDFS. Varnish Cache is used in proxying HTTP as it provides good efficiency. Scribe server aggregates real-time streaming log data from many different servers. It is built on top of Thrift.

In a typical file system an image file requires minimum 3 i/o to get stored. Haystack, Facebook's customized storage software developed to handle several pictures at a time requires only 1 i/o for each image file metadata. Haystack is an ad-hoc solution developed by Facebook primarily for its own usage. It brings in only low level optimizations along with append-only writes. Haystack is capable of scaling through 60,000 photographs in a second. Haystack is not open source software.

Apache HBase database management system is now used instead of Cassandra which was used for a long time. Facebook message works on its own architecture based on dynamic cluster management. Business logic is captured in small 'Cell'. New cells are added as per the demand. They can be easily upgraded too. The architecture is flexible enough to host these cells from different data centres. Metadata store failures will affect only users concerned with those particular cells and disaster management becomes quite easy. An automated system monitoring the smooth workflow escalates issues to humans, if it spots any outages it cannot repair automatically. Facebook Message search engine is developed with an inverted index. It is stored in HBase. Facebook chat, is created using Epoll server based on Erlang accessed using Thrift.

Server: Facebook owns nearly 60,000 servers. There are several data centres for the company each using their own self-designed hardware. The recent one was opened in Prineville, Oregon and advertised as Open Compute Project.

Facebook's Real Time Analytics system is controlled through Scribe. The incoming logs are stored in HDFS instead of using Puma to store them in HBase.

Technology Stack