# Facebook

Facebook is recognized as the largest online social network system in the last few years, which is come up with billions numbers of users in the last 2013. The system is recognized as distributed system in its design, infrastructure and architecture .The datacenters behind this network system are huge, robust, keeping the system scalable, reliable, secure, and let the Facebook accessible from anywhere with highly availability.

People from anywhere can use Facebook to stay connected with friends and family, they can share such contents of data and multimedia such as audios/videos and express what matters to them by comments and likes. Facebook systems at the first are responsible for processing large quantities of data, named as "Big Data", which is ranging from simple reporting and business intelligence to the huge measurements and reports executed from different perspectives [ (Asma Mohammad Salem, n.d.)], this numerous large of data located on different geographically distributed datacenters and being processed under highly equipment's servers, which they architected in high technologies to improve the whole performance of the system ,Facebook inspired by Hadoop and Hive systems supported by its integrated components which Facebook was built on top of this technologies.

Nowadays, Cloud computing is the main topic for supporting systems and realizing applications. Facebook system is as a geographically distributed system is recently being integrated with its feature and services by cloud computing solutions. Ending the system design with cloud technology solutions, this paradigm shift in technologies would server an alternative solution that could keep system in Facebook dynamically scale in the future, and maintain the rapid growth while keeping performance metrics in bounds and saving the system stability and functionality[ (Asma Mohammad Salem, n.d.)].

## Features of Facebook

Starting at early system features and services, Facebook as an example of commercial Online Social Network (OSN), and a hosted application that attract users with a set of features and attracts advertisers, who pay for the privilege of displaying ads targeted to these users. OSN interconnecting users though friendship relations, and allow for synchronous and asynchronous communications of user generated content ,such as text ,multimedia from audio, video and such third party OSN application updates over a social graph ,these services attract users and will be the main reason for huge traffic that flow through the system parts. Facebook features are being provided by starting with a registration phase, requiring a user name and password, this registration in done once and a login should be done after registration to start using the website, starting with inviting  based on 2 building a friendship to start sharing your status, media and news with them, most of website are:-

**Wall:** it is the original profile space for a user where contents posted there, including photos and videos, and files, user can attach any content on his or her wall and being visible to anyone, by choosing the space of visibility on the wall user can limit visibility to the wall contents, which were in early versions of Facebook as text only.

**News Feed:** it is a home page in which users can see a continue updates list of their friend's activities. They can explore information that includes profile changes, updates and coming events, users can explore the conversations that taking place between the walls of a user's friends.

**Timeline:** a space in which all photos, videos, posts, and contents are categorized according to instant of time in which they were uploaded or created.

**Friendship:** this feature is what Facebook is based on, "Friending" someone is the act of sending another user a friend request on Facebook or accepting friendship request. A user has full control to manage his or her friend list.

**Likes and tags:** It is positive feedback; users can apply likes on updates, comments, photos, status and links posted by their friends, these likes make the content appear in their friend's pages notifications and updates.

**Notifications:** keeping track of all the most recent actions or updates. It is an indicator to inform the user that an action has been added to profile page, his or her wall or time line, any Comment or like, shared media that being tagged in.

**Networks, groups, and pages:** Facebook allow users to build their networks, groups and creating pages which combine them around an idea or specific community. They can used for posting items or issuing messages for a group of users who join these communities.

**Messaging and outbox:** a service allows users to send messages to each other. Users can send a message to any number of friends at a time. Managing messages also provided .By the year of 2010, Facebook announced a new Facebook Messages service which give a user an account under the Facebook.com ,This system is available to all of the users, providing text messages, instant messaging, emails and regular messages

All these features and more are being served on Facebook, adding the different applications that are: events, marketplaces, notes, places, questions, photos, videos, and Facebook pages.

# Design

In just few years Facebook distributed system has a traditional design, in which Hadoop and Hive were working together to perform tasks for storage and analysis of large data sets .these analysis are classified in to two categories, most of them are offline batch jobs to maximize the throughput and efficiency and the others are online jobs. These workloads are read and write large amount of data form disks sequentially.

- **Memcahed servers:** Recent design of Facebook, let Hadoop performing a random access workloads that provides low latency access to HDFS, by using a combination of large clusters of MySQL databases and caching tiers built using memcached ,that will be support a better in performance while all results from Hadoop are directed to MySQL or memcached for consumption by the web tier side[ (Asma Mohammad Salem, n.d.)].

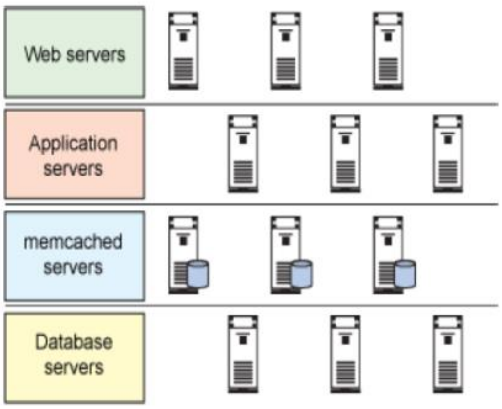



**Figure:** *memcached servers*

Recently, a new generation of applications has been applied at Facebook in which requires very high write throughput and cheap and elastic storage, while keeping low latency and disk efficient sequential and random read performance. We sum-up the whole system components in the figure below.

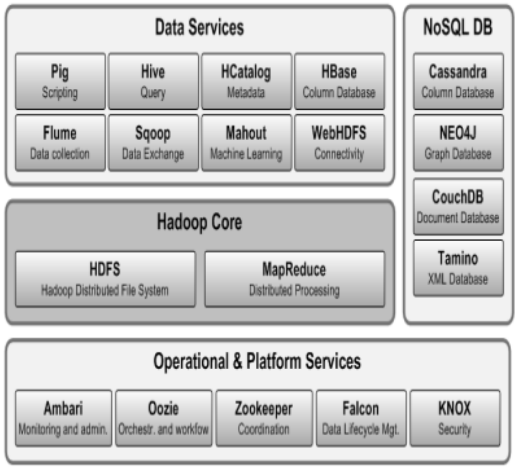



**Figure:** *whole recently system infrastructure*

- **Cloud computing support:-** Redundant cluster servers are used to hold the whole system in Facebook, now the system in consist of physical server that needed to be extended day by day, this scheme of datacenters hosted Facebook servers is subjected to be at risk one day and subjective to many problems being as limitations for growth. Nowadays cloud computing is offering a powerful environment to scale web applications without difficulty. Using such schemes of resources on demand for many scaling points as web applications, storages, and servers Cloud computing aims to deliver services over the network it provides ability to add capacity as needed. While scalability is a measure of ability of an application to expand to meet enterprise business needs ,resources under demand are anything could be required or shared by the system users ,it is ranging from processor, storage space and network bandwidth ,these resources will affect primary the system performance and degrade the application behavior when they are running on shortage of them ,when application is not scale well it is encounter the performance and service availability as demand increase[ (Asma Mohammad Salem, n.d.)].

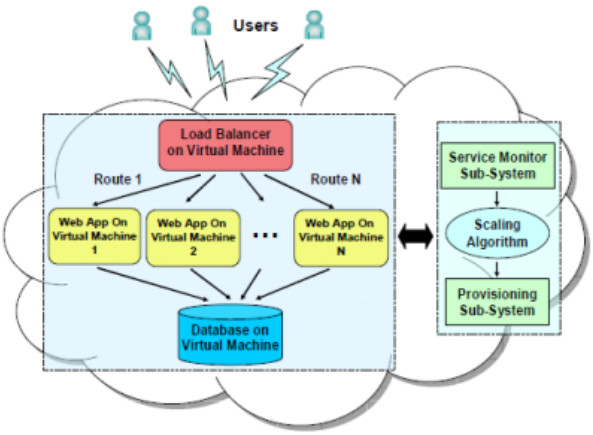



**Figure:** *Infrastructure to scale web applications in a Cloud*

# Architecture

Facebook, the online social network (OSN) system is relying on globally distributed datacenters which are highly dependent on centralized U.S data centers, in which scalability, availability, openness, reliability and security are the major System requirements. When founded in 2004 it was such a dream to be the largest OSN by the year of 2013 putting the system on the surface of risk unless it well designed and protected against failure and attacks. The architecture of the system, the scheme here is 3 tier architecture or more (4 tier), in which the data flow originated form clients requests that are served by the following steps (Asma Mohammad Salem, n.d.):

- Initially by dedicated web servers, these web servers are highly connected in high available scheme to handle billions of requests and aggregate the logs coming from different webservers.
- Then they are redirected in uncompressed format to the Scribe–Hadoop Clusters they are dedicated for logs aggregations , the latter is then communicate the Hive –Hadoop servers cluster, these servers are divided in two categories ,the Production and the Adhoc ,they are clusters of servers that are balancing according to the priority of jobs, for example the Production servers are dedicated to the jobs that being strict in delivery deadlines time constrains, while the adhoc cluster is serving the low priority batch jobs as well as any adhoc analysis that the user want to do on historical data sets.
- Federated MySql is the data base engine which hold the data bases holding up the whole system. These tier parts are described in the figure.

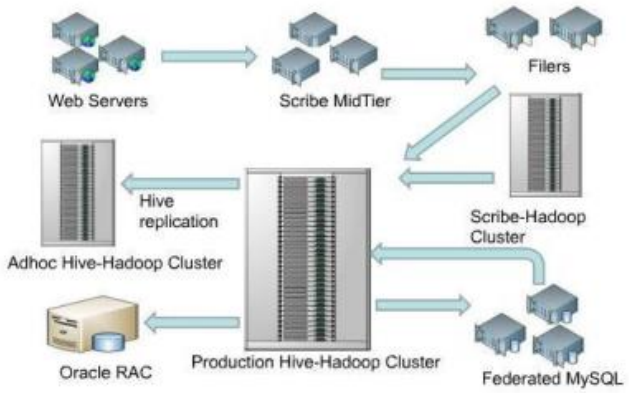



**Figure:** *Facebook system architecture*

# Layout

Ten years ago today, Mark Zuckerberg gave birth to The Facebook and launched an online social revolution in his Harvard dorm room. We all know what happened next—thanks in no small part to an eight-time Academy Award-nominated film. With such humble beginnings, no one could have predicted the wild success (and cries of outrage) the site's many iterations would bring over the years.

The original website is (vaguely) recognizable as the service we use today, but it is far clunkier-looking. There's no gloss, and there are strange aesthetic choices that belie a young rogue creator. The signature blue color scheme has endured; the binary code and random, creepy man's face at the top are long-gone. In terms of functionality, it's hard to compare a single-school online directory to the communication behemoth Facebook has become.

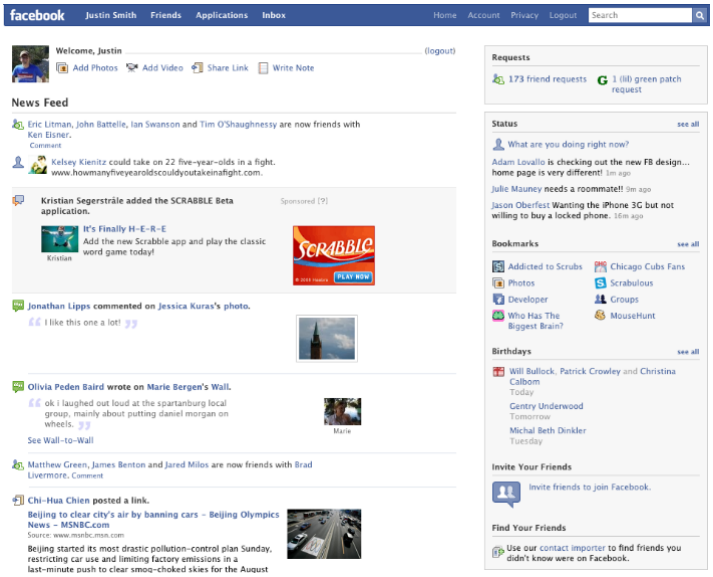
The homepage makes it clear that this is Zuckerberg's project: "A Mark Zuckerberg production" tags the bottom of each page. At first, Facebook was exclusive to Harvard, which was noted prominently on the site.

In 2005, a profile on The Facebook had many of the details you might find on today's Facebook profile, but still had a very MySpace vibe[ (Kate Knibbs, Mar 2, 2020)].


In 2006, Facebook opens its doors to everybody. News Feed also makes its debut, spurring the first of what will soon be seemingly endless debates on privacy settings[ (Kate Knibbs, Mar 2, 2020)].

One year later, Facebook got a more adult makeover, which included the addition of the Mini-Feed, or a timeline of a friend's Facebook activity directly on their profile. Several weeks after the Mini-Feed went live, Facebook moved beyond schools and opened to anyone with an email address[ (Chloe Albanesius, February 4, 2014)].


Summer 2008 was the dawn of "new Facebook," which included a cleaner menu bar up top, as well as tabs that separated your activity and other profile data.

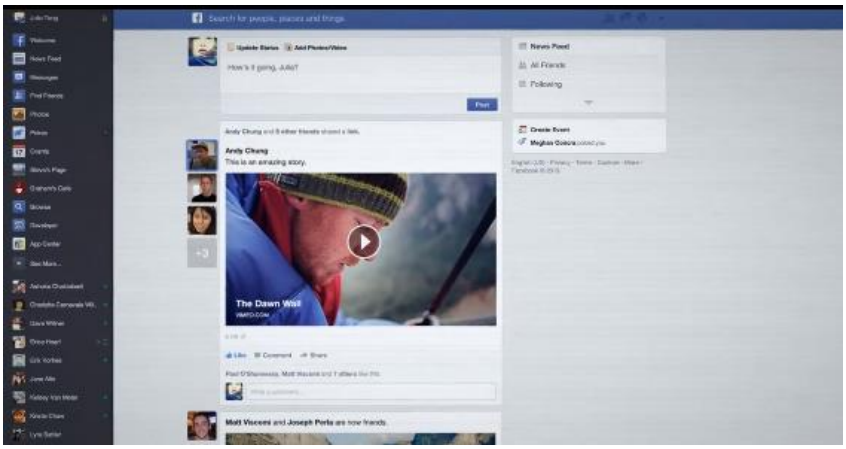


By March 2009, Facebook also rolled out a new homepage. It added the Publisher tool from user profiles to the main homepage to let you post links, photos, videos, or write a note without an extra click.


In 2010, it was time for another makeover. Probably the most interesting change was the array of photos that appeared near the top of the page (Asma Mohammad Salem, n.d.).

In 2011, Facebook introduces its biggest changes yet with Timeline and the front page News Tickers. The idea of the profile as a sort of digital scrapbook was easier to explain to investors than the Wall, and Timeline emphasized curating a collection of images and information about your life, while older versions of the profile were more about getting acknowledged by your friends. Zuckerberg kicked off the 2011 f8 developer conference with an ambitious new Facebook interface that "tells the story of your life": the Timeline. It featured a large cover photo and a Pinterest-like assortment of posts that scrolled back through your Facebook history[ (Chloe Albanesius, February 4, 2014)].


In 2013, Facebook gave Timeline and the News Feed a major revamp in 2013, with much larger pictures and an aesthetic influenced by Instagram[ (Kate Knibbs, Mar 2, 2020)].



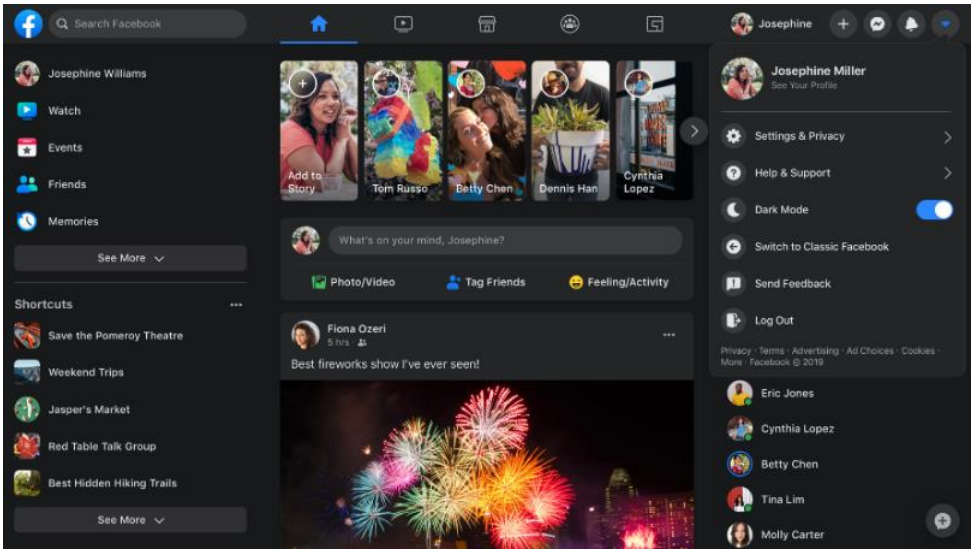
In 2020, the new interface uses a responsive design that displays content based on screen width and other parameters. If the width of the browser window is sufficient, Facebook displays a traditional but heavily modified three column design.

The design features rounded corners and is more colorful than the current design of desktop Facebook. Text appears larger and there is more grayspace; this may lead to extended scrolling sessions as less content is displayed at a time on the visible part of the screen.

Zooming in or out, or changing the size of the browser window may help display more content on the screen at the same time.

The icon bar at the top provides quick access to various sections on the site including the homepage, videos, marketplace, groups, and gaming[ (Brinkmann, March 23, 2020)].

In 2020, the long-awaited Facebook Dark Mode is available at one click. The classic light version will change into a snap into dark mode, which will be a good option when watching videos or Facebook Live.



# Working Procedure

Facebook uses a variety of services, tools, and programming languages to make up its core infrastructure. At the front end, their servers run a LAMP (Linux, Apache, MySQL, and PHP) stack with Memcache[ (Steve Campbell, Feb 27, 2010)].

- **Front end**

  - **Linux & Apache:** This part is pretty self-explanatory. Linux is a Unix-like computer operating system kernel. It's open source, very customizable, and good for security. Facebook runs the Linux operating system on Apache HTTP Servers. Apache is also free and is the most popular open source web server in use.

  - **PHP:** Facebook uses PHP because it is a good web programming language with extensive support and an active developer community and it is good for rapid iteration. PHP is a dynamically typed/interpreted scripting language. The facebook sdk (system development kit) for php is a library with powerful features that enable php developers to easily integrate facebook login and make requests to the graph API. It also plays well with the facebook sdk for javascript to give the front-end user the best possible user experience. But it doesn't end there, the facebook sdk for php makes it easy to upload photos and videos and send batch requests to the graph API among other things.


  - **Pipelining**: Bigpipe is a dynamic web page system developed by Facebook. The general idea is to perform pipelining of sections through the implementation of various stages within web browsers and servers. Browser sends an http request to web server. Web server parses the request, pulls data from storage tier then formulates an html document

and sends it to the client in an http response. Http response is transferred over the internet to browser. Browser parses the response from web server, constructs a tree representation of the html document, and downloads css and javascript resources referenced by the document. After downloading javascript resources, browser parses and executes them.

- **HIP-HOP:** It is a PHP compiler which is developed by Facebook. The processing time for PHP language is slow Created to minimize server resources.t converts PHP scripts into optimized C++ code.

- **Back-End**

Facebook's backend services are written in a variety of different programming languages including C++, Java, Python, and Erlang. Their philosophy for the creation of services is as follows:

1. Create a service if needed

2. Create a framework/toolset for easier creation of services

3. Use the right programming language for the task

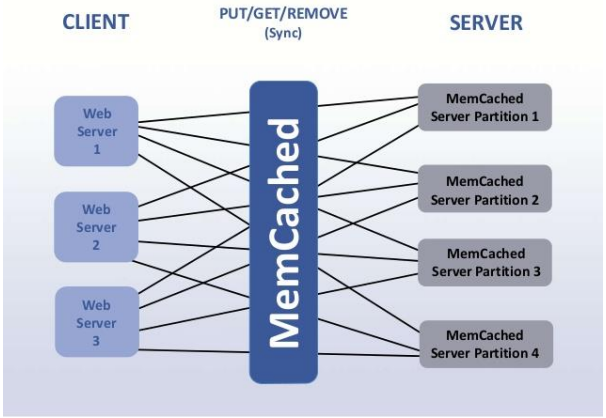
- **MySQL:** For the database, Facebook utilizes MySQL because of its speed and reliability. MySQL is used primarily as a key-value store as data is randomly distributed amongst a large set of logical instances. These logical instances are spread out across physical nodes and load balancing is done at the physical node level. As far as customizations are concerned, Facebook has developed a custom partitioning scheme in which a global ID is assigned to all data. They also have a custom archiving scheme that is based on how frequent and recent data is on a per-user basis. Most data is distributed randomly.

- **Scribe (log server):** Scribe is a server for aggregating log data streamed in real-time from many other servers. It is a scalable framework useful for logging a wide array of data. It is built on top of Thrift.

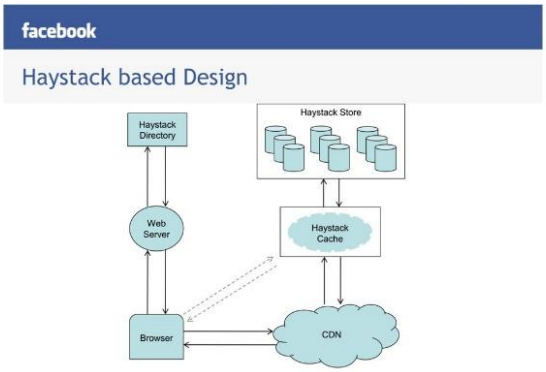
- **Thrift (protocol):** Thrift is a lightweight remote procedure call framework for scalable cross-language services development. Thrift supports C++, PHP, Python, Perl, Java, Ruby, Erlang, and others. It's quick, saves development time, and provides a division of labor of work on high-performance servers and applications.

- **Cassandra:** Cassandra is a database management system designed to handle large amounts of data spread out across many servers. It powers Facebook's Inbox Search feature and provides a structured key-value store with eventual consistency.


- **Memcache:** Memcache is a memory caching system that is used to speed up dynamic database-driven websites (like Facebook) by caching data and objects in RAM to reduce reading time. Memcache is Facebook's primary form of caching and helps alleviate the database load. Having a caching system allows Facebook to be as fast as it is at recalling your data. If it doesn't have to go to the database it will just fetch your data from the cache based on your user ID.



- **Haystack:** Haystack is an object store that is designed for sharing photos on Facebook where data is written once, read often, never modified, and rarely deleted and replaced. It contains efficient storage of billions of photos. It is highly scalable and uses extensive caching in its main memory. The new photo infrastructure merges the photo serving tier and storage tier into one physical tier. It implements a HTTP based photo server which stores photos in a generic object store called Haystack. The main requirement for the new tier was to eliminate any unnecessary metadata overhead for photo read operations, so that each read I/O operation was only reading actual photo data (instead of file system metadata). Haystack can be broken down into these functional layers: HTTP server, Photo Store, Haystack Object Store, File system and Storage[ (Daniel Barrigas, Hugo Barrigas, Melyssa Barata, Fedro Furtado, Jorge Bernardino, May 2014)].

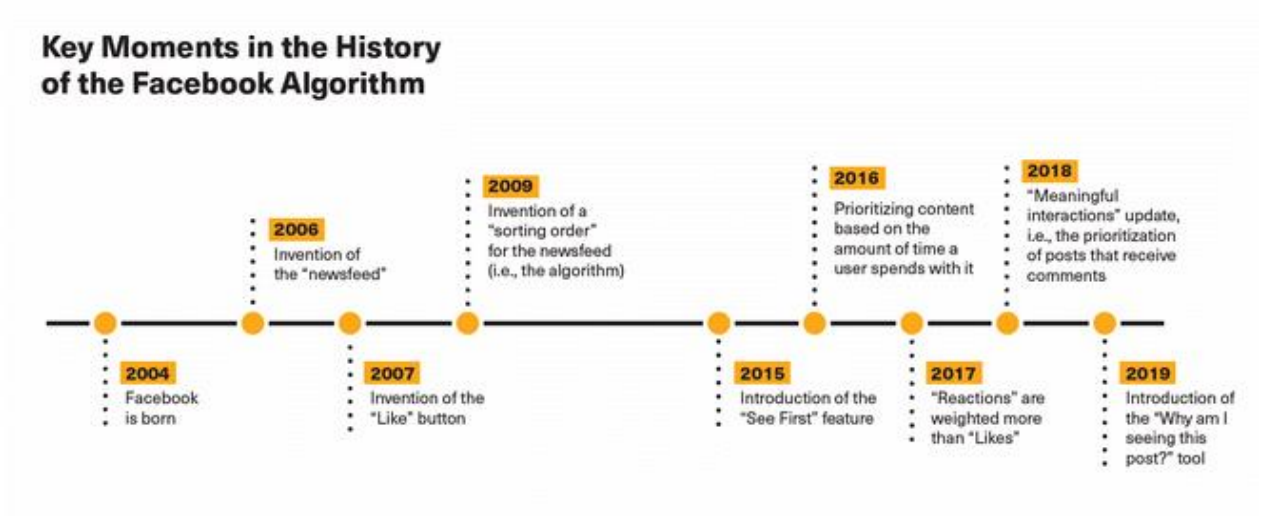



- **HADOOP & HIVE:** Hadoop is ideal for scalable systems with an enormous need to store and process large amounts of information. It scales horizontally. Instead of replacing existing storage systems, Hadoop complements them. This process allows existing systems to serve data in real time or provide transactional interactive business intelligence.

# Algorithm

It is believed that Facebook's new algorithm is based on the Vickrey-Clarke-Groves algorithm, which “operates as a closed auction.”

The Facebook algorithm, the set of calculations Facebook uses to decide what content you see, has a lot of sway and influence. Long gone are the days of a chronological feed, today you get what you’re given (Boyd, Jan 2, 2019).




The Facebook algorithm ranks all available posts that can display on a user’s News Feed based on how likely that user will have a positive reaction[ (Swan, Apr 29, 2020)].

Facebook now ranks and prioritizes content posted from friends over publishers, with a focus on what the algorithm determines as “meaningful interactions.”

Here are the four factors that determine if a story is relevant for a user’s News Feed:

**1. Inventory:** Inventory represents the stock of all content that can display to a user on Facebook’s News Feed. This includes everything posted from friends and publishers (Swan, Apr 29, 2020).

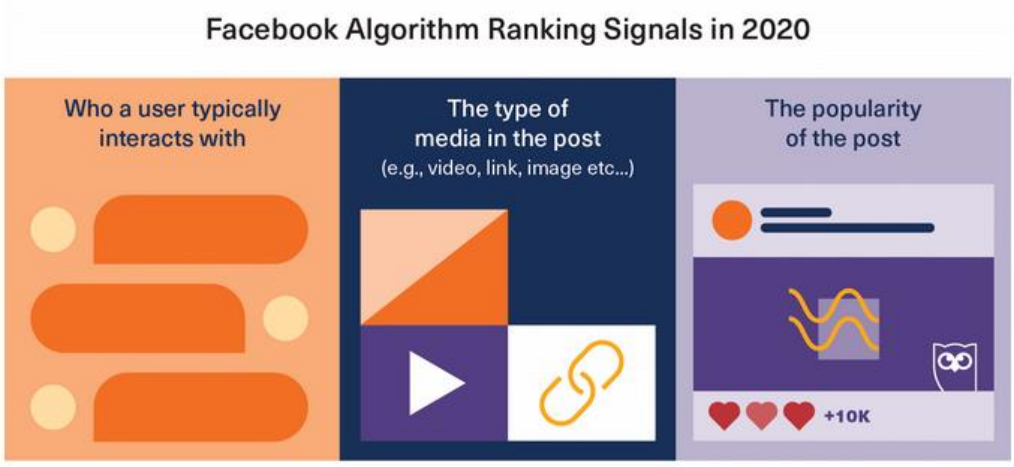
**2. Signals:** Signals represent the information that Facebook can gather about a piece of content. Signals are the single factor that you have control over. You want your content to signal to Facebook that it’s meaningful and relevant to your target audience (Swan, Apr 29, 2020).

These are what Facebook uses to help choose which content goes out. It includes a list of criteria. We’ve listed all the ones we know here, with the bolded ones being more heavily-weighted by Facebook.

As you can see Facebook is all about proper engagement from people, such as comments and shares on Messenger (Boyd, Jan 2, 2019).

- Comments and likes on a person's status of photo
- Engagement with publisher content posted by friends
- Shares on Messenger
- Replies to comments on a video
- Who posted the content
- When was it posted
- What time is it now
- Technology (what type of phone, how strong the internet connection is)
- Content type
- Average time spent on content
- How informative the post is
- Completeness of a profile

It's also worth noting that Facebook will give more weight to conversations between people than they will to those between a page and a person.




**3. Predictions:** Predictions represent the behavior of a user and how likely they are predicted to have a positive interaction with a content piece.This is where Facebook uses your profile and previous behaviors to decide what to show you. They attempt to work out how likely you are into like or interact with content, keeping stuff they think you won't engage with out of your timeline. Although it's worth noting a Pew Research Center study found 27% of people think the algorithm gets them wrong (Boyd, Jan 2, 2019).

**4. Score:** Score is the final number assigned to a piece of content based on the likelihood the user will respond positively to it. This refers to a value assigned to a piece of content that will determine its 'relevancy' to the user. The higher the score, the more likely it will appear in the feed. Obviously this means content will get different scores for individual people (Boyd, Jan 2, 2019).

# Challenges and how to overcome them

On February 4, 2004, Facebook was set up by Mark Zuckerberg and his fellow Harvard classmates. According to Zuckerberg, Facebook took its name from the school's directory, which listed facial portraits of students (Wenyan, 04-Feb-2019 ).

Facebook has become the world's largest social network, boasting billions of active users across the globe. Facebook system was founded with a mission to give people the power to share and make the world more open and still connecting them with friendship relationship.

Facebook's track record is a bit mixed and 2018 was a particularly challenging year for the company. These challenges included the collection of personally identifiable information of "up to 87 million people" by the political consulting and strategic communication firm Cambridge Analytica, which may have tried to exert influence on American voters using such information.
Facebook's struggles with user data underscore the common challenges faced by the Internet giants, which also includes Facebook's peers and competitors such as Google. Similarly, because China has the world's largest mobile Internet users with a population of 800 million (more than the entire population of Europe), the leading technology companies in China are scrambling to search for similar solutions, too.

What's alarming is that the individual users – and their personal data – are increasingly collected, amassed and grouped by a few “platform companies” like Facebook. These companies provide search, e-commerce, and social network all in one place, with an expansive vision for an immersive world, both online and offline (Wenyan, 04-Feb-2019 ).

How Facebook is addressing our most pressing challenges (Anon., February 4, 2019):

# Findings

The high growth rate of web based systems having hundreds of thousands of users accessing it continuously led to response time problems for users who are trying to receive information at the same time.

- The current architecture of Facebook is very large and consists of consists of many technologies and thousands of servers. The architecture of the system, the scheme here is 3 tier architecture or more (4 tier), in which the data flow originated form clients requests.

- Facebook needs to handle large amounts of information every second. Facebook continues to be a LAMP (Linux-Apache-Memcached-PHP) website, but had to change and expand its operation to incorporate other elements and services. Personalized and implemented systems also exist inside Facebook, such as Haystack.

- Facebook algorithm in a nutshell:
  Facebook's algorithm uses four main factors or signals:
  - A listing ("inventory") of the posts available to show
  - Signals from each post that tell Facebook what it is
  - Prediction on how users will react each post in the inventory
  - A final based on all of the ranking signals

  Facebook looks at:
    - Who users typically interact with
    - The type of content: videos, images, links etc.
    - The post's popularity
- The succession of Facebook is due to asynchronous (email, messages) and synchronous (face time, voice call etc.) communication and sharing of content such as text, audio, video, 3rd party apps and constant updates on wall, news feed and timeline.

# Summary

We have explored Facebook as a case study for distributed system discussed the system features and providing a detailed system design architecture, layout, working procedure, algorithm and challenges. The system is built on top of highly equipped data centers that are provide the system the availability and reliability, the Hadoop project is an example of this system that Facebook in built on top of its technology . Using the clusters for the data base systems, load balancing webservers and application servers that are responsible for replying on user's requests, the ability to compress the traffic between servers to save the bandwidth and the isolation between jobs that are derived from user's requests. Hadoop projects and whole components are example of success story that provide Facebook system with its requirements to be the most popular social network

by the year of 2013. Memcahed severs are also another example of these enhancements to decrease the load of accessing the data base in each case that require access to the data base. Cloud computing is model example that Facebook used to integrate with its features and services .this integration is done without any infrastructure modifications or any architectural changes, this is because cloud computing is offering an acceptable solution for integrating Facebook with such examples of cloud applications .the most interesting examples of these solutions the social cloud being built by the virtualization organizations that provided ,these are being scaled dynamically and on demand .

## Remarks

Facebook has gained a tremendous amount of popularity in the last years and has become the most successful social networking website ever created. If we look past all of the features and innovations, the main idea behind Facebook is really very basic “keeping people connected”.
Facebook realizes the power of social networking and is constantly innovating to keep their service the best in the business.

# References

Anon., February 4, 2019. *What Is Facebook Doing to Address the Challenges It Faces?.* [Online] Available at: https://about.fb.com/news/2019/02/addressing-challenges/

Asma Mohammad Salem, n.d. Facebook Distributed System Case Study For Distributed System Inside Facebook Datacenters. *International Journal Of Technology Enhancements And Emerging Engineering Research, Vol 2, Issue 7 152.*

Boyd, J., Jan 2, 2019. *The Facebook Algorithm Explained.*
Available at: https://www.brandwatch.com/blog/the-facebook-algorithm-explained/

Brinkmann, M., March 23, 2020. *This is how Facebook's new desktop design looks (and how you can restore the old Facebook)).*
Available at: https://www.ghacks.net/2020/02/11/this-is-how-facebooks-new-desktop-design-looks-and-how-you-can-restore-the-old-facebook/

Chloe Albanesius, February 4, 2014. *10 Years Later: Facebook's Design Evolution.*
Available at: https://www.pcmag.com/news/10-years-later-facebooks-design-evolution

Daniel Barrigas, Hugo Barrigas, Melyssa Barata, Fedro Furtado, Jorge Bernardino, May 2014. Overview of Facebook scalable architecture. *Conference: ACM International Conference on Information Systems and Design of Communication ISDOC'14.*

Kate Knibbs, Mar 2, 2020. *daily dot.*
Available at: https://www.dailydot.com/debug/old-facebook-profiles-news-feeds/

Steve Campbell, Feb 27, 2010. *How Does Facebook Work? The Nuts and Bolts [Technology Explained]*.
Available at: https://www.makeuseof.com/tag/facebook-work-nuts-bolts-technology-explained/

Swan, G., Apr 29, 2020. *How The Facebook Algorithm Works + 5 Best Practices.*
Available at: https://tinuiti.com/blog/paid-social/facebook-algorithm/

Wenyan, M., 04-Feb-2019 . *15 years on, what challenges does Facebook face today?.*
Available at: https://news.cgtn.com/news/3d3d514f3163444e32457a6333566d54/index.html