Case Study: Facebook

Submitted To :

Meskat Jahan

Lecturer

Department of CSE

Comilla University

Submitted By:

Farah Nawar

Id : 11608001

Student ,

Department of CSE

Comilla University

Date of Submission : 19 January,2021

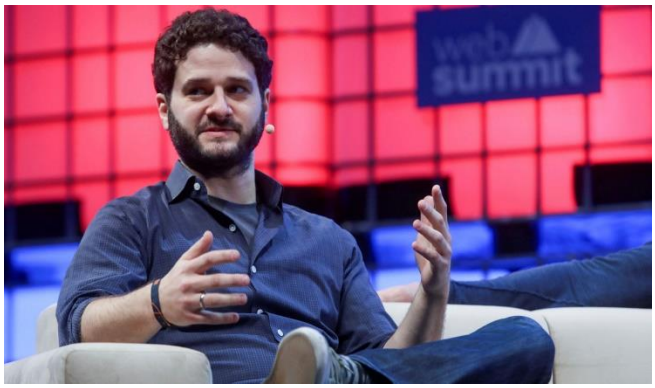
1. Introduction

The rise of the social network site (SNS) Facebook has been one of the most important social trends of the past decade. Although it only opened to the public in 2006, Facebook reports already serving one billion monthly active users at the end of 2012. (1)Moreover, 80% of these users reside outside the US and services are available in 70 languages, making Facebook a worldwide platform. While there are concerns about the accuracy and trustworthiness of these numbers (the number of accounts that are frequently used may differ from the real number of individuals using the platform) and neutral information is not available, one should agree the size of this SNS is at least substantial and the growth rate impressive. It is this growth rate that has attracted the attention of scientists from widely different fields of inquiry. In February 2013, using Facebook as a keyword on the ISI Web of Knowledge provided no less than 3068 hits.

2. Background of the Organization

Facebook was launched as TheFacebook on February 4; 2004.It was founded by Mark Zuckerberg with his college roommates and fellow Harvard University students Eduardo Saverin, Andrew McCollum, Dustin Moskovitz and Chris Hughes.

(a)



(b)

(c)

(d)

(e)

Fig.1: *(a) Mark Zuckerberg, founder and CEO of Facebook (b) Dustin Moskovitz , Co-founder of Facebook (c) Chris Hughes, Co-founder of Facebook (d) Andrew McCollum, Co-founder of Facebook (e) Eduardo Saverin,Co-founder of Facebook*

Fig.2. Facebook's co-founders (from left): Dustin Moskovitz, Chris Hughes and Mark Zuckerberg

In February 2004, Mark Zuckerberg launched "The facebook", as it was originally known; the name taken from the sheets of paper distributed to freshmen, profiling students and staff. Within 24 hours, 1,200 Harvard students had signed up, and after one month, over half of the undergraduate population had a profile. The network was promptly extended to other Boston universities, the Ivy League and eventually all US universities. (2)

Former PayPal CEO Peter Thiel gets a lot of credit for being the first investor in Facebook, because he led the first formal Facebook round in September of 2004 with a $500,000 investment at a $5 million valuation. But the real "first investor" claim to fame should actually belong to a Harvard classmate of Mark Zuckerberg's named Eduardo Saverin. (3)

It became Facebook.com in August 2005 after the address was purchased for $200,000. US high schools could sign up from September

2005, and then it began to spread worldwide, reaching UK universities the following month.

Fig.3 Zuckerberg, left, cofounded, Facebook with his Harvard roommate, Dustin Moskovitz, center. Sean Parker, right, joined the company as president in 2004. The trio was photographed in the company's Palo Alto office in May 2005.

As of September 2006, the network was extended beyond educational institutions to anyone with a registered email address. The site remains free to join, and makes a profit through advertising revenue. Yahoo and Google are among companies which have expressed interest in a buy-out, with rumored figures of around $2bn (£975m) being discussed. Mark Zuckerberg has so far refused to sell.

Fig.4 Zuckerberg, photographed in March 2006 at the headquarters of Facebook in Palo Alto.

The site's features had continued to develop during 2007. Users had the facilities to give gifts to friends, post free classified advertisements and even develop their own applications - graffiti and Scrabble are particularly popular.

In 2008, Octobor, Facebook sets up its international headquarters in Dublin, Ireland.

Fig.5.Facebook headquarter in Dublin, Ireland

In 2009, Facebook activates the Facebook like button and announces a feature whereby people can @-tag friends in their status updates and comments.

In 2010, Facebook introduces the option to like individual comments. Facebook Messenger is launched for Android and IOS. October 19, 2011.

In February 2011, Facebook announced plans to move its headquarters to the former Sun Microsystems campus in Menlo Park, California. In March 2011, it was reported that Facebook was removing about 20,000 profiles daily for violations such as spam, graphic content and underage use, as part of its efforts to boost cyber security. Statistics showed that Facebook reached one trillion page views in the month of June 2011, making it the most visited website tracked by DoubleClick. According to a Nielsen study, Facebook had in 2011 become the second-most accessed website in the U.S. behind Google.

Fig.6. Facebook headquarter in Menlo Park, California

In 2012, April Facebook acquires Instagram for $1 billion.

In 2015, August 27, Facebook announces that it has hit the milestone of 1 billion users accessing it on a single day.

On April 12, 2016, Zuckerberg outlined his 10-year vision, which rested on three main pillars: artificial intelligence, increased global connectivity, and virtual and augmented reality

On August 14, 2020, Facebook started integrating the direct messaging service of Instagram with its own Messenger for

both iOS and Android devices. After the update, an update screen is said to pop up on Instagram's mobile app with the following message, “There’s a New Way to Message on Instagram” with a list of additional features. As part of the update, the regular DM icon on the top right corner of Instagram will be replaced by the Facebook Messenger logo.

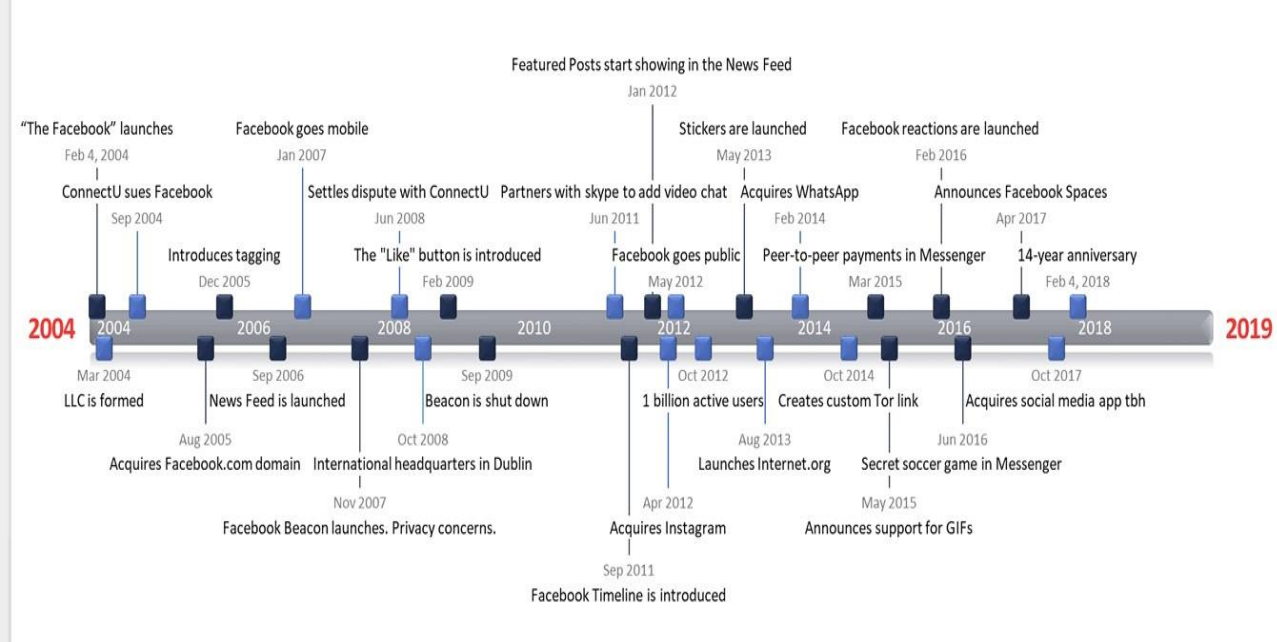


Fig7. Facebook History timeline by 2019

2.1. Controversies

During its eight years in business, Facebook has suffered numerous privacy controversies, partly as a result of how the service works: users of Facebook create online profiles, which contain a great deal of personal and sensitive information including their name, their interests, the names of their friends, photos and videos they upload, and content they add to their friends’ profiles by sending comments and sharing photos. Users may also “tag” their friends’ images (i.e., identify them by

name) without prior consent from those friends and install games and other applications developed by third parties that obtain access to the profile information of both the users and their friends. In short, Facebook, by its very nature, raises fundamental privacy challenges because it enables users to disclose unprecedented volumes of highly personal information, not only to friends and friends of friends, but, depending on one's privacy settings, to very large and unfamiliar audiences as well.

3. Design

The Facebook system can be divided into five components:

i. Data source;
ii. Data Collection, Processing and Loading (CPL)
iii. Data Analysis and Aggregation
iv. Interface and Visualization
v. Job and Model Specification

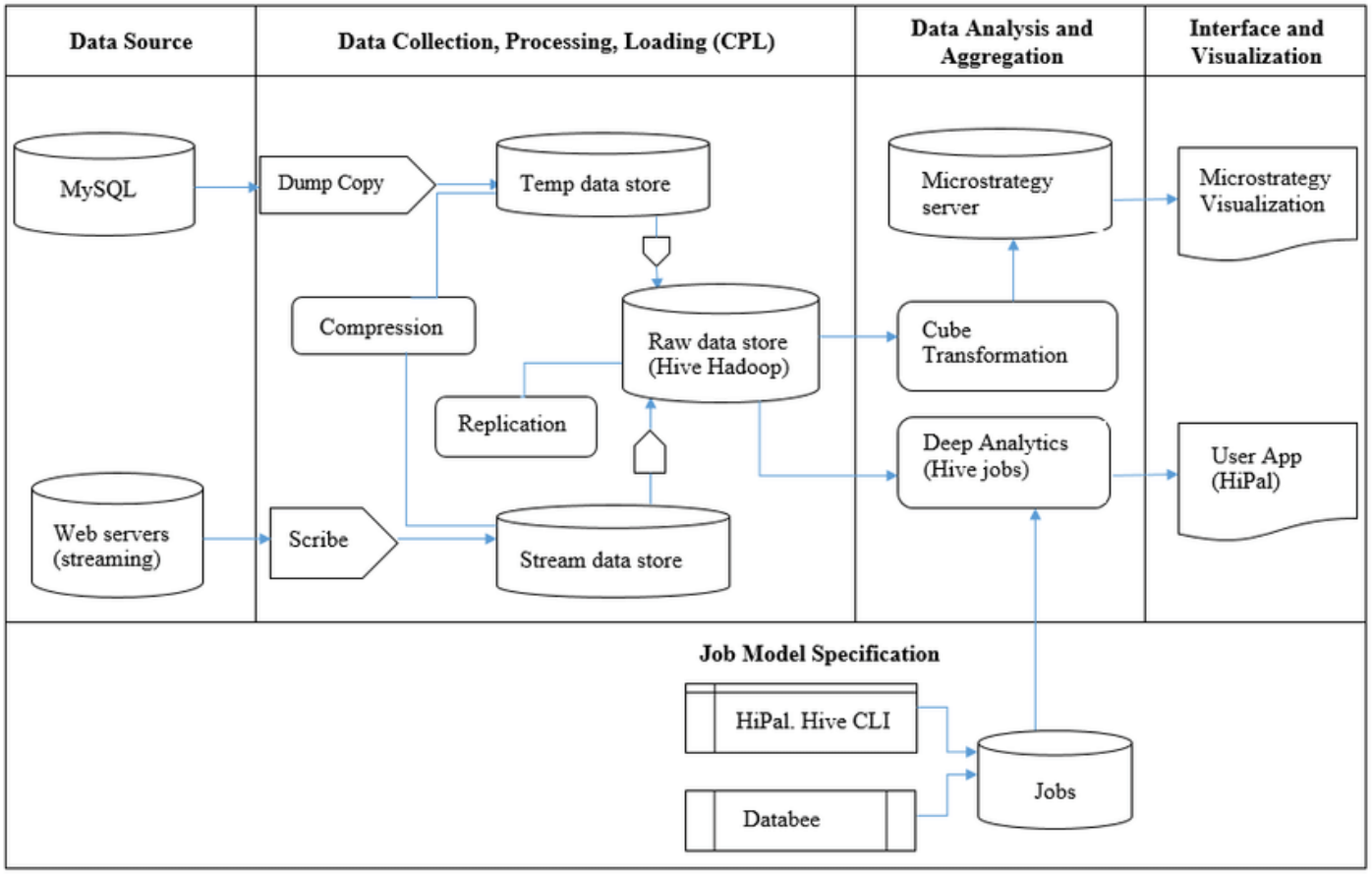


Fig 7. System design of Facebook (4)

i. **Data Source:** It refers to the original source of data to be collected. This can be traditional data such as relational data or streaming data. Data can be structured, unstructured, semi-structured or streaming. Facebook collects data from two sources:
   a. A federated MySQL tier containing user data
   b. Web servers for event log data

ii. **Data Collection, Processing and Loading:**

   a. *Data Collection* represents getting data from the (normally multiple) sources for storage or analysis. There are several techniques for this such as Snapshot.

In Facebook, The structured data from the Federated MySQL is copied, compressed and stored in the Production Hive-Hadoop cluster.

b. Data Processing covers the executions required for processing the source data before moving into the Loading stage.Data Processing represents functionalities such as data cleaning, replication, filtering, algorithm service, etc.

c. *Data Loading* can then be executed, meaning the collected and processed data are now loaded into a data storage such as Hadoop HDFS.

These three items (Data Collection, Processing and Loading) are grouped in in one component because each item closely links to each other. For example, data collection requires data processing for cleaning or formatting as well as loading the data into storage. At high level architecture, this improves logical linkage and flow between items closely related, and hence provides clearer and effective overall architecture.

iii. **Data Analysis and Aggregation:** Data analysis refers to the related data analysis tasks and processes whereas the Aggregation refers to the data storage (including multidimensional) which stores the results of the analysis.
In Facebook, The Scribe servers aggregate event logs and process the data in Hadoop Distributed File System (HDFS). HDFS data is compressed periodically, and moved to Production HiveHadoop clusters for further processing. Facebook differentiates low or high priority clusters for data analysis. High priority jobs are executed in the Production Hive-Hadoop cluster whereas lower priority jobs and ad hoc analysis jobs are executed in the Ad hoc Hive-Hadoop

cluster. In addition, data is replicated from the Production cluster to the Ad hoc cluster. . The results of data analysis are stored in the Hive-Hadoop cluster or MySQL tier for Facebook users.

iv. **Interface and Visualization:** they represent the end users as well as applications such as dashboards. HiPal, a graphical user interface with Hive CLI (a Hive command-line interface) is used to accommodate ad hoc analysis. Databee (a python framework) is used for execution and scheduling of periodic batch jobs in the production cluster. Microstrategy Business Intelligence (BI) tools are used for cube dimensional analysis.

v. **Job and Model Specification:** This covers models trained, specifications and scheduling of jobs with their storages

Databee (a python framework) is used for execution and scheduling of periodic batch jobs in the production cluster.

Microstrategy Business Intelligence (BI) tools are used for cube dimensional analysis.

3.1. Facebook Social Graph

Facebook has more than 1 billion active users who record their relationships, share their interests, upload and comment on text, images and videos. Facebook stores majority, if not all, of users' data, such as profiles, friends, posts and comments, inside a single giant *social graph*.

There are two elements inside a social graph:

i. **Nodes:** A node represents an entity, such a user, a post, a comment and a location.

ii. **Edges:** An edge represents the relationships between the nodes. For example, one edge could mean that a particular user created a particular post.

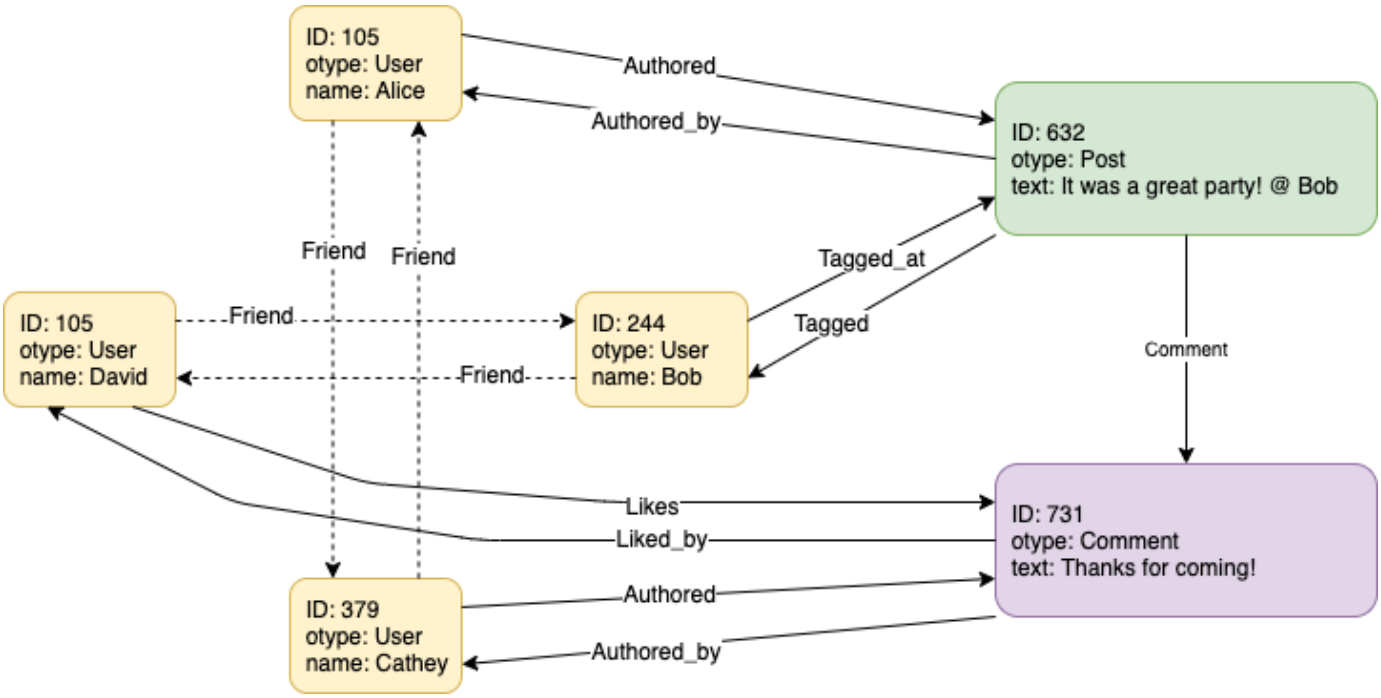


Fig.8.Facebook data model sample for social graph

Facebook uses only two database tables to represent the social graph that captures the activities of its one billion users. They are:

i. Object table
ii. Association table.

i. Object table

Object table has a very simple schema. It has 3 columns.

- **The id column**: stores the unique id of the object.
- **Otype:** stores the object type.

Each object/node can have a list of key-value pairs. otype specifies the possible keys and value type. For instance, otype of User means there could be a key name with value type string. The list of key-value pairs are serialized and stored in the data column.

ii. Association table

The schema of the association table is similar. It has 4 columns

a) Id1 and id2 represents the source and destination of the edge
b) atype is the edge type
c) Data, stores the optional list of key-value pairs associated with the edge

3.2. Facebook Backend system

There are two major components in Facebook's backend system.

i. TAO
ii. The database.

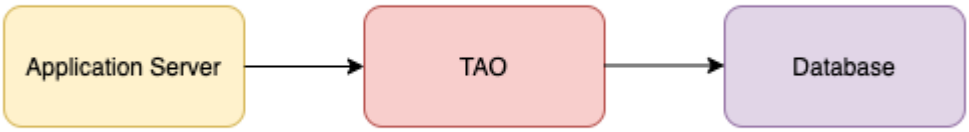


Fig.9.Facebook backend design

i. TAO

TAO is Facebook's distributed data store. It serves two primary purposes:

a) **Define data access API**

TAO exposes a list of APIs to query and mutate objects and associations. Facebook's application servers will talk to TAO instead of the database. The APIs can be classified into 2 categories.

- Object related APIs
- Association related APIs.

b) Behaves as a write-through cache

It aggressively caches the objects and associations to reduce latency and load on the database system.

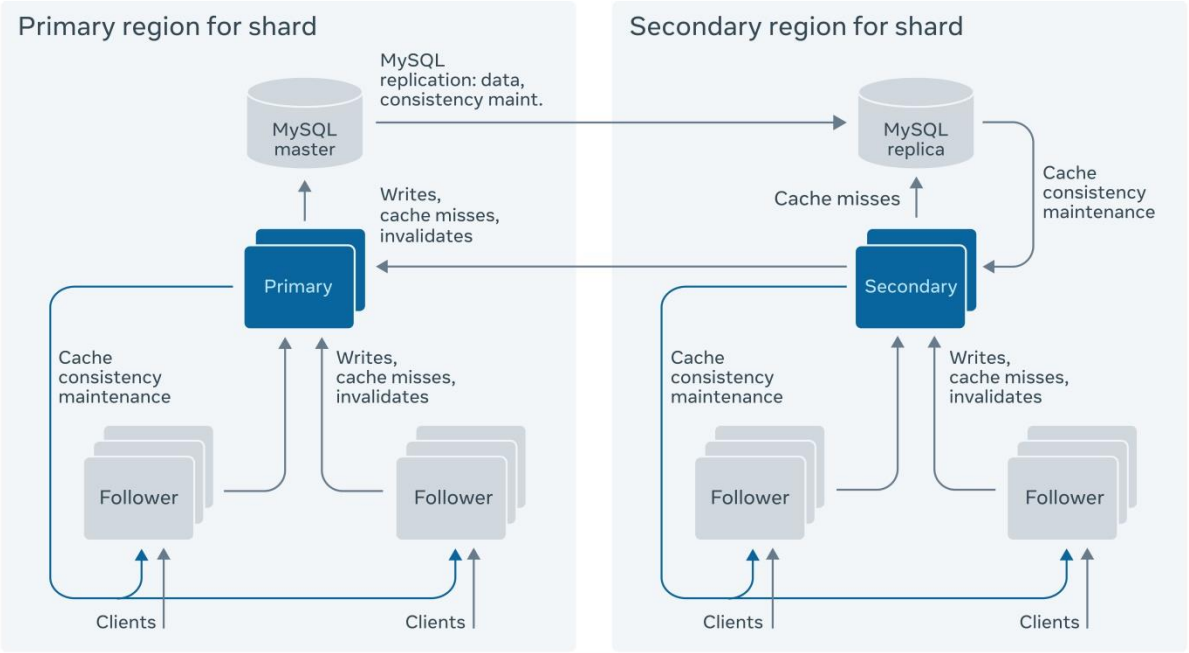


Fig.10.Tao system design

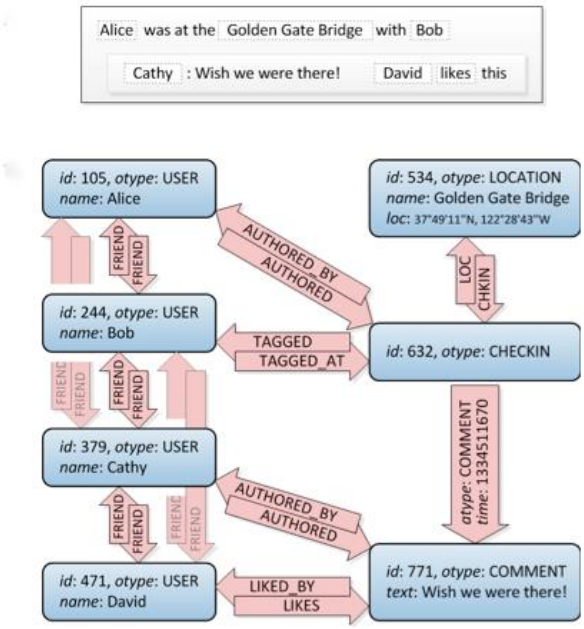


Fig.11.Tao datamodel and API

Before Tao, Facebook's servers directly accessed MySQL to read/write the social graph. Memcached was used as a look-aside cache. But it had several issues:

- **Inefficient edge list** - A key-value store is not a good design for storing a list of edges.
- **Distributed Control Logic** - In look-aside cache architecture, the control logic runs on the clients which increase the number of failure modes.
- **Expensive Read-After-Write Consistency** - Facebook used asynchronous master-slave replication for MySQL which introduced a time lag before latest data would reflect in the local replicas.

In early 2009, a team of Facebook infrastructure engineers started to work on TAO ("The Associations and Objects") to handle these issues. TAO has now been in production for several years. It runs on a large collection of geographically distributed server clusters. TAO serves

thousands of data types and handles over a billion read requests and millions of write requests every second.

4. Architecture

Facebook has one of the biggest hardware infrastructure in the world with more than 60.000 servers. Facebook born in 2004 on a LAMP stack,A PHP website with MySQL datastore. (5)

LAMP is an open source Web development platform that uses

- Linux as the operating system
- Apache as the Web server
- MySQL as the relational database management system
- PHP as the object-oriented scripting language.

Because the platform has four layers, LAMP is sometimes referred to as a LAMP stack.

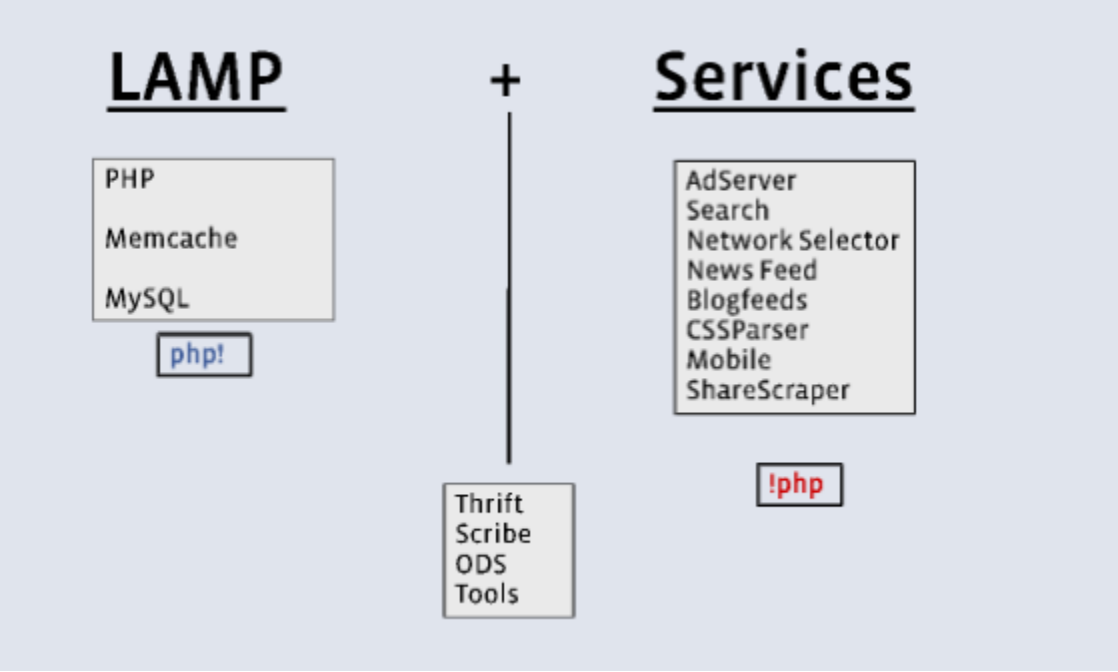


Fig.12. Facebook architecture overview

Although Facebook is a LAMP website, it had to change and extend its operation to incorporate a lot of other elements and services, and modify the approach to existing ones.

For example:

- Facebook still uses PHP, but it has built a compiler for it so it can be turned into native code on its web servers, thus boosting performance. The customization they did in PHP are:
    - Op-code optimization
    - APC improvement
    - Custom extensions
- Facebook uses Linux, but has optimized it for its own purposes (especially in terms of network throughput).
- Facebook uses MySQL, but primarily as a key-value persistent storage, moving joins and logic onto the web servers since optimizations are easier to perform there.
    - Custom partitioning scheme
    - Custom archiving scheme

Caching System: Facebook uses Memcached as a caching layer between the web servers and MySQL servers. The Memcached infrastructure is based on more than 800 servers with 28TB of memory.

Facebook improved the open source version of memcached and used it as a building block to construct a distributed key-value store for the largest social network in the world. (6) The customizations that were made on Memcached are the followings:

- Memcached over UDP
- On demand aggregation of per thread stats
- Multiple Kernel changes to optimize Memcache usuage (7)

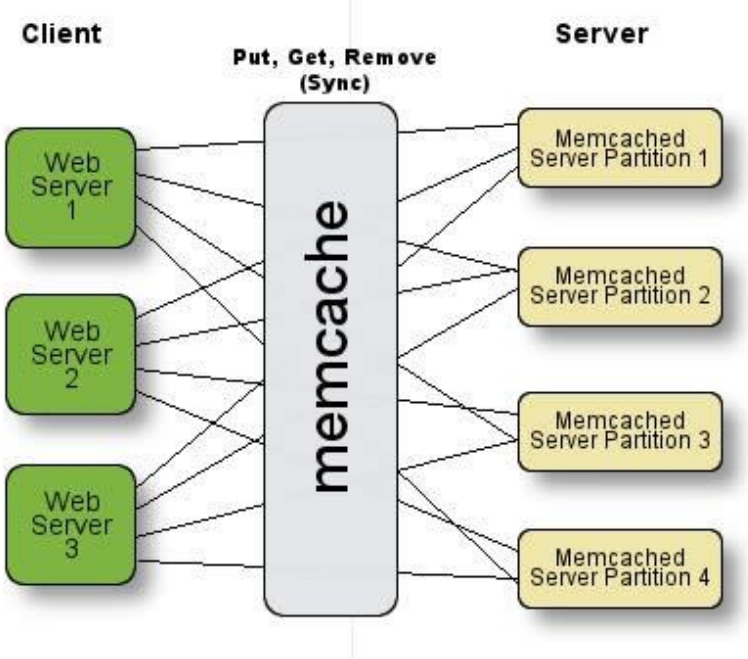


Fig.13. Memcached architecture

So the overall facebook architecture consists of some major data sturctures. They are:

i. Front end:
   a. HipHop
   b. BigPipe

ii. Service aggregiators:
   a. Scribe
   b. Thrift

iii. Datastore/Backend:
   a. MySQL

b. HBase
c. Haystack
d. Memcached

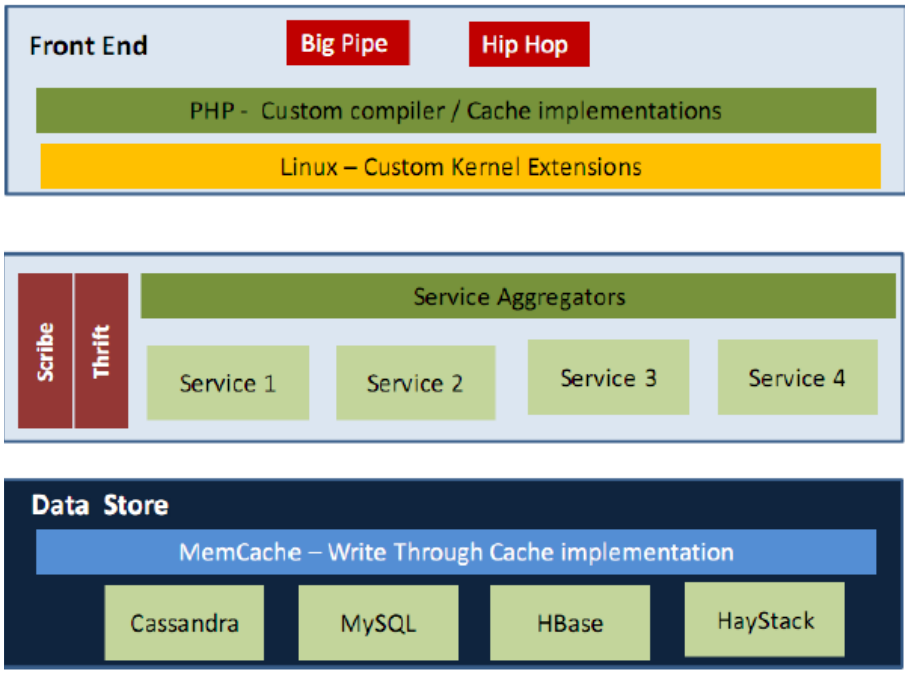


Fig.14.Architectural and technological overview of Facebook

4.1. Major Data structure

i. HipHop for PHP and HipHop Virtual Machine (HHVM)

HipHop for PHPPHP, being a scripting language, is relatively slow when compared to code that runs natively on a server. HipHop converts PHP into C++ code which can then be compiled for better performance. This has allowed Facebook to get much more out of its web servers since Facebook relies heavily on PHP to serve content.

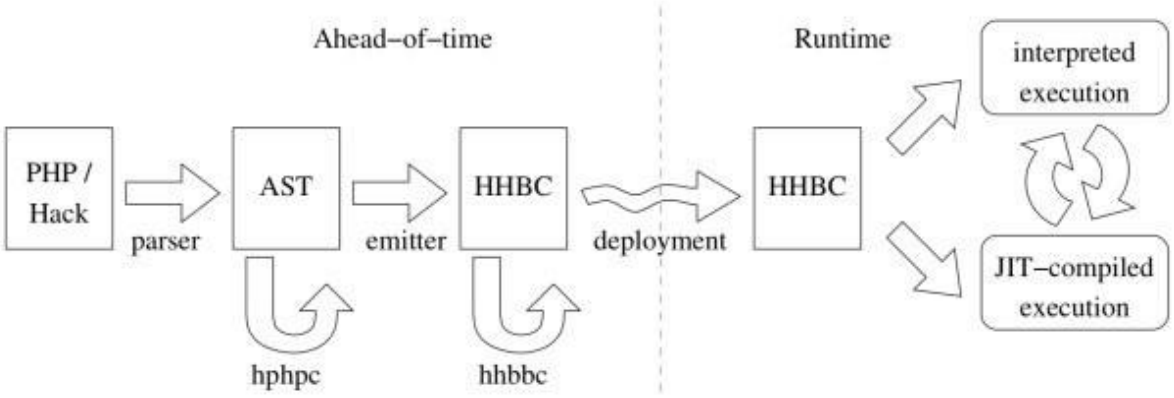


Fig.15. Overview of HHVM architecture

A small team of engineers (initially just three of them) at Facebook spent 18 months developing HipHop, and it was used for a few years. The project was discontinued back in 2013 and then replaced by HHVM (HipHop Virtual Machine).

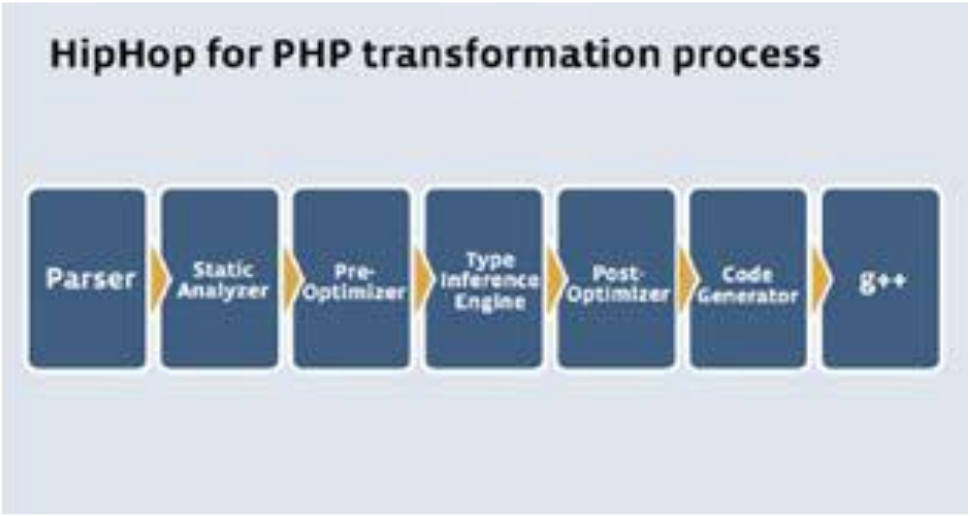


Fig.16. HipHop transformation process

ii. Haystack

Haystack is Facebook's high-performance photo storage/retrieval system (strictly speaking, Haystack is an object store, so it doesn't necessarily have to store photos). It has a ton of work to do; there are more than 20 billion uploaded photos on Facebook, and each one is saved in four different resolutions, resulting in more than 80 billion photos.

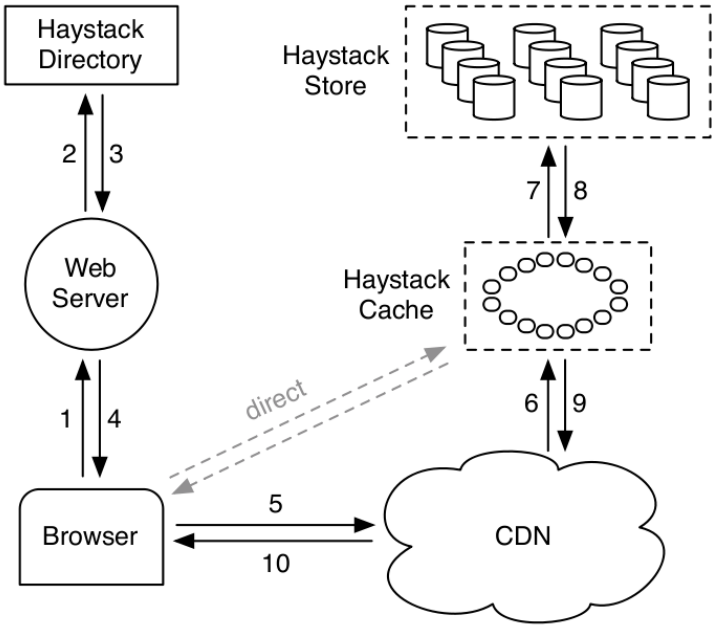


Fig.17.Haystack architecture

And it's not just about being able to handle billions of photos; web performance is critical. As we mentioned previously, Facebook users upload around 147,000 photos every minute which makes it 2,450 photos per second.

iii. BigPipe

BigPipe is a dynamic web page serving system that Facebook has developed. Facebook uses it to serve each web page in sections (called "pagelets") for optimal performance.

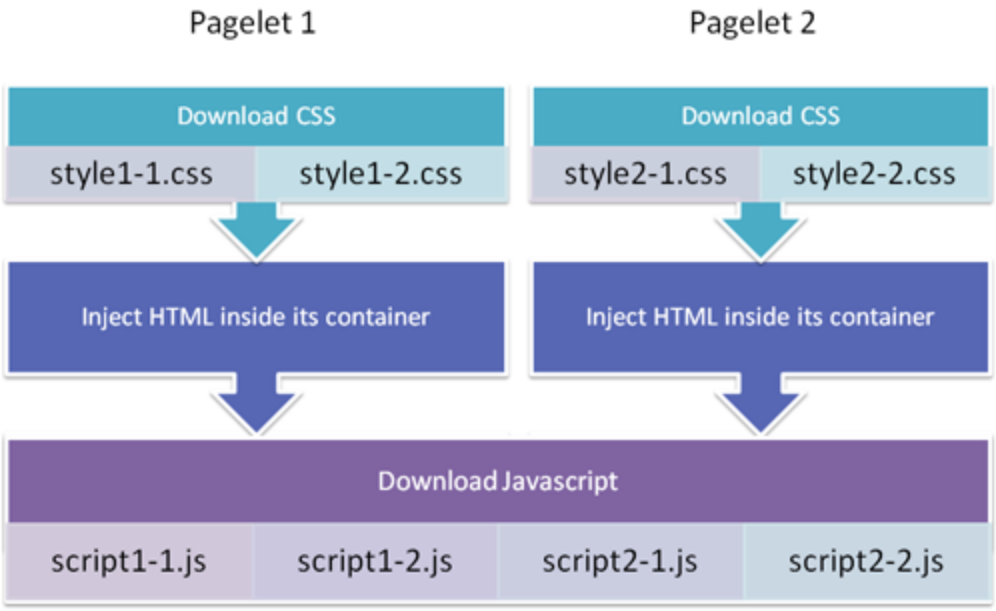


Fig.18.BigPipe architecture

For example, the chat window is retrieved separately, the news feed is retrieved separately, and so on. These pagelets can be retrieved in parallel, which is where the performance gain comes in, and it also gives users a site that works even if some part of it would be deactivated or broken.

iv. Scribe

Scribe was a flexible logging system that Facebook used for a multitude of purposes internally. It's been built to be able to handle logging at the scale of Facebook, and automatically handles new logging categories as they show up (Facebook has hundreds). As of 2019, Scribe's GitHub repository states that this project is no longer supported or updated by Facebook which probably means that it's not in use anymore.

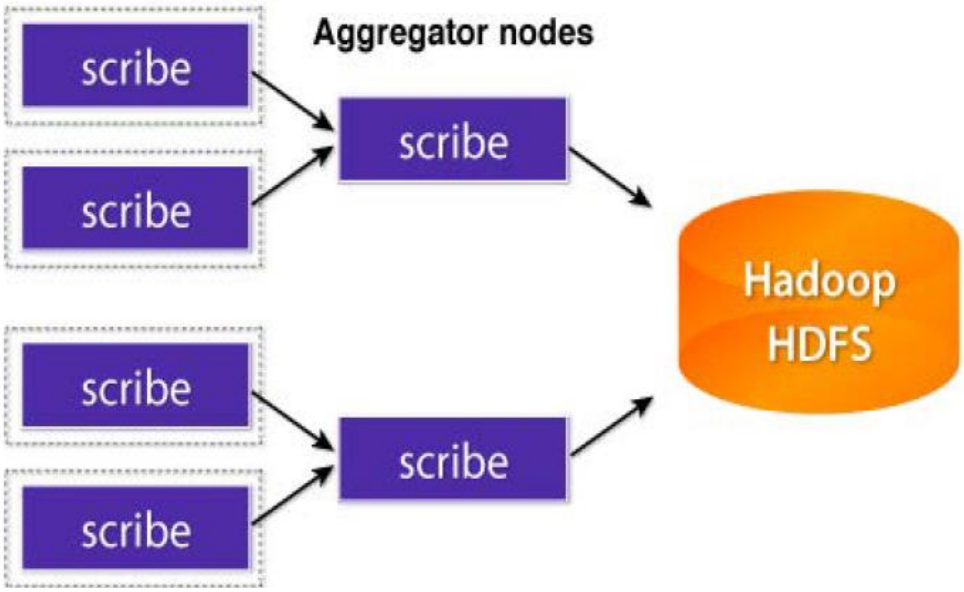


Fig.19.Scribe logging system

v. Hadoop and Hive

Hadoop is an open source map-reduce implementation that makes it possible to perform calculations on massive amounts of data. Facebook uses this for data analysis (and as we all know, Facebook has massive amounts of data). Hive originated from within Facebook, and makes it possible to use SQL queries against Hadoop, making it easier for non-programmers to use.

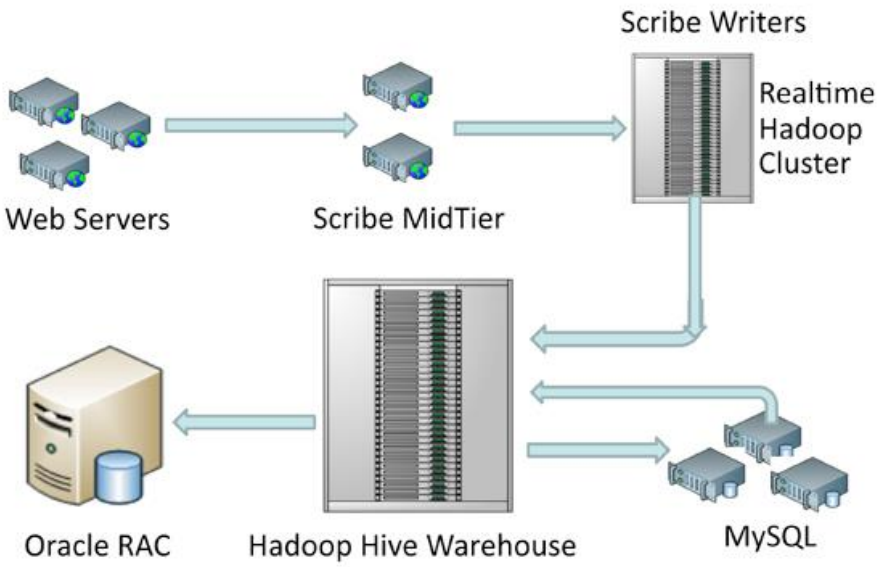


Fig.20.Hadoop, Hive and Scibe architecture overview

Both Hadoop and Hive are open source (Apache projects) and are used by a number of big services, for example Yahoo and Twitter.

vi. Thrift

Facebook uses several different languages for its different services. PHP is used for the front-end, Erlang is used for Chat, Java and C++ are also used in several places (and perhaps other languages as well). Thrift is an internally developed cross-language framework that ties all of these different languages together, making it possible for them to talk to each other. This has made it much easier for Facebook to keep up its cross-language development.

Facebook has made Thrift open source and support for even more languages has been added.

vii. Varnish

Varnish is an HTTP accelerator which can act as a load balancer and also cache content which can then be served lightning-fast.

Facebook uses Varnish to serve photos and profile pictures, handling billions of requests every day. Like almost everything Facebook uses, Varnish is open source.

viii. React

React logoReact is an open-source JavaScript library created in 2011 by Jordan Walke, a software engineer at Facebook. Later, Facebook introduced React Fiber, which is a collection of algorithms for rendering graphics. Interestingly, React is now one of the world’s most widely used JavaScript libraries.

ix. HBase

HBase is Facebook's new real-time messaging system. It can store 135+ billion messages a month.

Facebook created Cassandra and it was purpose built for an inbox type application, but they found Cassandra's eventual consistency model wasn't a good match for their new real-time Messages product. Facebook also has an extensive MySQL infrastructure, but they found performance suffered as data set and indexes grew larger. And they could have built their own, but they chose HBase.

HBase is a scaleout table store supporting very high rates of row-level updates over massive amounts of data. Exactly what is needed for a Messaging system. HBase is also a column based key-value store built on the BigTable model. It's good at fetching rows by key or scanning ranges of rows and filtering. Also what is needed for a Messaging system. Complex queries are not supported however. Queries are generally given over to an analytics tool like Hive, which Facebook created to make sense of their multi-petabyte data warehouse, and Hive is based on Hadoop's file system, HDFS, which is also used by HBase.

5. Layout

Facebook has been updating and experimenting with its layout from the starting of it’s journey.

Fig.21.Facebook’s old layout(pre 2020)

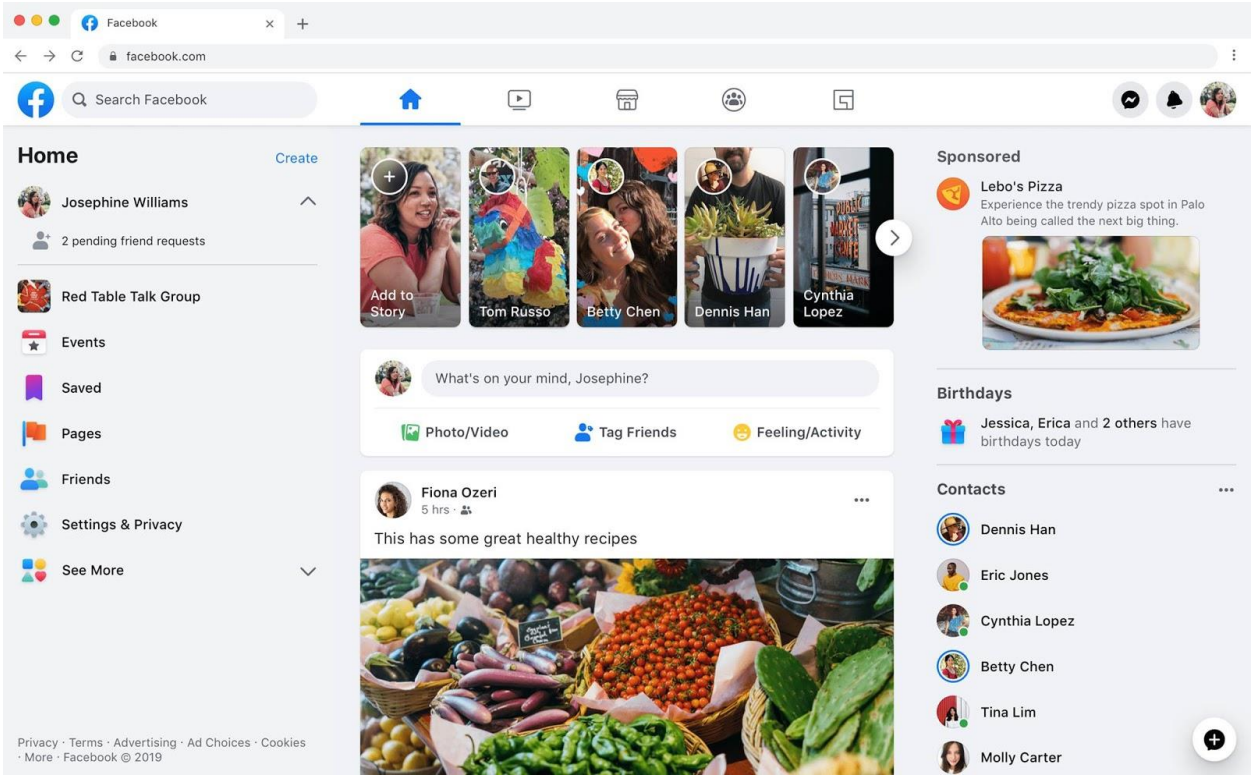


Fig.22. Facebook's new layout in 2020

To create a user account, one has to sign up in Facebook by filling up a form providing necessary information.

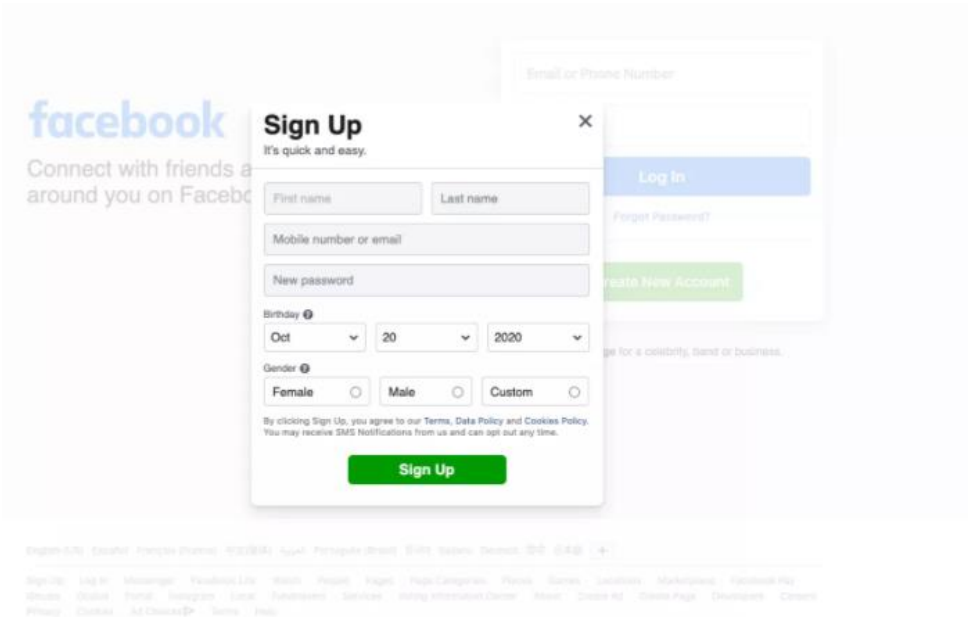


Fig.23.Creating user account

Facebook structure consists of the following features:

i. **Newsfeed:** The news feed is the primary system through which users are exposed to content posted on the network. Using a secret method (initially known as EdgeRank), Facebook selects a handful of updates to actually show users every time they visit their feed, out of an average of 1500 updates they can potentially receive

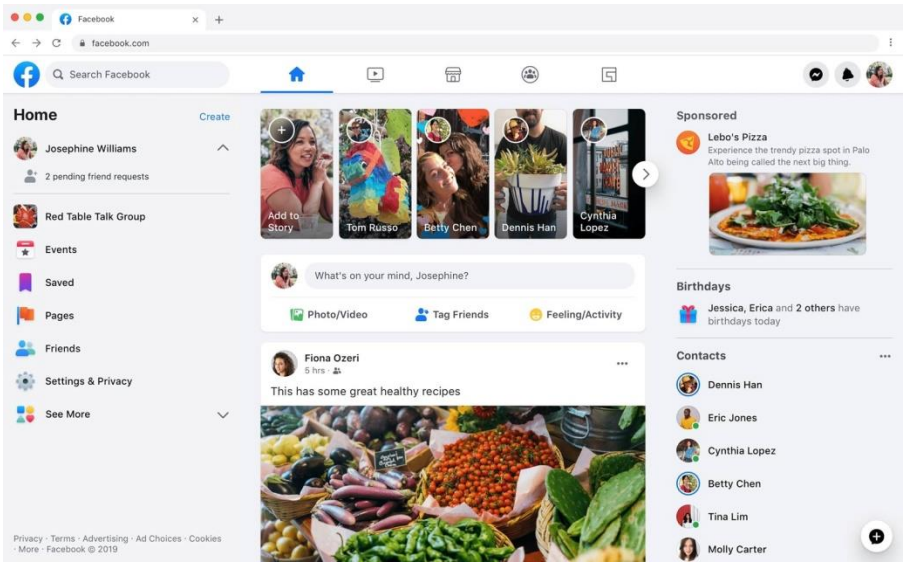


Fig.24.Facebook newsfeed

ii. **Friend:** "Friending" someone on the platform is the act of sending another user a "friend request" on Facebook. The two people are Facebook friends once the receiving party accepts the friend request. In addition to accepting the request, the user has the option of declining the friend request or hiding it using the "Not Now" feature. Deleting a friend request removes the request, but does allow the sender to resend it in the future. The "Not Now" feature

hides the request but does not delete it, allowing the receiver to revisit the request at a later date.

iii. **Wall:** The Wall is the original profile space where Facebook users' content until December 2011 was displayed. It allowed the posting of messages, often short or temporal notes, for the user to see while displaying the time and date the message was written. A user's Wall is visible to anyone with the ability to see their full profile, and friends' Wall posts appear in the user's News Feed.

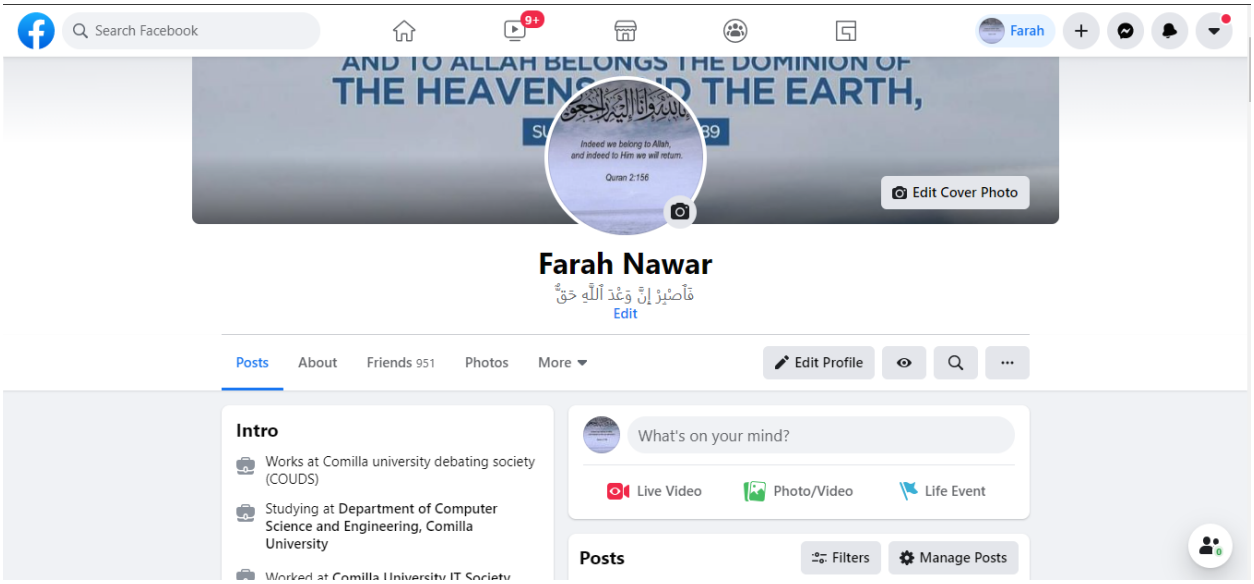


Fig.25.Facebook wall of a user

iv. **Timeline:**In September 2011, Facebook introduced "Timeline" at its developer conference, intended to revamp users' profiles in order to show content based on year, month and date. "Cover" photos were introduced, taking up a significant portion of the top of pages, and a redesigned display of personal information such as friends, likes and photos appeared on the left-hand side, while story posts appeared on the right. The new design introduced flexible sizing for story posts in the feed, along with more prominent location and photo placement. The Timeline also

encouraged scrolling, with constantly loading story posts of users' pasts.

v. **Likes and reactions:** The like button, first enabled on February 9, 2009,enables users to easily interact with status updates, comments, photos, links shared by friends, videos and advertisements. Once clicked by a user, the designated content appears in the News Feeds of that user's friends. acebook officially rolled out "Reactions" to users worldwide on February 24, 2016, letting users long-press on the like button for an option to use one of five pre-defined emotions, including "Love", "Haha", "Wow", "Sad", or "Angry".

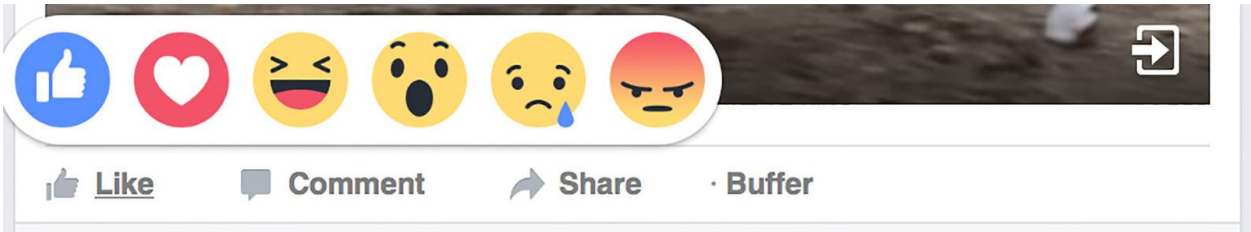


Fig.26.Facebook reactions in a post

vi. **Comments:** Users can share their views on other user's post through comments section.

vii. **Messages and inbox:** Facebook Messenger lets Facebook users send messages to each other. Complementing regular conversations, Messenger lets users make voice calls and video calls both in one-to-one interactions and in group conversations.

viii. **Notifications:** Notifications tell the user that something has been added to his or her profile page. Examples include: a message being shared on the user's wall or a comment on a picture of the user or on a picture that the user has previously commented on.

Initially, notifications for events were limited to one per event; these were eventually grouped category wise.

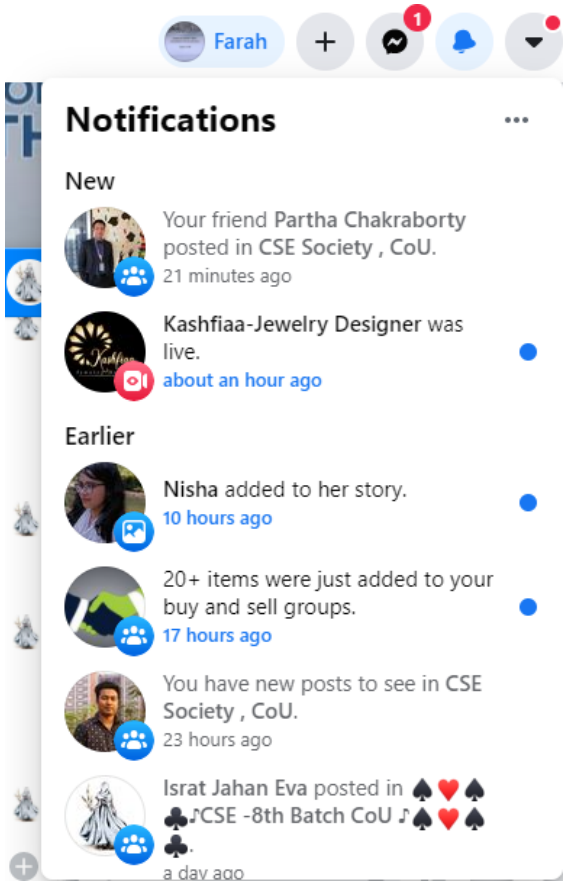


Fig.27.Notification

ix. **Groups:** Groups are used for collaboration and allow discussions, events, and numerous other activities. They are a way of enabling a number of people to come together online to share information and discuss specific subjects. They are increasingly used by clubs, companies and public sector organizations to engage with stakeholders, be they members of the public, employees, members, service users, shareholders or customers. Groups can have three different levels of privacy settings:

- "Open" means both the group, its members and their comments are visible to the public (which includes non-members) but they cannot interact without joining.
- "Secret" means that nothing can be viewed by the public unless a member specifically invites another user to join the group

6. Working Procedure

To explore Facebook, one must create a free account on the site. Facebook's terms of use state that members must be at least 13 years old, and any member between the ages of 13 and 18 must be enrolled in school. Facebook requires new members to provide a valid e-mail address before completing registration. After that Facebook will generate a profile for the user.

Facebook provides several ways to find friends:

- One can browse and join **networks**, which are organized into four categories:
  - **regions** (networks that are linked to specific cities or countries)
  - **Colleges**
  - **Workplaces**
  - **high schools**

  Once the user joins a network, he can browse through the list of members and search for people he knows. He can sort people by age, sex, relationship status, political views and other criteria.

- The user can let Facebook pull contacts from a Web-based e-mail account. To do this, he has to give Facebook his e-mail address and password. Facebook uses a program that searches

through his e-mail contacts and compares the list against its membership database. Whenever Facebook discovers a match, it gives him the option to add that person as a friend.
- The user can use Facebook's search engine to look for a specific person by typing the person's name into the search field, and Facebook will display any profiles that match the name.

Facebook is a LAMP website that uses PHP, but it has built a compiler for it so it can be turned into native code on its web servers, thus boosting performance. Facebook works on Linux, but it has optimised it for its own purposes (especially in terms of network throughput) .Facebook uses MySQL as a data sever with a key-value persistent storage, moving joins and logic onto the web servers since optimisations are easier to perform there. (Memcached layers) And some others developed internally: Haystack, a highly scalable object store used to serve Facebook’s immense amount of photos; Scribe, a logging system that can operate at it’s scale but it seems it was abandoned this year. So we expect soon to know about the new logging system.

Memcached at Facebook targets a workload dominated by reads (which are two orders of magnitude more frequent than writes). Facebook needs to support a very heavy read load, over 1 billion reads/second, and needs to insulate backend services from high read rates with *very high fan-out*. To this end, Facebook uses memcache as a "demand-filled look-aside cache". When a web server needs data, it first requests the value from memcache by providing a string key. If it is a cache hit, great, the read operation is served. If the item addressed by that key is not cached, the web server retrieves the data from the backend database and populates the cache with the key-value pair (for the benefit of the upcoming reads).

For write requests, the web server issues SQL statements to the database and then sends a delete request to memcache to invalidate the stale data, if any. Facebook chooses to delete cached data instead of updating it in cache because deletes are idempotent (which is a nice property to have in distributed systems). This is safe because memcache is not the authoritative source of the data and is therefore allowed to evict cached data.

For web page serving system they use BigPipe. It serves each web page in sections called “pagelets”. The chat window and the newsfeed in there is retrieved separately. For calculations on the massive amounts of data they use Hadoop in combinations with Hive. Both Hadoop and Hive are open source (Apache projects) and are used by a number of big services, for example Yahoo and Twitter.

They use an open source http accelerator Varnish which can act as a load balancer and also cache content fast.To serve billions of photos they use React as a JavaScript library - collection of algorithms for rendering graphics. And for all they use Thrift internally developed cross-language framework that ties all of these different languages together, making it possible for them to talk to each other.

7. Algorithm

- **Edgerank Algorithm**

EdgeRank is the name commonly given to the algorithm that Facebook uses to determine what articles should be displayed in a user's News Feed. As of 2011, Facebook has stopped using the EdgeRank system and uses a machine learning algorithm that, as of 2013, takes more than 100,000 factors into account.

A simplified version of the EdgeRank algorithm was presented as:

$$\Sigma_{edges}\ u_e w_e d_e$$

Where,

u_e is user affinity.

w_e is how the content is weighted.

d_e is a time-based decay parameter.

Three main parts to EdgeRank called

- **Affinity:** Affinity is your relationship with users. It means that the amount a user has interacted with your Page in the past affects how likely they are to see your posts. So if they've clicked your links, liked or commented on your posts in the past, EdgeRank assumes they're pretty keen to see more of your stuff, and gives it higher priority in that user's News Feed.
- **Weight:** is how much priority EdgeRank gives to your post, based on the post type. Facebook has a hierarchy of post types, since some types garner more engagement than others. Photos and videos take top priority. Links are second, and plain text status updates are at the bottom end. Weight doesn't end there, though.
- **Time decay:** Time decay simply means how old a post is. The longer your post has been on Facebook, the less likely it will be to show up in a user's News Feed. Again, this one's not quite cut-and-dried, though. For users who only drop-in to Facebook once or twice a week, your post will be relevant for longer than it would for those users who check in twice a day.

Although there are those three main areas of the EdgeRank algorithm, there are actually four specific action-points that are used by EdgeRank each time you post. These aren't necessarily fool-proof guarantees that your post will hit the News Feeds of your fans, but they're all good indicators of the likelihood of that.

The four specific points are these:

i. user's past interactions with the author
ii. user's past interactions with that post type
iii. reactions from other users for that particular post
iv. amount of complaints or negative feedback on that post

In March of 2013 Facebook announced the first major change to the Facebook News Feed in six years. The theme at the time was reduction of clutter by providing more choice and control over the content and stories users will see in their feeds.

i. **Story Bumping** – Story bumping considers showing not only all new stories, but considers all stories that are new to the user that may have driven engagement since the last time you visited. If you miss stories on one visit, they may be eligible to show on the next visit. This was tested with 80% of employees, older stories were at the top of the feed and interaction went up on organic content during internal testing. Stories read went from 57% to 70%.

ii. **Last Actor** – This is another new feature to the news feed algorithm and "captures your current state of mind". This new element takes into consideration the most recent factors from your Facebook activity. This ties to your last 50 interactions with content and gives those users a slight bump in news feed ranking. An example is if you interact with an individual in the morning

that same person’s content may be weighted slightly higher later in the day.
This has led to 1 to 2% bump in the number of interactions with posts. This feature is now live on mobile & web.

- **Deepface Algorithm**

DeepFace is a deep learning facial recognition system created by a research group at Facebook. It identifies human faces in digital images. It employs a nine-layer neural network with over 120 million connection weights and was trained on four million images uploaded by Facebook users.DeepFace shows human-level performance. The Facebook Research team has stated that the DeepFace method reaches an accuracy of 97.35% ± 0.25% on Labeled Faces in the Wild (LFW) data set where human beings have 97.53%.This means that DeepFace is sometimes more successful than the human beings. However, DeepFace model falls behind Google FaceNet which got 99.65% on the same data set.It also leaves behind the FBI's Next Generation Identification system which have 85% performance .One of the creators of the software, Yaniv Taigman, came to Facebook via their 2007 acquisition of Face.com.

8. Challenges and how to overcome them

The main challenge Facebook engineer's face, despite receiving billions of users on a daily basis, is keeping the website online and functional. Scalability is essential for web applications with massive numbers of users. The high growth rates observed in web based systems having hundreds of thousands of users accessing it continuously led to major response time problems for users who are trying to receive information

at the same time. As more and more users access a web site, one needs to know how the performance varies.

The current architecture of Facebook is very large and consists of many technologies and thousands of servers. (8) Facebook handles this large amount of data through different optimization algorithms, data warehouses,distributed data servers and memcached system.

Facebook improved the open source version of memcached and used it as a building block to construct a distributed key-value store for the largest social network in the world.

Facebook has also dome many customization in the LAMP service.

For example:

- Facebook still uses PHP, but it has built a compiler for it so it can be turned into native code on its web servers, thus boosting performance. The customization they did in PHP are:
    - Op-code optimization
    - APC improvement
    - Custom extensions
- Facebook uses Linux, but has optimized it for its own purposes (especially in terms of network throughput).
- Facebook uses MySQL, but primarily as a key-value persistent storage, moving joins and logic onto the web servers since optimizations are easier to perform there.
    - Custom partitioning scheme
    - Custom archiving scheme

9. Findings

We have found out the following things about the architecture of facebook:

1. **Layered** - components are independent and isolated.
2. **Service/API Driven** - each layer is connected via well defined interface that is the sole entry point for accessing that service. This prevents nasty complicated interdependencies. Clients hide behind an application API. Applications use a data access layer.
3. **Distributed** - layers are transparently distributed across clusters of machines
4. **Separate Application Logic** - Application logic is encapsulated in application servers that provide an API endpoint. Application logic is implemented in terms of other services. The application server tier also hides a write-through cache as this is the only place user data is written or retrieved, it is the perfect spot for a cache.
5. **Stateless** - State is kept in a database tier, caching tier, or other services, not in applications or stuffed in local pockets.
6. **Scalable Component Services** - Other services that themselves are scalable and highly available are used to implement this service. Messages also uses: Memcache, User index service, and HayStack.
7. **Full Stack Ops** - Tools were developed to manage, monitor, identify performance issues and upgrade the entire service, up, down and across the stack.

8. **Celled** - Messages has as the basic building block of their system a cluster of machines and services called a **cell**. If you need more power captain, then cells are added in an incremental fashion. A cell consists of ZooKeeper controllers, an application server cluster, and a metadata store. Cells provide isolation as the storage and application horsepower to process requests is independent of other cells. Cells can fail, be upgraded, and distributed across datacenters independent of other cells

10. Summary

Facebook is a social network founded in 2004 by Mark Zuckerberg. It is a website where anyone can have an online profile with their personal information, photos, videos, links, posts, etc. Members of this social network, interact by browsing profiles, making friends, establishing contacts, leaving comments, sending messages, and basically in one word, communicate. Within each member's personal profile, there are several key networking components. With over 2.7 billion monthly active users as of the second quarter of 2020, Facebook is the biggest social network worldwide. In the third quarter of 2012, the number of active Facebook users surpassed one billion, making it the first social network ever to do so.

For a web application like Facebook with massive numbers of users, Scalability is essential. The high growth rates observed in web based systems having hundreds of thousands of users accessing it continuously led to major response time problems for users who are trying to receive information at the same time. The main challenge Facebook engineer’s face, despite receiving billions of users on a daily basis, is keeping the

website online and functional. The current architecture of Facebook is very large and consists of many technologies and thousands of servers.

11. 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

1 Works Cited

1. *Facebook: A literature review.* **Caers, Ralf and De Feyter, Tim and De Couck, Marijke and Stough, Talia and Vigna, Claudia and Du Bois, Cind.** UK: London, England : Sage Publications Sage, 2013, Vol. 15.

2. **Phillips, Sarah.** A brief history of Facebook. *The Gurdian.* [Online] July 25, 2007.

3. **Mar, Nicholas Carlson.** At last — the full story of how Facebook was founded. *Business Insider.* [Online] March 2010.

4. *A reference architecture for big data systems in the national security domain.* **Cooper, John Klein and Ross Buglak and David Blockow and Troy Wuttke and Brenton.** s.l. : Proceedings of the 2nd International Workshop on BIG Data Software Engineering - BIGDSE '16, 2016.

5. **Mostosi, Andrea.** How it works: Facebook – Part 1. *Useful Stuff.* [Online] March 21, 2013.

6. *Scaling Memcache at Facebook.* **Rajesh Nishtala, Hans Fugal, Steven Grimm, Marc Kwiatkowski, Herman Lee, Harry C. Li,.** s.l. : 10th USENIX Symposium on Networked Systems Design and Implementation (NSDI ’13), April 3,2013.

7. **Agarwal, Aditya.** *Facebook: Science and the Social Graph.* [Video] s.l. : QCon SF, 2008.

8. *Overview of Facebook scalable architecture.* **Barrigas, Hugo, et al., et al.** s.l. : Proceedings of the International Conference on Information Systems and Design of Communication, 2014.