

Cloud Around Us

Kavita Tewani

Assistant Professor(CE)

ITM Universe,

Vadodara, India

kavitatewani012@gmail.com

Abstract—End-users share a wide variety of information on Facebook, without being aware of where the data is stored and how security issues are concerned. The paper discusses how Facebook magically implements cloud in real world and which cloud technologies are used to store the vast data user uploading on daily basis. Facebook was build on ground-up as a website or a “cloud application” SAAS (Software as a Service). It’s success with users is that it just works- no matter what device you are on, no matter what network connection you have, and no matter what type of media large or small you are sharing-IT JUST WORKS! Business grade Cloud IT services need to be the same! And depending on the complexity of the cloud service offered or looking to purchase, many still have some way to go in this respect. This paper gives the description of how Facebook use SQL database for storage purpose and all other technologies that Facebook uses to attract its users. The paper also covers some concepts of NoSQL (not only SQL), a broad class of database.

Keywords— Cloud Computing, NoSQL, Hadoop Distributed File System (HDFS), Mashup, Web 2.0

I. INTRODUCTION

Facebook has a huge amount of data and different application that utilizes cloud. It is a good example of cloud computing in real world. The data is stored without any limitation and the physical storage is not known to the user. Facebook also uses many technologies of web 2.0 like mashup for photos and videos. Many of the characteristics of cloud computing like ubiquitous network access, rapid elasticity, on demand self services, location independent resource pooling play important role in Facebook. The fact that Facebook is a cloud-based application SAAS is largely inconsequential- usurped by the fact that Facebook is Facebook. Its ecosystem is almost a redefinition of the Cloud phenomenon itself and is perhaps a more familiar term for many users. The Facebook Cloud is fast evolving into a Platform (Platform-as-a-service (PAAS)) as developers rush to develop apps for Facebook to tap its vast user-base concept. Whilst a single business cloud couldn’t hope to gain to notoriety and power of Facebook by serving the business IT services sector, there’s nothing to stop Cloud IT service providers creating more friendly and flexible technology ecosystems and – with the right formula – becoming known by name as a pioneer redefining the landscape of cloud computing as we know it.

II. WHAT IS CLOUD COMPUTING

The cloud computing is a computing service that charges you based only on the amount of computing resources we use. Cloud computing is internet based computing where virtual shared servers provide software, infrastructure, platform devices and other resources and hosting to customers on a pay as you use basis. All information that is digitized system has to

offer is provided as a service in the cloud computing model. Users can access these services available on the “internet cloud” without having any previous know-how on managing the resources involved. Thus users can concentrate more on their business processes.

Cloud using customers do not owe the physical infrastructure; rather they rent the usage from the third party provider (vendor). This helps them to avoid worrying about the huge software packages to be updated and to buy new one every time. They consume resources as a service and pay only for resources that they use. Most cloud computing infrastructures amongst can improve, as servers are not necessarily left idle, which can reduce costs significantly while increasing the speed of the application development.

III. CLOUD COMPUTING IN FACEBOOK FOR VARIOUS APPLICATIONS

As we observe Facebook keeps on updating the applications and design of the websites. Photos, videos, links and more can be posted on Facebook. One can upload the photos or videos of their wish and other users can simultaneously comment on it. This is known as multi tenancy, again an application of cloud computing, where many users can access and update at the same time. Many of us use cloud computing everyday with the sites like Facebook, Gmail, Yahoo etc. Photos and videos shared are stored on the servers, which can be accessed any time with an internet connection.

A. Mashup for photos and video

Mashup is one of the popular technologies of Web 2.0 which combines data sources from different places to provide new services on the World Wide Web. By using this technology, one can combine different features on a single page. As we can see on the timeline page of the profile, where the details about the status, current activities, photos etc is been placed on a single page.



Fig. 1. Example of mashup

B. Recommendation for pages and interest

Facebook has expanded its “Recommendations” feature on places pages to now show endorsements from non-friends. Previously users can only see “Recommendations from friends” on the right side bar. While recommendations from friends are probably more convincing than those of strangers, Facebook seems to have determined that both are worthwhile additions to pages. Now one can even see the recommendation given by Facebook itself on liking a particular page. It simply recommends few more pages according to the pages you have liked.

C. Social plugins

Social plugins are simple tools that can be dropped into any website to provide people with personalized and social experiences, just like recommendation and like the page or list of other people including your friends who liked this page. While these buttons and boxes appear on other websites, the content populating them comes directly from Facebook. The plugins were designed so that the website you are visiting receives none of this information. These plugins may be seen as an extension of Facebook. One can only see personalized experience with the friends if he/she is logged in. If user is not logged in, then log in page will be prompted directly before using the site

At a technical level, social plugins work when external websites put an iframe from Facebook.com on their site-as if they were agreeing to give Facebook some real estate on

their website. If one is logged in, the Facebook iframe recognize the user and show personalized content within the plugin as if the user were on Facebook.com directly. Even though the iframe is not on Facebook, it is designed with all the privacy protections as if it were.

IV. HOW DATA IS STORED IN FACEBOOK

So how much Big Data is in Facebook and where it is kept? The popular details making the rounds of Hadoop and Big data conferences focuses mainly on the huge clusters running Facebook data warehouses running on Hadoop and Hive. Facebook’s corporate blog about their massive hadoop migration last year to a larger data center. However on daily basis, the repository and platform you interact with is still powered by MySQL databases. Given the publicity around how traditional “relational” databases can’t handle internet scale, and that NoSQL databases are the way to go, the fact that Facebook still operates MySQL as the backend is eye opening.

A survey shows the details that how Facebook is not just relying purely on MySQL, and that they have a massive layer of memcached servers that are being used as an in-memory database highlighting that MySQL servers on their own couldn’t possibly handle the real load of live Facebook traffic. For functionality such as the Hadoop and HBase are used instead. Additionally Hadoop is used as the backup for the MySQL data.

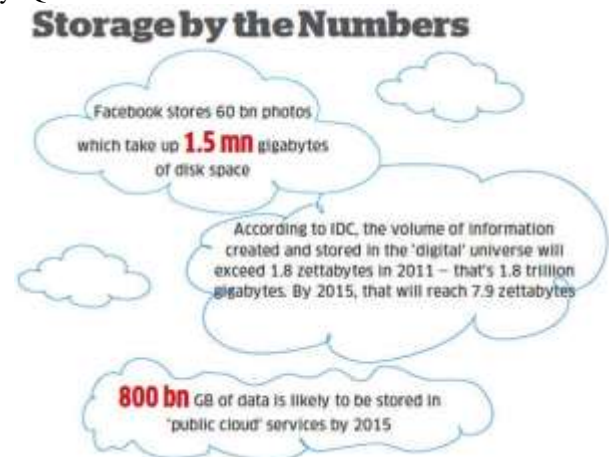


Fig. 2. Facebook storage

A. How NoSQL differs from MySQL

NoSQL is a new way of thinking about the database. NoSQL is not a relational database. The reality is that a relational database model may not be the best solution for all situations. The easiest way to think NoSQL is that of a database which does not adhering to the traditional relational database management (RDBMS) structure. Sometimes you will also see that it is referred to as ‘not only SQL’.

It is not built on tables and does not employ SQL to manipulate data. It also may not provide full ACID (atomicity,

consistency, isolation, durability) guarantees, but still has a distributed and fault tolerant architecture. The NoSQL taxonomy supports key-value stores, document stores, BigTable, and graph databases. NoSQL is gaining momentum and is supported by Hadoop, MangoDB and others. The NoSQL Database site is a good reference for someone looking for more information.

V. MIGRATION OF VIRTUAL MACHINES

For migrating data a couple of different migration strategies are used. One was a physical move of the machines to the new data center. Moving all machines to the datacenter is an easy task, however this is not a viable option as the users and analysts depend on the data 24/7, and downtime would be too long. Another approach was to set up a replication system that mirrors changes from the old cluster to the new, larger cluster. Everything is simply redirected to the new cluster. This approach is more complex as the source is a live file system, with files being created and detected continuously. Due to the unprecedented cluster size, a new replication system that could handle the load would need to be developed. However, because replication minimizes downtime, it was the approach that is used for massive migration.

Once the required system for migration is developed, the replication approach is executed in two steps. First, a bulk copy is transferred, including most of the data from the source cluster to the destination. Most of the directories were copied via DistCp-an application shipped with Hadoop that uses a MapReduce job to copy files in parallel. Hadoop engineers made code and configuration changes to handle special cases with Facebook's dataset, including the ability for multiple mappers to copy a single large file, and for the proper handling of directories with many files. After the bulk copy is done, file changes after the start of the bulk copy were copied over to the destination cluster through the new replication system. File changes are detected through a custom Hive plug-in that recorded the changes to an audit log. The replication system continuously poll the audit log copies modified files so that the destination would never be more than a couple of hours behind. The plug-in recorded Hive metadata changes as well, so that metadata modifications such as the last accessed time of Hive tables and partitions were propagated. Both the plug-in and the replication system were developed in-house by members of the Hive team.

At the final migration switchover time, we set up camp in a war room and shut down Hadoop Job Tracker so that new files would not be created. Then, the replication system is allowed to catch up. During this time, user's directories are copied as well. Once replication is caught up, both clusters become identical, and the DNS entries are changed so that the host name referenced by Hadoop jobs pointed to the servers in the new cluster. The Job Tracker in the new data center is started and the jobs are able to run as usual, with no modifications required.

VI. FUTURE SCOPE

Right now we are in the early days of cloud computing, with many organisations taking their first, tentative steps. But by 2020 cloud is going to be major – and permanent – part of the enterprise computing infrastructure.

. With increased demand of cloud in enterprise, there will be development in the technologies that support clouds, with rapid increases in processing power making cloud projects even cheaper while technologies currently limited to supercomputing will make it into the mainstream.

Following points shows where cloud computing will be in few years.

- Software will be floated away from hardwares. The director of HP recently quoted :” Cloud computing is the final means by which computing becomes invisible”.
- Datacentres will beome ecosystem: The twinned technologies of abstracted software and commodified hardware should combine to make datacentres function much more like ecosystems, with an overarching system ruling equipment via software, with hardware controlled from a single point, but growing and shrinking according to the workloads.
- Medical treatments will be simplified: The future of cloud computing is not confined to entertainment and gaming options as it can contribute massively in the fields of medical science as well. As most of the contemporary treatment require computer assistance, as data have to be searched for various things like DNA samples and other biochemical procedures and hence cloud computing is going to play its part in most of the therapies. In addition, it will make easy the task of data processing.
- A new age safety system: With the help of cloud computing, records of cars including number, driving license and address details of the owner can be stored on the cloud in case the car is stolen one and recovered by some security agencies in far off locations, they can instantly inform the owner of the car. It also lessens the burden of the various police and security organizations in a given region.

VII. CONCLUSION

As cloud computing is internet based technology and it is getting popularity very swiftly, the day is not far when many of the enterprises will use cloud to store data and to use other features of cloud. Not just Facebook, but many companies like Google, Microsoft also make use of cloud. Cloud computing is a style of computing in which dynamically scalable and often virtualized resources are provided as a service over the internet. Cloud computing has become a significant technology trend, and many experts expect that it will reshape IT process and the IT marketplace as availability and scalability are the main

features of cloud computing, it will become helpful for many enterprises. The seemingly limitless potential of cloud computing's huge, cost-efficient data centers without minimizing the problems facing this emerging platform. In short a fundamental shift in digital services is taking place.

REFERENCES

- [1] Enterprise Cloud Computing: Technology, Architecture, Applications by Gautam Shroff
- [2] Mr. Gurmeet Singh, Mr. Vineet Kumar Sachdeva, Impact and Challenges of Cloud Computing in Current Scenario,
- [3] International Journal of Social Science & Interdisciplinary Research, vol. 1 Issue 10, October 2012
- [4] Harjit Singh Lamba, Gurdev Singh, Cloud Computing-Future Framework for e-management of NGO's, International Journal of Advancements in Technology, vol. 2, Issue 3, July 2011.
- [5] nosql-database.org
- [6] Cloud Computing Bible by Barrie Sosinsky
- [7] www.roseindia.net/cloudcomputing
- [8] searchcloudcomputing.techtarget.com/definition/Hadoop
- [9] Rajesh Piplode, Umesh Kumar Singh, An overview and Study of Security Issues & Challenges In Cloud Computing, vol. 2, Issue 9, September 2012.