A Survey of Cloud Computing and Social Networks

Sanjay P. Ahuja[1] & Bryan Moore[1]

[1] School of Computing, University of North Florida, Jacksonville, USA

Correspondence: Sanjay P. Ahuja, School of Computing, University of North Florida, Jacksonville, FL 32224, USA. E-mail: sahuja@unf.edu

Abstract

In recent years there has been rapid growth in cloud computing and social networking technologies. Cloud computing shifts the computing resources to a third party, eliminating the need to purchase, configure and maintain those resources. With the incentive of lowered operational costs in software, hardware and human effort, many companies are considering the use of cloud services. Likewise, social networks have seen massive growth, with millions of Internet users actively participating across various social networking websites. Even corporations have begun using social networks as a means to market and reach their customers. This paper will survey the current issues in cloud computing and social networks and how these technologies are being used together.

Keywords: cloud, social networks

1. Introduction

A social network is a structure of entities interconnected through a variety of relations. These entities are typically referred to as "users". The relationships between these users have a number of different names across different social networks such as friends, or followers. Through these relationships users share messages and media amongst themselves. There exist a number of online social networking websites such as the popular Facebook, LinkedIn and Twitter. These social networking sites have well over 100 million active members. With such a great number of users using these services, social networks present an interesting area of study in a variety of ways (Falahi, 2010).

Cloud computing refers to the access of computing resources across a network. These resources include but are not limited to networks, storage, servers and services. This model can provide a number of advantages. Chief among them is the reduction of costs. An organization can utilize cloud computing services from some third party when such resources are required and scale up and down as needed without needing to invest in costly infrastructure. Another major benefit is that applications and data can be accessible at any time through the Internet (Motta, 2012).

Cloud computing and social networking has intermingled in a variety of ways. Most obviously social networks can be hosted on cloud platforms or have scalable applications within the social networks. Recent research has also proposed cloud based applications using social networks for user management and authentication in a system called social clouds (Chard, 2010).

In this paper, we examine the trends and issues in cloud computing and social networking. The paper is organized as followed: Section 2 briefly reviews social networks and cloud computing, Section 3 describes the current applications and uses of both social networks and cloud computing, Section 4 examines the cloud architecture that is being employed, Section 5 reviews current applications, and Section 6 concludes the paper.

2. Background

Cloud computing refers to the access of services across a network, usually the Internet. Clouds typically have a number of common characteristics such as being allocated as needed, accessible across various internet-capable devices, and metered and billed based upon resource usage. These services have a number of different labels, but essentially they can be divided into three categories – Software as a Service, Platform as a Service and Infrastructure as a Service. Software as a Service describes a cloud application that is hosted in the cloud, and is provided to its users through the Internet. This model removes the need to install an application on the end user's

system and lowers the software cost through usage pricing. Platform as a Service encompasses the entire software development lifecycle. This includes the development environment, and production environment to deploy the application. Infrastructure as a Service provides the use of resources such as virtual machines, storage etc. Costs are again reduced by removing the need to procure, install and configure infrastructure. These are by no means an exhaustive list of cloud services, and may be referred to as a different name by different vendors. Cloud services generally follow four types of deployment models: Private Cloud, Public Cloud, Community Cloud and Hybrid Cloud. Private Clouds are used by a single organization. Public clouds are services provided to the public by a cloud service provider (CSP). Infrastructure costs rest on the CSP. Community clouds are where services are shared amongst multiple organizations and Hybrid Clouds mix the previous three models. One example of a hybrid cloud would be an organization that relies on both private and public clouds (Motta, 2012).

Social networks are networks of users connected through relationships such as friendship, following or otherwise. Through these relationships, users are able to share content amongst themselves. There are numerous existing social networking websites such as Orkut, Facebook, Linked.in, and Google+. On these sites, one of the greatest concerns has been the security and privacy of personal data. That is to control the personal information that is being shared to other users and social applications, as well as how information is being shared with third-parties. (Falahi, 2010).

3. Related Work

As social networks and applications reside on the web, cloud computing as an infrastructure or platform is a possible solution for these technologies. Cloud computing presents the same general advantages to social applications: a significant decrease in operational and infrastructure costs, along with the ease of scalability to meet the increasing or decreasing needs of the applications. These advantages would carry over to social network data analysis. With the various cloud resources and services available, the configuration between these technologies can vary.

Cloud Computing and social networks have numerous examples of being used together. Typically these involve the social network being hosted on a cloud platform or social applications being hosted on the cloud. Recent research has explored the idea of building cloud infrastructure leaning on the social network for the established relationships and user management it provides. In this type of system, users would provision their own resources or third party resources to other users based upon their previously established relationships in the social network. This type of cloud system would be built on top of existing social network as a social application. The major advantage in this type of configuration would be eschewing the management capabilities off the application and using the existing social networks capabilities (Chard, 2010).

With social networks having millions of users and even more user relationships they present large datasets for analysis. As with other data analytic applications, social network analysis can provide useful information about users. This could include things such as sediment analysis, locating key opinion users as well as more targeted applications such as mapping disease outbreaks, and natural disasters. In order to perform these types of analysis, considerable computing power may be required. Once again, cloud computing is a possible solution. Ting et al investigated techniques for cloud based analysis and data warehousing of social networks. Two algorithms were examined: MapReduce and Bulk Synchronous Parallel (BSP). MapReduce is a technique for processing large datasets developed by Google. MapReduce consists primarily of two functions: map and reduce. The map function divides the input and distributes to the nodes that process the data they receive. The reduce function combines and merges those results to generate the output. BSP is a modification of MapReduce using a parallel processing technique. They concluded that the BSP technique performed considerably better than the MapReduce. Although BSP displayed better performance in this case, the ease of use and maturity of existing applications and systems for MapReduce simplifies parallel programming complexity for processing datasets (Ting, 2011). Vokali et al developed a cloud based framework for the analysis of social networking trends. Their framework, Cloud4Trends, adapted and ported trend detection into a cloud application. Additionally their system incorporated status-detection where resources would be increased based upon job completion times. They concluded that the cloud-based social trend detection was a suitable solution given that trend detection is typically a computational intensive process due to the size of the datasets being analyzed. In this way, the amount of resources could be scaled up and down based on their system's needs (Vakali, 2012).

As with most other technologies, security is of great importance in cloud computing and social networking. Cloud computing security refers to the technology that is used to protect the data and applications of the cloud from threats such as unauthorized access, disruption of services, modification and others. In cloud security, the

common goals or objectives of information security still apply: confidentiality, integrity, and availability. With social networks attention should be given to the sharing of data between authorized users. Tran et al proposed such a framework for securely sharing data. This was based upon a proxy re-encryption process where a key is shared between both the user and proxy. If a user is removed from a group by the administrator they would be unable to reaccess that group. This model had two weaknesses: security could become vulnerable should the proxy and a user participate in collusion, and the potential load on the proxy for encryption and decryption of data (Tran, 2010). Wooten et al. developed a social cloud system for healthcare. Once again security is of great concern especially because of the sensitivity of personal health data. They used a trust-aware role based access control system. Trust ratings are calculated based upon the user's activities compared to their peers. Access is allowed based upon these trust ratings and the user's role (Wooten, 2012).

4. Architecture

The cloud architecture that is commonly used through social applications does not differ from the typical cloud architectures. PaaS is commonly used for social applications as a total solution for social app development. Social applications can be designed as applications on top of existing social networks or as separate applications. In the Social Cloud, the system consists of a Facebook application that is used to share the resources provided by the users. In this system, Facebook's built-in capabilities were leveraged with user management and authentication. The established relationships within the social network are used to map certain resources and services to particular users. For example, resource sharing can be done only with friends, or members of the same group. The application itself serves as the type of marketplace where the actual services or resources can be obtained. In their implementation, the resources were provided as storage as a service (Chard, 2010). In social cloud for healthcare, an existing social network was not used. Instead the entire social cloud application was hosted on the CSP Amazon's EC2. The major components here were the social cloud, the access control and the database. Users access the social cloud through the access control, and data is persisted in the backend database (Wooten, 2012).

In terms of architecture, once again close attention must be taken in regards to security of the cloud whether using either type of social application. Both between the user and the cloud service provider. Both entities need to be aware of the possible threats to security. These include both outside and inside threats. Since public clouds tend to have multiple users on the same system, it is important to ensure that there is proper authorization, authentication and access control to maintain each user's security on their cloud data and resources. This would require the cloud system to be secure from end to end, on the virtual environments, API calls, network communications and otherwise. These requirements are similar yet different from traditional information security due to the additional complexity and flexibility of the cloud. To ensure proper security, providers and users should also define policies and requirements before moving to a cloud solution especially due to the sensitive nature of personal data on the social network. This means that if a social application is sitting on top of an existing social network, special scrutiny should be given to the communication between the two (Behl, 2011). In one framework for a secure social network on the cloud, the CSP is transparent to the users. A proxy sits between the cloud provider and the end user. Data is encrypted with a key from the key manager before being sent to the proxy and then decrypted once it reaches the user with a corresponding key (Tran, 2011).

5. Existing Applications

There exist a number of social applications that are making use of cloud computing technologies. As previously discussed, these applications typically involve using the existing user management capabilities of the social network to use cloud resources much like the content that is already being shared by social networking users. Box.net is one such cloud storage provider. They have created a variety of apps aimed at sharing their stored data across numerous social networks. These include Twitter, Linked.in and Facebook. The application interfaces with social networks and posts links that allow users access to the stored data (Cassavoy, 2011). The flexibility of cloud services to scale up and down to meet the resource need fits well with the dynamic nature of the social network.

5.1 Facebook

Facebook is a social networking website that provides users a personal profile page where they can post messages, photos and other media. These materials can be shared with other users who they have 'friended.' Other features include: groups and friend lists. As of September 2012, Facebook has surpassed one billion active users who use their service. A cloud storage provider, Dropbox has introduced Facebook integration. Facebook allows storing and sharing files within groups. Now Dropbox has been integrated such that files from the CSP can be uploaded directly from Dropbox to Facebook (Taylor, 2012). Facebook has also partnered with Heroku, a

PaaS provider, for hosting Facebook applications using a variety of languages such as PHP, Ruby and Python. Their system is integrated within Facebook to provide a user friendly experience for novices to be introduced to application development on Facebook (Lee, 2012). Internally, Facebook hosts the largest in volume Hadoop cluster that consists of 4,400 nodes and over 100 PB of data (Menon, 2012).

5.2 Twitter

Twitter is a social networking service that provides users a personal page where they can post messages that are no longer than 140 characters called "tweets". Users are able to communicate with each other through adding an username prefixed with the "@" symbol. As of December 2012, Twitter announced they had over 200 million active monthly users. Twitter uses Hadoop clusters to do off-line batch processing of user relationship data to power their *People You May Know* feature (Ryaboy, 2012).

5.3 LinkedIn

LinkedIn is a social network geared towards professional networking. Users are provided with a profile page where they can maintain a list of connections with other users on the service. Other features include: resume posting and job postings. As of January 2013, LinkedIn had more than 200 million users on their network. LinkedIn's architecture is made of several components. For features such as People You May Know, Hadoop, Hive and Pig are used to batch process off-line data. Other features such as recommendation products and rate limiting are powered by the distributed data store Voldemort. LinkedIn has about 10 Voldemort clusters, across over 100 nodes (Auradkar, 2012).

5.4 YouTube

YouTube is a video sharing website where users can upload, view, share and comment on videos. Users are provided with a profile page that lists their videos and messages. Users are able to subscribe to other users to receive updates on their videos and comments. As many as 1 billion unique users visit YouTube in a month. YouTube makes use of a delivery cloud that is responsible for serving video content. YouTube uses two methods of load distribution across this cloud. Based upon the user's location, users are directed to video cache servers in close proximity. During peak hours, they may be directed to a farther cache if located in a heavy usage area. The second method is just a redirection to another user if the current server being used is busy. This delivery cloud has three components: video id space, video servers, and a physical server cache. The video id is a fixed length unique identifier for each video. The video server organization consists of several DNS namespaces representing a set of logical video servers. The physical server cache is a hierarchy of physical servers grouped into primary, secondary and tertiary locations (Adhikari, 2011).

5.5 Flickr

Flickr is a video and image hosting website that allows users to share and comment on photos. As of 2012, the site has hosted over 6 billion images. Flickr uses federation architecture for their user data such as favorites, where the database is distributed across servers as shards. These database slices, shards, are arranged in a master-master ring replication. Each shard holds roughly 400,000 members' data, with the entire database sized at over 12TB (Pattishall, 2008).

5.6 eHarmony

eHarmony is an online dating website that is aimed to matching couples thorough their common interests. eHarmony has a membership of over 30 million members. To support this matching feature, eHarmony processes event log files that are parsed with Hparser. These files are stored in a staging table that is roughly 30 billion rows at 1.2 TB worth of data as of 2012. At this point in the process, tools such as Hive are used to query the data for relationship discovery. The processed files are stored in a Hadoop Distributed File System for one year before being archived in Amazon's S3 service (Chiguluri, 2012).

6. Open Issues and Challenges

6.1 Privacy and Trust

Privacy has been a subject of great concern with social networks. The protection of a user's identity varies across the various social network services available across the internet. Some websites, such as Facebook, encourage the use of real names and thus make a connection between their social network and public identities. Others sites, such as dating services provide some weak anonymity by using only first names or a user-created name instead. Even though Facebook does not provide anonymity, it does provide options to restrict access only to those you allow access. Other than access by other users, there are questions on how these social networking services may be using the vast amounts of data that users are providing. Facebook's policy states that information may be

shared with third parties that does not identify or expose the user's identity. In this case, it may be marketing research companies who use the information to target advertisement to certain users. In terms of privacy, there are questions as to what is being removed from the data being shared that makes the users "unidentifiable". There are means to deduce identities based upon the social network graph topology, and distorted and removing data could affect the quality of data analysis and mining of the information that is being shared. These issues raise questions as to how these social network services handle their data to balance the needs of third party data consumers and the expectations of their users (Bianco, 2009).

6.2 Ownership of Content

The massive amounts of data that exist on social networking services are mostly user-generated. Different social media sites have different policies. For example, Facebook's policies state that it will use user's information in promotion or connection with its service. When dealing with items such as images, the content remains private if set as private by user preferences. However, Facebook does not have extensive copyright options or preferences much like the image sharing site Flickr. With Flickr, a user can set different policies through licenses: creative commons, no derivative works, etc. While users may be the owners of this data, license agreements based upon the use of the services' network may allow these sites to retain data even after users initiate removal or deletion (McCarthy, 2009).

6.3 Data Retention and Failures in the Cloud

Although cloud technologies present much value, there are several concerns about centralizing data and data control in the cloud. Should valuable data be placed in the cloud and lost, there is little that can be done to recover that data. This is not different from the traditional model where data is managed by the organization itself. However when that data is sent into the cloud, organizations relinquish some measure of control and that requires trust that the cloud service provider will manage the data properly. One example of such a failure occurred when social bookmarking service Ma.gnolia had system failures across primary and backup servers, effectively losing all user data in 2009 (Bianco, 2009).

7. Conclusion

We are currently living in the age of communication where millions of people are connected through the Internet. Several of these people maintain relationships online through social networking sites like Facebook, LinkedIn, Twitter and others. The rapid growth of these social networks has given rise to marketing and customer relationship opportunities for businesses, and large datasets for analytics. For the future of social networking there will be a continued focus on user privacy and data control. To meet these demands, social networking services will be inclined to adopt policies and data protection settings so that users will be able to manage their data, and access to it. Policies concerning lifespan of data will also need to be examined to clarify what happens should users terminate their account or die. This would also include transparency of how these services are using their data internally as well as providing to third parties and other organizations. As far as cloud computing for hosting social applications, it will remain as an attractive option especially with integration with social networks and their APIs. It provides an inexpensive solution that reduces the effort to create an application. Social media has also given rise to several applications of data analytics. This trend is expected to continue with the idea that organizations can derive useful information such as trends and user profiling. Social media has a strong tie with big data as these services produce and consume big data. Several of these services are driving current development of sophisticated big data architectures and technologies to power the features that they provide to their users. This can be related to cloud computing since these technologies typically make use of distributed computing resources are typically the cloud. As the social networks grow, there will be a growing need for increasing amounts of computing resources and cloud computing remains as a viable solution to meet those needs.

References

Adhikari, V. K., Jain, S., & Zhi-Li, Z. (2011). Where Do You "Tube"? Uncovering YouTube Server Selection Strategy. *Computer Communications and Networks (ICCCN), 2011 Proceedings of 20th International Conference on* (pp. 1-6). http://dx.doi.org/10.1109/ICCCN.2011.6006028

Auradkar, A., Botev, C., Das, S., De Maagd, D., Feinberg, A., Ganti, P., … Zhang, J. (2012, April). Data Infrastructure at LinkedIn. *Data Engineering (ICDE), 2012 IEEE 28th International Conference* (pp.

1370-1381). http://dx.doi.org/10.1109/ICDE.2012.147

Behl, A. (2011, December). Emerging security challenges in cloud computing: An insight to cloud security challenges and their mitigation. In *Information and Communication Technologies (WICT), 2011 World Congress on* (pp. 217-222). IEEE. http://dx.doi.org/10.1109/WICT.2011.6141247

Bianco, J. (2009). Social Networking and Cloud Computing: Precarious Affordances for the 'Prosumer'.

Cassavoy, L. (n.d.). *Box.net OpenBox Service: Facebook*. Retrieved March 22, 2011, from http://www.pcworld.com/article/233949/boxnet_openbox_service_facebook.html

Chard, K., Caton, S., Rana, O., & Bubendorfer, K. (2010, July). Social cloud: Cloud computing in social networks. In *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on* (pp. 99-106). IEEE. http://dx.doi.org/10.1109/CLOUD.2010.28

Chiguluri, S., & Parsamyan, G. (2012). *How eHarmony Turns Big Data into True Love*. Retrieved from https://community.informatica.com/mpresources/Communities/IW2012/Docs/bos_65.pdf

Falahi, K. A., Atif, Y., & Elnaffar, S. (2010, December). Social networks: Challenges and new opportunities. In *Proceedings of the 2010 IEEE/ACM Int'l Conference on Green Computing and Communications & Int'l Conference on Cyber, Physical and Social Computing* (pp. 804-808). IEEE Computer Society. http://dx.doi.org/10.1109/GreenCom-CPSCom.2010.14

Lee, C. (n.d.). *Facebook and Heroku: an even easier way to get started*. Retrieved September 2011, from http://developers.facebook.com/blog/post/558/

McCarthy, C. (2009). *Facebook faces furor over content rights*. Retrieved from http://www.cnn.com/2009/TECH/02/17/facebook.terms.service/

Menon, A. (2012). Big data @ facebook. In *Proceedings of the 2012 workshop on Management of big data systems* (MBDS '12) (pp. 31-32). New York, NY, USA: ACM. http://dx.doi.org/10.1145/2378356.2378364

Motta, G., Sfondrini, N., & Sacco, D. (2012, May). Cloud Computing: An Architectural and Technological Overview. In *Service Sciences (IJCSS), 2012 International Joint Conference on* (pp. 23-27). IEEE. http://dx.doi.org/10.1109/IJCSS.2012.37

Pattishall, D. (2008). *Federation at Flickr: Doing Billions of Queries per Day*. Retrieved from http://www.scribd.com/doc/2592098/DVPmysqlucFederation-at-Flickr-Doing-Billions-of-Queries-Per-Day

Ryaboy, D. (n.d.). *Twitter at the Hadoop Summit*. Retrieved June 2012, from http://engineering.twitter.com/2012/06/twitter-at-hadoop-summit.html

Taylor, C. (n.d.). *Facebook Integrates With Dropbox to Power File-Sharing Within Facebook Groups*. Retrieved September 26, 2012, from http://techcrunch.com/2012/09/26/facebook-integrates-with-dropbox-to-power-file-sharing-within-facebook-groups/

Ting, I. H., Lin, C. H., & Wang, C. S. (2011, July). Constructing A Cloud Computing Based Social Networks Data Warehousing and Analyzing System. In *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on* (pp. 735-740). IEEE. http://dx.doi.org/10.1109/10.1109/ASONAM.2011.25

Tran, D. H., Nguyen, H. L., Zha, W., & Ng, W. K. (2011, December). Towards security in sharing data on cloud-based social networks. In *Information, Communications and Signal Processing (ICICS) 2011 8th International Conference on* (pp. 1-5). IEEE. http://dx.doi.org/10.1109/ICICS.2011.6173582

Vakali, A., Giatsoglou, M., & Antaris, S. (2012, April). Social networking trends and dynamics detection via a cloud-based framework design. In *Proceedings of the 21st international conference companion on World Wide Web* (pp. 1213-1220). ACM. http://dx.doi.org/10.1145/2187980.2188263

Wooten, R., Klink, R., Sinek, F., Bai, Y., & Sharma, M. (2012, May). Design and Implementation of a Secure Healthcare Social Cloud System. In *Cluster, Cloud and Grid Computing (CCGrid), 2012 12th IEEE/ACM International Symposium on* (pp. 805-810). IEEE. http://dx.doi.org/10.1109/CCGrid.2012.131

Copyrights