International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Relevance of Big Data Technologies to and users in Context of Social Networking Applications

Pawan Saxena, Rajiv Pandey

Abstract*: Unremitting generation of data by various data analytics platforms, ubiquitous ,edge nodes and social networks in the concurrent scenario has shaped the exceptional amount of data in volume, velocity, veracity, variety and value. Exceptional data have made traditional information technology and method unfeasible to cope up amid. This exceptional data has been termed as Big Data. Social media is one of the most important sources of Big Data. social media is a constituent of Big Data. Besides Big Data plays a vital role in moving forward the Social Networking Applications to innovate and enhance the experience of users. Various technologies are factored for Big Data storage, processing and analysis in the context of social networking. This paper investigates these technologies which are being used by social networking applications with their relevance to the end users. The research article provides a relevance computation of various social media platforms. It further summarizes a visualization of the use of the platforms in their contribution to the big data.

Keywords: Big Data, Social Networking Applications, Hadoop, Hive, Corona, Presto, Cassandra, Technologies of Big Data, KPA of Big Data.

I. INTRODUCTION

Big Data is an axis point for countless researchers in all spheres of expertise. Developed countries like USA and UK are putting in millions of dollars in the research related to Big Data as well as rapid growing economies like India is also catching up by funding various government departments in various fields like biotechnology using Big Data and machine learning.

Big Data [24,25] needs to be stored and analyzed for the gain of meaningful information. This had escorted to the development of various technologies related to Big Data. These Big Data technologies are Apache Hadoop, Apache Hive, HBase, and Cassandra.

Human beings around the world need to communicate and share their views and thoughts with each other and this desire had been realized by social networking applications in the modern era. These applications tend to generate enormous data of varying types. The data may be structured, semi-structured or unstructured. This data is stored and processed by the Big Data Technologies in using different methodologies.

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Pawan Saxena*, AIIT, Amity University, Lucknow, India paxena1@amity.edu
Rajiv Pandey, AIIT, Amity University, Lucknow, India rpandey@lko.amity.edu

This paper analyses Big Data Technologies with focus on the use of these technologies by the end users with respect to social networking applications and their features. Facebook, the most popular social networking application, and its Big Data technologies are the primary focus of this paper. Certain questions with points-based answers are asked to end users with respect to features using the Big Data Technologies. This data set is used to make broad picture for technologies relevant to them.

This paper had been divided into six sections. Section I being an introduction to the subject of Big Data, the second section highlights the related work done in the context of this paper and its technologies. Third section explains the concept of social networking application with a focus on major market leaders like Facebook, Google plus, Twitter and LinkedIn. The fourth section discusses the method and technique used for finding out relevance of Big Data Technologies. It is followed by fifth section which analyses the result. The sixth section makes the conclusion and provides the future scope.

II. RELATED WORK

International Data Corporation (IDC), a reputed market intelligence organization defines Big Data technologies as key transformable technologies for extracting valuable data from huge volumes with range of varieties at intense velocity of data association and its associated analysis. These technologies had been centered on three key pillars which are data itself, analysis part and final presentation of analysis. Technologies related Big Data are focus on either one pillar or combination of pillars. Research had been on various aspects like computational model and storage system.

The termed Big Data was coined by Gartner Inc. in 2007. They defined Big Data as data which beyond the capability of the traditional system to storage and process using different Vs (Velocity, Volume, Value). Since then there are many Big Data technologies which had been developed to exploit Big Data. Some of these technologies are as discussed in below points.

A. Apache Hadoop

The Apache Hadoop is a framework of Big Data which have facilities including storage, processing and analysis. It is inspired by Google MapReduce and Distributed file system. It has many subcomponents like HDFS, MapReduce and YARN. HDFS is the distributed storage system for data and MapReduce is the computational programming model.

B. Apache Hive

The Apache Hive is data warehouse software for Big Data. It was developed by Facebook and now available under Apache License. It works with the Hadoop framework. It uses SQL query statement for operation in the Hadoop based environment.

C. HBase

The Apache HBase is a database used by Big Data. It is inspired by Google BigTable. It allows exploiting the function of NoSQL with features of Hadoop. It uses key/value pairs for data storage in HDFS of Hadoop.

III. SOCIAL NETWORKING APPLICATIONS: FACEBOOK, GOOGLE+, TWITTER, LINKEDIN

Social networking had become integral part of modern society. There are social networks with a number of users greater than population of most countries. The unparalleled potential of the web through which individual interact with one another to share information, ideas, personal messages, professional information, up-to-the minute thoughts and other content such as images, videos is known as social networking. Some of the popular social networking sites are Facebook, Twitter, LinkedIn and Google+. With an internet connection, anyone can interact with millions of people online through social networking. The primary focus of social networking is on building and reflecting of social relations among people.

Social networking is a data-driven application and is the prime source of generating Big Data. The data derived from these sites are in various ways. By storing, processing and analyzing data, these sites offer various products to the billions of active users. These functions use various technologies.

The technologies used in the social networking for storing, processing and analyzing Big Data are discussed in below sections.

Corona

Previously, Facebook had been using MapReduce to manage its massive data, but MapReduce failed to handle Facebook's growing needs. Later they developed a new data framework called Corona[31]. The concept of a cluster manager had been introduced to manage all the resources. Each job was tracked by an individual job tracker.

Presto

As data are growing enormously to the point of petabytes, querying the database is a major concern. We shall be able to run more interactive queries and get the result faster.

Haystack

It is high performance photo storage/retrieval system. It is a highly scalable storage used to serve Facebook's huge amount of photos. Haystack stores photo data inside 10 GB bucket with 1 MB of Meta data for every GB stored.

These technologies support the features of social networking sites delivered to the end users.

Social Networking Applications

Social networking portal [19] had been providing a virtual platform for the user to explorer themselves in new techniques. It is strong support by Big Data technologies. Some of these most social networking portals are Facebook, Google plus, twitter and LinkedIn. A brief description of these portals is mentioned as below :

Facebook

Today Facebook [1] is the most popular social media with more than 800 million active users. Some key features of Facebook are Photos, News Feed, Like Button, Messenger, Time Line and Events.

Google+

Google Inc.'s new product in the social networking world when Facebook gained much more popularity compared to Google's ORKUT. It had been launched in 2011 and became hugely popular with 40 million users within one month of being launched. Users can add fellow users in their circles and interact with them. Google+ has a "+1 button" with the same functionality as Facebook's "Like" button. Users can add pictures and edit them on the website itself. Key features of Google+ are "Circles", "Real-Time Stream", "Hangouts", "Photo Uploads", "Sharing and Tagging".

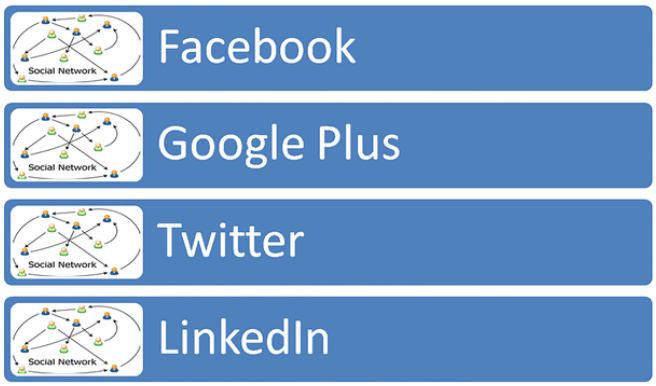


Figure 1: Diagram showing the popular social networking applications (Designed by author)

Twitter

Twitter is a social media to share what you think, what you view or what is your opinion about the things happening in people's life and in the world. Text-based posts had been used to express your feelings and thoughts. Users can follow other people based on their interests. Today 200 million people are connected using Twitter. Popular features of Twitter include Timelines, Following & Followers, Conversations, ReTweets and Hashtags.

LinkedIn

LinkedIn is a social media for professionals of various fields to interact with each other. It is business oriented and a complete career management, social networking platform. It has more than 400 million members active today. Key features of LinkedIn are Posting Updates, News Signals, Tags, Personalizing Messages, Groups and Multi-media.

IV. METHODOLOGY USED FOR RELEVANCE COMPUTATION

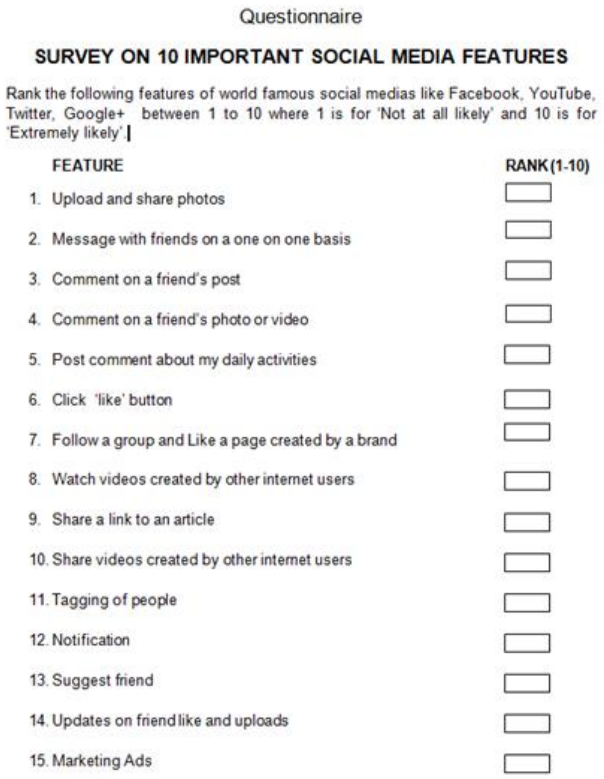
The methodology used in this paper is the questionnaire as a tool for determining the relevance of social media technologies to end users.

The users of the social media were selected randomly and asked to give rank with respect to questions in the questionnaire.

A questionnaire consisting of 15 questions with rank based answers had been developed for this research purpose. This questionnaire as shown in Fig. 1 relates to some of the prominent features of the social networking sites under our study. A survey was conducted through this questionnaire amongst active users of these social media.

The users who use social networking applications were required to answer the questionnaire. These users are 200 students approximately belonging to random Indian university. The users are students who are studying courses like B.Tech. M.Tech., B.B.A., M.B.A., B.Law and B. Architecture. They are in age between 19 to 27 years. These students devote from 10 minutes to 6 hours on social networking applications in a single day. The mobile phones are most favored platform and few users prefer other platforms like Laptops, PCs and Tablets connect to social media sites.

The users were suggested to give their response by ranking these features between numbers 1 to 10 where 1 is for 'Not at all likely' and 10 is for 'Extremely likely'. The sample format of the survey questionnaire had been shown in Fig. 2.



Questionnaire

SURVEY ON 10 IMPORTANT SOCIAL MEDIA FEATURES

Rank the following features of world famous social medias like Facebook, YouTube, Twitter, Google+ between 1 to 10 where 1 is for 'Not at all likely' and 10 is for 'Extremely likely'.

FEATURE	RANK (1-10)
1. Upload and share photos	
2. Message with friends on a one on one basis	
3. Comment on a friend's post	
4. Comment on a friend's photo or video	
5. Post comment about my daily activities	
6. Click 'like' button	
7. Follow a group and Like a page created by a brand	
8. Watch videos created by other internet users	
9. Share a link to an article	
10. Share videos created by other internet users	
11. Tagging of people	
12. Notification	
13. Suggest friend	
14. Updates on friend like and uploads	
15. Marketing Ads	

Figure 2 Form of survey questionnaire which used for observation (Designed by author)

The responses of the users were examined, and the following observations were drawn: -

Table 1 Average rank of survey result

#	Question	Rank
Q.1	Upload and share photos	84.2
Q.2	Message with friends on a one on one basis	67.3
Q.3	Comment on a friend's post	68.4
Q.4	Comment on a friend's photo or video	63
Q.5	Post comment about my daily activities	62.2
Q.6	Click 'like' button	66
Q.7	Follow a group and Like a page created by a brand	46.9
Q.8	Watch videos created by other internet users	51.2
Q.9	Share a link to an article	46.7
Q.10	Share videos created by other internet users	40.7
Q.11	Tagging of people	65.8
Q.12	Notification	70.4
Q.13	Suggest friend	49.3
Q.14	Updates on friends like and uploads	55
Q.15	Marketing Ads	58.5

The normalized rank ranges from 40.7 to 84.2 and after examination provides insight with a wide spectrum of result set. The values of the set are 84.2, 67.3, 68.4, 63, 62.2, 66, 46.9, 51.2, 46.7, 40.7, 65.8, 70.4, 49.3, 55 and 58.5. The median of the rank is 62.2.

On sorting the data set smallest to largest order we get data set X, then

X = {40.70, 46.70, 46.90, 49.30, 51.20, 55.00, 58.50, 62.20, 63.00, 65.80, 66.00, 67.30, 68.40, 70.40, 84.20}

Formula for calculating median is

Median element position = {(count + 1) ÷ 2}, where count is the total elements

Total count of elements in X is 15.

Let, median be M, then

M = {(15+1) ÷ 2}

M = {16 ÷ 2}

M = 8

The median of the rank is 62.2 as it is at eighth position in data set X. So, "Post comment about my daily activities" is the common relevant technologies for most of the users. In the next section, further analysis of the obtained data was conferred.

V. RESULT AND ANALYSIS

The result of questionnaire shows varying results which could be used different analysis and mapping of patterns. Like button feature of Facebook uses Hadoop, which was used for determining the orientation, satisfaction, intelligence, emotional stability and many other parameters related to the end users.

On analyzing the result, the technologies which offer the end users comments are more relevant while the technologies that give the data or options to end users are less popular. For example, "like" and "upload and share" features take the inputs from the end users hence a have high-ranking score. But, share video from other, has a low rank as it takes minimum inputs from the end users. The most users had issue of security while sharing the video as many of the links shown as video have resulted in malware. They were also issuing of the size of the video as it consumes a lot of bandwidth.

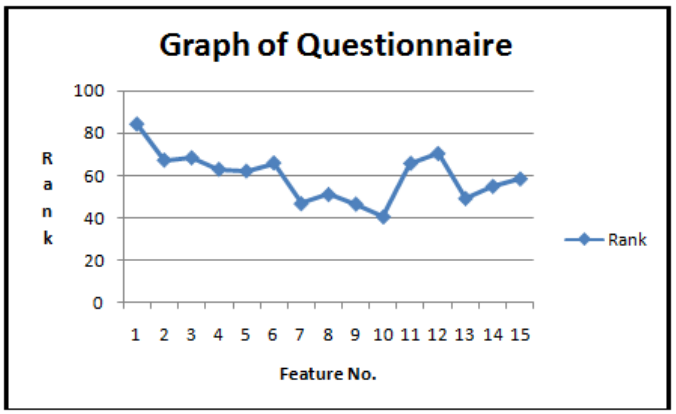


Figure 3: Graphical visualization of rank with respect to questionnaire (Designed by author)

Fig. 3 is showing a line chart of the result of the survey conducted through the questionnaire. The graph depicts rank on the Y axis and the feature number on the X axis. The graph clearly shows that the comments taking features has higher ranks reinforcing the earlier analysis.

VI. CONCLUSION AND FUTURE SCOPE

The paper explores the significance of the Big Data technologies with special focus on the social networking environment. The images and videos sharing technologies are most relevant to the user. They are uploading millions of images in friction of the seconds which was processed by technologies like Hadoop, Corona, Presto, Haystack, Memchache and Cassandra. The effective photo uploading system is vital to the success of any social networking application apart from text message sharing. The security aspect of video post by another user remains an issue on which some preliminary work had been done. The video streaming by popular portal like YouTube has been a success story but similar story is missing in social networking application.

These discussed results helped in establishing the base for future research and development with market driven technology development in the field of Big Data. In future, more research could be done on the technologies found to be critical to the end user as well as the reason for the technologies with low rank can be identified in order to improve the end user experience as well as technology up-gradation.

ACKNOWLEDGEMENT

We are grateful to all those students which participated in the survey. They have contributed their valuable time and effort in answering the questionnaire.

REFERENCES

1. 10 Features That Made Facebook The Most Used Social Media Site, Retrieved April 20, 2016 from https://www.vcpost.com.
2. Agnivesh and Dr. Rajiv Pandey. "Elective Recommendation Support through K Means Clustering Using R Tool" Proc. of IEEE International Conference on Computational Intelligence and Communication Networks (CICN - 2015). Web<http://www.cicn.in/Proceedings.html>. ISBN 978-1-5090-0076-0
3. Agnivesh and Dr. Rajiv Pandey. "Shiny Based Elective Recommendation Web App through K Means Clustering" Proc. of IEEE International Conference on Communication Systems and Network Technologies (CSNT-2016). Web<http://www.csnt.in/Proceedings.html>. ISBN 978-1-4673-9950-0
4. Agrawal, D., Das, S., & El Abbadi, A. (2011, March). Big data and cloud computing: current state and future opportunities. In Proceedings of the 14th International Conference on Extending Database Technology (pp. 530-533). ACM.
5. Apache Thrift TM, Retrieved April 16, 2016 from https://thrift.apache.org
6. Archenaa, J., & Anita, E. M. (2015). A survey of big data analytics in healthcare and government. Procedia Computer Science, 50, 408-413.
7. Beyer, Mark. "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data" Retrieved April 14, 2016 from http://www.gartner.com/newsroom/id/1731916.
8. Bradburn, Norman M., et al. Improving interview method and questionnaire de-sign: Response effects to threatening questions in survey research. University Microfilms, 1992.
9. Chen, C. P., & Zhang, C. Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences, 275, 314-347.
10. Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 36(4).
11. ChitreshVerma and Dr. Rajiv Pandey."An Implementation Approach of Big Data Computation by Mapping Java Classes to MapReduce."IndiaCom 2016. Proc. of IEEE INDIACom - 2016: Computing For Sustainable Global Development, India, at New Delhi. IEEE Delhi Section.Web. 30 Mar. 2016. <http://bvicam.ac.in/news/INDIACom 2016 Proceedings/Main/index.html>. ISSN 0973-7529; ISBN 978-93-80544-20-5
12. ChitreshVerma and Dr. Rajiv Pandey. "Big Data Representation for Grade Anal-ysis Through Hadoop Framework." Proc. of IEEE Confluence-2016 - Cloud Sys-tem and Big Data Engineering. ISBN: 978-1-4673-8202-1
13. ChitreshVerma and Dr. Rajiv Pandey. "Comparative Analysis of GFS and HDFS:Technology and Architectural Landscape." Proc. of IEEE International Conference on Communication Systems and Network Technologies (CSNT-2016).Web.<http://www.csnt.in/Proceedings.html>. ISBN 978-1-4673-9950-0
14. Danah M. Boyd, Nicole B. Ellison, Social Network Sites: Definition, History and Scholarship.
15. Emani, C. K., Cullot, N., & Nicolle, C. (2015). Understandable big data: A survey. Computer science review, 17, 70-81.
16. Facebook Technology Stack, Retrieved April 19, 2016 from https://www.facebook.com/press/info.php?statistics.
17. Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A. Y., ... & Bouras, A. (2014). A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE transactions on emerging topics in computing, 2(3), 267-279.
18. Fang, H., Zhang, Z., Wang, C. J., Daneshmand, M., Wang, C., & Wang, H. (2015). A survey of big data research. IEEE network, 29(5), 6-9.
19. Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144.
20. HaewoonKwak, Changhyun Lee, Hosung Park, and Sue Moon, What is Twitter, a Social Network or a News Media?, Proc. Of International World Wide Web Conference Committee(IW3C2), WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA., ACM 978-1-60558-799-8/10/04.
21. Here's How Facebook Manages Big Data. The CIO Report RSS. Web. 28 Apr. 2016. <http://blogs.wsj.com/cio/2013/10/31/heres-how-facebook-manages-big-data/>.
22. Martin Traverso, Presto: Interacting with petabytes of data at Facebook, Retrieved April 14, 2016 from https://www.facebook.com
23. Person, and AvantikaMonnappa. "How Facebook Is Using Big Data: Good, Bad, and Ugly | Simplilearn." Simplilearn.com. Web. 28 Apr. 2016. <http://www.simplilearn.com/how-facebook-is-using-big-data-article>.

24. Priyanka, K., & Kulennavar, N. (2014). A survey on big data analytics in health care. International Journal of Computer Science and Information Technologies, 5(4), 5865-8.
25. Sagiroglu, Seref, and DuyguSinanc. "Big data: A review." Collaboration Tech-nologies and Systems (CTS), 2013 International Conference on.IEEE, 2013.
26. Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., & Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-art. Journal of Network and Computer Applications, 71, 151-166.
27. Singh, D., & Reddy, C. K. (2015). A survey on platforms for big data analytics. Journal of Big Data, 2(1), 8.
28. Top Ten Google + Features, Retrieved april 17, 2016 from http://www.dummies.com
29. Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications (Vol. 8). Cambridge university press.
30. Welcome to Apache™ Hadoop®! (n.d.)., Retrieved April 14, 2016 from https://hadoop.apache.org/
31. Why Facebook ditched Hadoop'sMapReduce and built a better mousetrap called Corona to handle its data, Retrieved April 15, 2016 from http://thenextweb.com/facebook/2012/11/08/facebook-engineering-team-builds-corona-for-mapreduce-jobs/#gref.
32. Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong.Analysis of topological characteristics of huge online social networking services.In Proc. of the 16th in-ternational conference on World Wide Web.ACM, 2007.

AUTHOR PROFILE

Mr. Pawan Saxena, is working as Assistant Professor in Amity Institute of Information Technology, Amity University Uttar Pradesh, Lucknow Campus, He is M.Tech. (IT). The author is Pursuing Ph.D. in Information Technology from Amity Institute of Information Technology, Amity University, Lucknow Campus, Uttar Pradesh. His research areas include Big data Analytics, Cloud Computing, IoT and Machine Learning. His papers are published in IEEE, International Conferences and Scopus Indexed journals.

Dr. Rajiv Pandey, Senior Member IEEE is a Faculty at Amity Institute of Information Technology, Amity University, Uttar Pradesh, Lucknow Campus, India. He possesses a diverse background experience of around 30 years to include 15 years of Industry and 15 years of academic. His research interests include the contemporary technologies as Semantic Web Provenance, Cloud computing, Big-Data, and Data Analytics. He has been on technical Committees of Various Government and Private Universities. He is intellectually involved in supervising Doctorate Research Scholars and Post graduate Students. He is also an active contributor in professional bodies like IEEE, IET and LMA. He is a member of Machine Intelligence Labs.