

Impact of Social Media on Big Data and Green Computing and Solution

Ashwini M Chalawadi

Department of Computer Science and Engineering, VTU Belagavi, Karnataka, India.

mca.ashwini@gmail.com

Abstract- Big data's one of the major challenge is volume of data. With high speed 4G internet and with the occurrence of smart hand held devices like Smart cell phones, iPads, Notebook computers, increased the number of internet users, that too online social user. This has led to online social data generated in tremendous rate never seen before. This huge data generated encountered various challenges like storage, maintenance and green computing for software industries. So, in this paper I proposed an algorithm which provides a form of optimal solution to control unwanted data generation from social media interaction and also discussed its impact on storage and on Green Computing and respective solution.

KeyWords: Big data, Social media, Green Computing, Data Center, PUE.

I. INTRODUCTION

The term Big Data deals with data generated across the world in an unexpected rate never occurred before. Big Data is a branch of computer /information science and is an interpretation of collection, organization and analysis of large data sets. This is done mainly to discover information patterns and other required information sets. Big Data analysis is well known for helping companies and organizations to stay ahead of competitors.

The data generated from various social forums is

complex in nature. It's a combination of homogeneous and heterogeneous content. Organizing and storing such a huge complex data it is a tedious task and so minimization of data leads to less maintenance, less burden with reference to storage.

Every day, large amount of data being generated from various sources like cell phones, twitter, instagram and other social forums leads to storage and maintenance problem, so it is time to consider this issue seriously, to minimize unwanted data generation and storage.

Big Data can be defined in terms of V's as below.

1. Volume – Volume in Big Data implies the amount of data produced and stored. It is estimated that on an average, 2.3 trillion gigabytes of data is produced from various sources every single day. Analysis of such large quantity of data is time consuming and impractical. Therefore, Volume is one feature which needs to essentially considered while dealing with the term 'Big Data'.

2. Variety- The term Variety deals with nature of data. The data produced from various sources, especially online users may be of any form like DOC files, PDF's, TXT files, ZIP files, JPEG, PNG,MP3,MP4 XLSX,EXE and more. Obviously the data is sometimes structured, semi structured and unstructured in nature. We need advanced tools to process these varieties of data to deal with Big Data.

3. Velocity- The term Velocity refers to speed at which data flows from different sources. The main sources of data generation are Networks, SocialMedia Websites, Business Processing Units, Application

logs, Road/Air/Marine-Traffic sensors, Cyber Security sensors, Army Security sensors, Weather forecasting sensors, Hand held devices like smart phones, iPads etc. To deal with 'Velocity' the data generation speed must match with data processing speed.

4. Veracity - Biases, noise and abnormality in data is known as Veracity. Is one of the major challenges of Big data. To deal with this problem we require advanced filtering techniques and tools to eradicate biases, noise and abnormality of data.

Technologically, Big Data made lot of difference in human lives and also help us to make standard decisions. The following are the application of big data.

Areas of Applications:

- Health sectors.
- Public opinion Policy making and organizations.
- Smart cities development organization for countries.
- E-education
- Artificial Intelligence, Robotics and Machine Learning.
- Agriculture Sciences.
- Internet of Things.
- Advertizing Analysis (eg: google advertizing)
- Predictive Analytics (Predicting future trends)
- Customer Churn Analysis
- Customer purchasing trends etc.

After analyzing different sources of data generation in concerning with velocity, volume, variety, veracity. The following are the interesting facts observed.

Volume

The data generated by Facebook and Twitter together is 500 times greater than the data generated by New York Stock Exchange (NYSE). 90% of people out of 100 %(world population) have cell phones. It is

estimated that 43trillion Giga bytes of data will be generated by 2024.

Variety

Facebook content exchange every month is 300 billion and tweeter content generated every month is 400 million.

Velocity

High Speed internet facility had made users to transform their data with tremendous speed across different devices. According to statistics NYSE captures 1 Terabytes of Trading information during each session.

Big data concerned with diverse, complex, large sets of information that enlarged at ever-increasing rate. Big data often produced from numerous sources and arrives in numerous formats. It is comprised with Volume, Velocity, Variety, and Veracity.

Social media is an online group of people across the world for communication, especially to share formal/informal information, Knowledge, Messages through social websites and device applications.

Green computing means eco friendly use of computers and their resources. Also we can define it as a study of Manufacturing/Building/Using Engineering devices in such a way that, they consume less environmental resources and also produce less bad impact on environmental factors like air, heat production, electrical usage etc.

Data center is a large group of networked computer servers typically used by organizations for the remote storage, processing, or distribution of large amounts of data.

PUE (Power Usage Effectiveness) means ratio of amount of energy required to carry out data center, to the energy required to run the server in it. PUE represents efficiency of Data centers.

The presented research work classified as follows, Section I represents Introduction of actual work, Section II represents related work carried out,

Section III contributes possible outcome of the work and section IV concludes the paper.

II. RELATED WORK

The following algorithm decides the significance of social media chats, posts. In order to minimize the amount of storage.

ALGORITHM

```
Begin,  
message_send()  
mark imp, not_imp  
Social_Media _account X to Social_Media _ account  
Y(sender to recipient)  
if valid  
    send()  
    if (imp)  
        Store until user delete it manually.  
    else(not_imp)  
        store till 8.64e+7 milliseconds.(one day) and  
delete permanently.  
else  
invalid recipient can't send message  
end
```

The flow of Algorithm

When users of social media begin to send a message, they create a text message, audio, video, images files, when they create a message they call send message method, Before sending, it asks two options important and not important. Once a send message option is selected it's also checks the validity of users, whether they are authorized user or not, if users are authorized then only users can send message otherwise they will be sent with message of invalidity of account. If valid, message send successfully with option, imp and not imp, If it is a important then it will be remain in storage of organization /company permanently until user delete it manually , if not_imp then message get deleted after 8.64e+7 milliseconds. Here another prime aspect is checking of authorization of users. Whether the users are authorized, created with valid e-mail id, address etc. So that it reduces frequency of fake user accounts, and unwanted data storage, that leads storage with electricity. Usually user messages like birthday wishes, morning and night wishes with

www.asianssr.org

videos , images and plain texts will remain in database long time without having much significance , maintaining these form of messages really burden, time and energy consuming , so this kind of messages depending on user preference and scenario they can be marked as important and not important. This approach certainly addresses the problem of volume of big data.

III. RESULT

Impact of Social Media Chats, Posts, Tweets on Energy Computing

Due to billions of users of social media and their minute to minute chats and day to day chats, online discussion forums, leads to high volume in data , which leads to establishment of more databases ,storage devices and servers and intern which again leads to building of large data centers ,their maintenance issues on both companies/organizations Because companies/organizations has to pay lot of electric expenses to maintain data centers ,also to implement privacy and security and it should be available all the time to the user. To achieve this, all these before said objective companies/organizations basically spending lot of electric energy and time which is against green computing.

Data centers are known as huge power consumption. So the above said algorithm provides optimal solution in order to reduce load of volume of big data and also supports "Green Computing". Means when there is a lesser data, which might cause establishment of fewer data centers. The efficient energy savings may lead to usage of energy for other significant work.

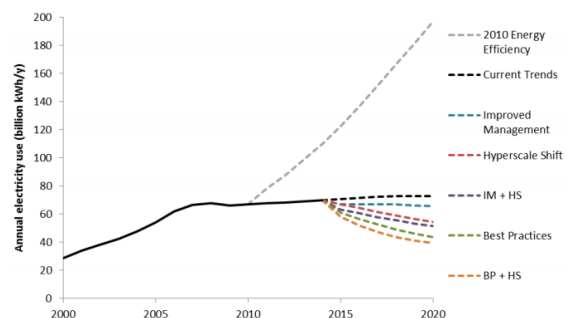


Fig. 1: Electric usage by datacenters across the world.

On demand establishment of data center leads to consumption of more electrical energy. The above graph shows the year wise electrical consumption by datacenters across the globe, consumption rate is gradually increasing with increase in data centers, so the data center and electrical energy consumption are directly proportional to each other.

According to world survey on impact of Information Technology on Green Computing is stated that, more than 50 billion devices to be connected by 2020. Again statistics says that more than 150 billion information technology devices further added by next 5 years i.e. by 2025. . Due to this increase in size and number of data centers, data centers are the biggest energy consumers.

After referring various research papers, Data centers using 20% of all available electricity in the world by 2025.

Due to social media chats, posts and information exchange across the world using various device platforms has led to production of large volume of data and to process these, we need data centers. Data centers are using high amount of electric energy for storing, processing, cooling. Every year 3% of generated electricity will be used by data centers.

With the view to implement Green Computing, industries must implement less PUE ratio. Also ICT industries embraced with advanced power management solution techniques such as smart UPS, smart battery monitoring equipment and intelligent racks. Along with these techniques the social media forums also implement technique proposed in this paper, so that most effectively, big data problems can be addressed.

IV. CONCLUSION

In 4v's of big data challenges, volume plays equally major role. So Social Media's like Whatsapp, Messengers, Facebook, Twitter etc consumes major storage space on public, hybrid and private cloud. So above algorithm focuses on the data storage and its validity. In this algorithm data is significant or not it

is decided by end user, So it reduces burden on storage space(which involves unwanted, unimportant, noisy data) and will be automatically decided by end users. It reduces maximum cloud storage issues leads to minimum storage. Also when storage is minimized, directly it affects the minimization of number of data centers. So, obviously it positively impact on Green Computing. In this paper, addressed volume challenge to some extent. This leads to serious users that efficiently use these social platforms for personal, professional and social concern. In future, the issues can be addressed on impact of end users if this mechanism is implemented.

REFERENCES

- [1] Avinash Aslekar and Pramod Damle "Improving Efficiency of Data Centres in India: A Review" Indian Journal of Science and Technology, Vol 8(S4), 44-49, February 2015 , ISSN (Online) : 0974-5645 ,ISSN (Print) : 0974-6846 .
- [2] Miyuru Dayarathna, YonggangWen, Senior Member, IEEE, and Rui Fan "Data Center Energy Consumption Modeling: A Survey " IEEE Communications Surveys & Tutorials, Vol. 18, No. 1, First Quarter 2016.
- [3] Maria Avgerinou, Paolo Bertoldi and Luca Castellazzi "Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency. 2016
- [4] Paolo Bertoldi, Maria Avgerinou, Luca Castellazzi "Trends in data centre energy consumption under the European Code of Conduct for Data Centre Energy Efficiency" research gate publication 2017.
- [5] Hong Shu "Big data analytics and Six Techniques" Geo-spatial Information Science ISSN: 1009-5020 (Print) 1993-5153 (Online) Journal.
- [6] C.L. Philip Chen , Chun-Yang Zhang "Data-intensive applications, challenges, techniques and technologies: A survey on Big Data" International Journal of Information Sciences Elsevier publication Information Sciences 275 (2014) 314–347.

[7] Nada Elgendy , Ahmed Elragal “Big Data Analytics: A Literature Review Paper” Conference Paper in Lecture Notes in Computer Science · August 2014.

[8] S. Justin Samuel etl “A Survey On Big Data And Its Research Challenges” ARPJ Journal of Engineering and Applied Sciences , Vol. 10, No. 8, May 2015 ISSN 1819-6608.

[9] Dr. S.Vijayarani and Ms. S.Sharmila “Research In Big Data – An Overview” Informatics Engineering, an International Journal (IEIJ), Vol.4, No.3, September 2016.

[10] Althaf Rahaman.Sk, Sai Rajesh.K, .Girija Rani K “Challenging tools on Research Issues in Big Data Analytics” International journal of IJEDR Volume 6, Issue 1 | ISSN: 2321-9939, 2018.

[11] Samitha Sahu P , Dr.V.IlangoP Dr.R.Chinnayan P, “Research Issues and Challenges of Big Data –A Review” International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-5, May 2016 ISSN: 2395-3470.

[12] D. P. Acharjya , Kauser Ahmed P “ A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools”, International Journal of Advanced Computer Science and Applications, Vol. 7, No. 2, 2016.

[13] Benjamin W Wah ,Yuanzhuo Wang et al , “Significance and Challenges of Big Data Research” Elsevier Publication Nov 2017.

[14] T.Veeramani, P. Srinuvasarao,B. Rama Krishna, R. Thilagavathy, “Impact of Social Media Networks Big Data Analysis for High-Level Business” , International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7 Issue-5S2, January 2019.

[15] Prashant Sahatiya, “Big Data Analytics on Social Media Data: A Literature Review”, International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 05 Issue: 02 | Feb-2018.