



CLOUD BASED MINING: A NEW APPROACH FOR SUGGESTIONS

GURPREET SINGH, SUNIL KUMAR CHAWLA, SUMIT KUMAR
and SHIPRA THAKUR

Department of Computer Science and Engineering
CGC-College of Engineering
Landran, Mohali, Punjab, India
E-mail: gurpreet.3529@cgc.edu.in
sunil.3550@cgc.edu.in
sumit.coeapplied@cgc.edu.in
shipra.4390@cgc.edu.in

Abstract

The paper is about how data mining can be used in cloud computing so that large area of customers can be targeted and hence improving way of suggestions. Data mining till date is a concept used to analyze the behavior of customers during their visit to shopping sites in order to provide them valuable suggestions to improve their online shopping experience. Combining data mining with daily activities is becoming a normal behavior. Mixing data mining with cloud computing leads the users to capture required and meaningful information from a large, virtual and integrated data warehouse thereby decreasing the overall cost.

1. Introduction

In today's world, Internet becomes a vital tool which helps us in our day to day activities, both official and personal, as the no. of users are increasing daily. It should not be taken as a surprise that the world wide web has taken over the normal routine of business and give a new approach to reach customers globally. Although the cloud computing is one of the main trends of marketing in the past few years. The Cloud, as it is always known by, also provides its services (IAAS, PAAS, and SAAS) [2] over the Internet. Most organizations in today's era choose cloud as a preferred option in place of

2010 Mathematics Subject Classification: 62M45, 68P15.

Keywords: Cloud Computing, Data Mining, DM Cloud, IaaS, PaaS, SaaS.

Received October 11, 2019; Accepted November 8, 2019

spending money on creating their own infrastructure in order to host storage space for database (i.e. hardware) or software and hence allowed themselves to have access to their resources over the Internet. Cloud Computing is getting popular due to its properties of allowing access to data during motion, availability every time everywhere (24X7) and cheapest alternative. Just like all other things cloud computing also have many limitations as there are various security concerns related to data stored on third party servers which may pose a threat to the company's assets. In the same age, data mining techniques also have grown up and became more useful in discovering knowledge in various fields of databases such as: business, medicine, science and engineering, spatial data etc. The rising Cloud Computing trends provides exclusive benefits for its users an unmatched access to valuable data that can be converted into valuable imminent that will help them to acquire their professional objectives. Data mining combined with cloud computing will make this picture more clear in terms of business for various companies or organizations. This paper will throw light on various technologies used in cloud and data mining and finally how these two terms when comes in combination will lead to profitable approach for various companies or organizations.

2. Data Mining

The term refers to mining the data to obtain the information which is hidden inside the large databases which makes it a powerful technology with more capability to help organizations to concentrate on the most important information in their data warehouses [1]. Tools used for mining the data in order to make predictions of various futuristic trends and behaviours, there by allowing various businesses to make bold decisions based on extracted knowledge. The analyses done automatically by various software tools offered by data mining make the analyses go ahead of past events provided by retrospective tools typical of decision support systems. As with time the amount of data increasing direct hand on analyses has been replaced by automated indirect data processing techniques. To uncover patterns which are hidden, the process of data mining has been applied on large data sets [1]. The technique of mining the data is used to order through data to identify patterns and making relations. Various parameters included in Data mining

are:

1. Association – Association Looks for patterns in data where one incident is in relation to another incident.
2. Analysis of ordering or paths -It looks for patterns where one event initiates another event.
3. Classification - Recognizing new patterns.
4. Clustering - Facts which are not known previously, will be recognized collectively and documented visually.
5. Forecasting - Discover patterns in data that can lead to reasonable predictions about the future. This area of data mining is known as predictive analytics.

3. Cloud Computing Overview

Cloud computing structure has been divided into three layers:

I. Infrastructure as a Service (IAAS),

II. Platform as a Service (PAAS),

III. Software as a Service (SAAS). Out of the above SaaS is king of all the services.

IaaS:

- Provides computer infrastructure as a beneficial service, typically in an environment which is not present Physically but existing virtually.
- Provides a large embryonic for enhancing and scaling.

PaaS

- It provides answer to infrastructure related queries using cloud infrastructure.
- Sits on a top of the IaaS architecture and combine the development and middleware capabilities as well as database, messaging and queuing functions.

SaaS :

- It opens the gates for accessing a huge number of softwares over the internet via a cloud Infrastructure.
- Lying above IaaS and PaaS Layer it acts as a leader from the front.

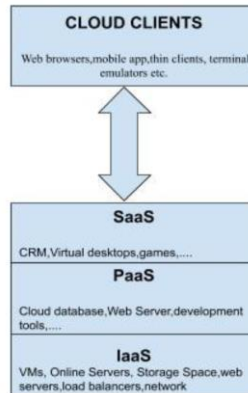


Figure 1. Layers of Cloud Computing.

4. Reducing Cost of Data Mining Using SaaS

SaaS (Software as a service) model helps in decreasing overall cost of using Data mining concepts by outsourcing the hardware requirements. At SaaS software cost was reduced as software need not be installed at the user end, it is provided by the software service provider over cloud infrastructure. It means that the overall responsibility related to software and hardware and updates related issues is beared by service provider and the client only needs to pay for the same. In cloud mining, the mining done by the servers which are on cloud.

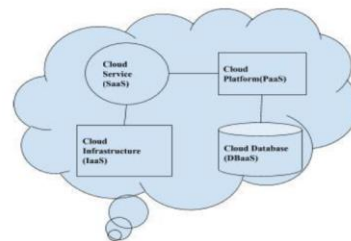


Figure 2. Cloud Computing Basic Architecture.

The Cloud used in cloud mining may be a public cloud (like Google cloud, Amazon's AWS, Microsoft Azure [8]) or a private cloud (such as SAP HANA Cloud, DELL Cloud etc. [4]). However both have pros and cons of their own for example in the case of public cloud, customers will get all the required resources by just paying a minimal amount as compared to their actual cost, secondly customers need not to worry about updations in the softwares as those are maintained by the Cloud service provider themselves. However on the dark side as the entire data is on third party's servers privacy issues may arise, moreover to make data secure customer may have to spend more as compared to the original cost. In case of Private cloud security issues are lesser as compared to public cloud but the organization has to bear license and updations costs, moreover trained persons required to maintain their public cloud.

5. Key Reasons behind the Adoption

Following are some of the main reasons why cloud mining is the coin of the day:

- a. Increased reliability and dependability on web/mobile applications.
- b. Increased security which is well trusted by the customers.
- c. Increased and improved bandwidth. Low cost of ownership.

6. Mining of Data Using Cloud

There are various companies which take initiative towards providing data mining using cloud based services, such as Layered technologies a leading company in providing on demand infrastructure has developed a hybrid cloud (hybrid cloud is a mixture of characteristics of both public and private cloud) infrastructure to its customers through virtualized environment on dedicated servers within Layered Tech. data centers [7], also it provides flexibility and security to access their VPDC (Virtual Private Data Centers) through dedicated lines either using VPN or Internet. Layered Tech has overcome the user concerns regarding the security of their data at enterprise level. Customers have options to choose from various platforms they need such as 3-Tera's App Logic, VMware etc. Layered Tech also provides various basic Data mining tools for its customers on this hybrid

cloud platform. Microsoft is another name of company which is working in the field of data mining through cloud with its cloud platform for data mining named as “DMCloud”. This platform allows its users to perform basic data mining tasks in connection to cloud based analysis. DM Cloud [5] is a valuable service which can consider SQL based mining without putting a burden on users (which are technical professionals) to first Install additional analysis services (as they are already inbuilt in DM Cloud). Clients of DM Cloud can use its services no matter where they are geographically as long as they have a working internet connection. DM Cloud provides some basic data mining [6] operations as in traditional data mining in excel. Some of these basic tasks are:

- (a) Analyze Key Influencers
- (b) Detect categories
- (c) Forecast
- (d) Prediction Calculator
- (e) Shopping Basket analysis

The work is still continued on extending various other features of data mining through cloud.

7. Conclusion and Future Scope

Mining of data with the help of Cloud allows companies to easily manage their assets with the addition of effective, reliable and secure services for their users. In this paper we explored how various tools used for mining the data which works upon SaaS, PaaS and IaaS that are used in cloud computing to retrieve the information. Various big names like Amazon AWS, Microsoft Azure, Openstack and various other big brands are moving towards cloud based data mining. Data mining through cloud helps people to build information listing, get information about various topics by searching in different forums etc. Companies can use cloud mining to see what kind of information is floating on the Internet for their products or services and hence can take actions based on the data presented. Finally we can conclude that as data mining is the need of the time and cloud computing is the future, thus combining both will give a hope for a bright future.

References

- [1] A. Sarkar, A. Bhattacharya, S. Dutta and K. K. Parikh, Recent Trends of Data Mining in Cloud Computing, In: *Emerging Technologies in Data Mining and Information Security, Advances in Intelligent Systems and Computing*, vol 813. Springer, Singapore.
- [2] R. Y. Bhandayker, An Overview of The Integration of All Data Mining At Cloud-Computing, In *Airo International Research Journal Volume XVI, ISSN: 2320-3714 June, 2018*.
- [3] K. Rajamani and D. Sheela, Data Mining Techniques and Algorithms in Cloud Environment-A Review In *International Journal of Pure and Applied Mathematics*, 119(7) (2018), 599-602.
- [4] L. Yu, J. Zheng, W. C. Shen, B. Wu, B. Wang, L. Qian and B. R. Zhang, BC-PDM data mining, social network analysis and text mining system based on cloud computing, In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1496-1499 (2012).
- [5] A. Dhote and S. P. Deshpande, Data Mining with Cloud Computing: - An Overview In *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 1, January 2016*.
- [6] D. Chappell, A short introduction to cloud Platforms: An enterprise-oriented view, White Paper, 13 Pages, San Francisco, Chappell and Associates, 2008.
- [7] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2000.
- [8] Ajmera Rajesh and Siripuri Kiran, Anomaly Detection Using Data Mining Techniques in Social Networking in *International Journal for Research in Applied Science and Engineering Technology*, Volume-6, Issue-II, February 2018, 1268-1272 [ISSN :2321-9653],