International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 4 Issue: 2 099 - 102

Cloud Computing with Reference to Data Mining

Amiya Kumar Sahu,
Research Scholar,
JJT University,
Jhunjhunu(Rajasthan) INDIA

Abstract:-The performance information of knowledge mining techniques throughout Cloud computing can allow the users to induce back of import information from much incorporated data warehouse that reduces the prices of infrastructure and storage. This analysis describes however data processing is employed in cloud computing by scrutiny key options of cloud computing firms. The mix of knowledge mining techniques into traditional regular behavior has become common place. Data processing techniques and applications square measure considerably fascinating within the cloud computing pattern. The information mining in Cloud Computing permits organizations to mix the management of software package and data storage, with guarantee of well organized, consistent and safe services for his or her users.

Keywords: *DM Cloud, CDM, IaaS , SaaS & PaaS*

*****

1. Introduction:

The data mining in Cloud Computing allows organizations to consolidate the management of software and data storage, with guarantee of well organized, consistent and safe services for their users. Data mining technologies provided through Cloud computing is an absolutely necessary characteristic for today's businesses to make proactive, knowledge driven decisions, as it helps them have future trends and behaviors predicted. Data mining is the process of extracting structured information from unstructured or semi-structured web data sources. The data mining in Cloud Computing allows organizations to centralize the management of software and data storage, with assurance of efficient, reliable and secure services for their users. All this accumulated data is potentially hiding in (useful) information, such as the buying preferences, financial situation, interests, political views etc. of users or clients, which can significantly improve the decision-making. But, how to get to this hidden and potentially useful information, which is hidden in the "sea of data" when processing and storing large amounts of data, which is daily multiplying, represents a significant problem and reveals certain limitations of the traditional information and communication technologies and tools? The answer is, of course, the application of modern technology. Cloud infrastructure can be effectively used for intensive and demanding operations with data that is typical for processes of data mining. It is necessary to have available scalable data warehouses and scalable computing resources that are capable to accept, efficiently store and deeply analyze such large amounts of data, and Cloud offers that, without huge investments that are necessary if one wants to build a DM system within the IS company or organization.

1.1 Data Mining

Data mining, the extraction of hidden predictive information from large databases, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviors, allowing businesses to make proactive, knowledge-driven decisions. Data Mining, or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools can answer business questions that traditionally were time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectations.

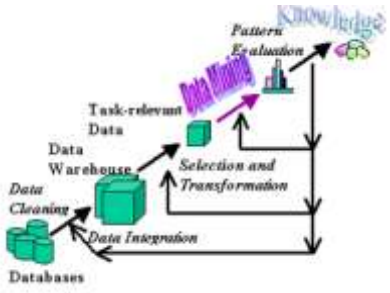

Figure-1 Data Mining is the core of Knowledge Discovery Process [24]

Data mining parameters include:

- Association - looking for pattern where one occurrence is associated to another occurrence.
- Series or path analysis - looking for patterns where one event leads to another later event.
- Categorization - looking for new patterns (May result in a change in the way the data is organized but that's ok)
- Clustering - Verdict and visually documenting groups of facts not before known.
- Forecasting - discovering patterns in data that can lead to reasonable predictions about the future (This area of data mining is known as predictive analytics.) [24]

1.2 Cloud Computing

We can perceive the cloud computing loosely whereas understanding the essential word of the cloud computing because the cloud contains numerous jargons associated with cloud computing that makes the understanding concerning the cloud additional accessible. so a Cloud computing may be a model for enabling omnipresent, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications and services) that may be speedily

provisioned and free with negligible management effort or service supplier interaction.

1.3 Data mining in Cloud Computing

CDM (Cloud Data Mining) offers tremendous potential for analyzing and extracting the (useful) information in various fields of human activities: finance, banking, medicine, genetics, biology, pharmacy, marketing, etc. The application of this technology should enable that with just a few clicks of the mouse one can reach the desired information about customers, their habits, interests, purchasing power, frequency of purchases of certain items, location and so on. Cloud should enable everyone to use this potential, providing, in the form of service, what was recently reserved only for the big (and rich) companies. Small and medium-sized companies that do not have sufficient funds to invest in (too) expensive systems, now have the opportunity to rent a Cloud provides technology that can "handle" huge amounts of data, which cannot be processed efficiently and at reasonable cost using standard technologies and techniques. Analyzing data which courses to social networks, pattern recognition, processing of large-scale images, encryption and description and, of course, data mining is just one of the examples of the tasks that are ideal for implementation in the Cloud. Data mining in cloud computing is the process of extracting structured information from unstructured or semi-structured web data sources. The data mining in Cloud Computing allows organizations to centralize the management of software and data storage, with assurance of efficient, reliable and secure services for their users." [6] As Cloud computing refers to software and hardware delivered as services over the Internet, in Cloud computing data mining software is also provided in this way. The main effects of data mining tools being delivered by the Cloud are:

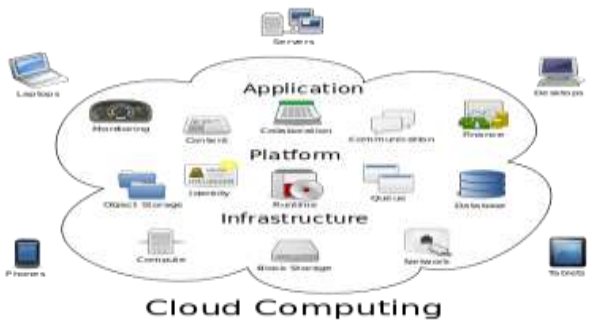

Figure 2: Logical Cloud Computing

• the customer only pays for the data mining tools that he needs – that reduces his costs since he doesn't have to pay for complex data mining suites that he is not using exhaustive;

• the customer doesn't have to maintain a hardware infrastructure, as he can apply data mining through a browser – this means that he has to pay only the costs that are generated by using Cloud computing.

Using data mining through Cloud computing reduces the barriers that keep small companies from benefiting of the data mining instruments. "Cloud Computing denotes the new trend in Internet services that rely on clouds of servers to handle tasks. Data mining in cloud computing is the process of extracting structured information from unstructured or semi-structured web data sources. The data mining in Cloud Computing allows organizations to centralize the management of software and data storage, with assurance of efficient, reliable and secure services for their users." [6] The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage. The architecture of cloud services can be divided into three levels: infrastructure, platform, and application software. Application software builds the user interface and shows the application system's functions. To build a cloud computing application as a service requires infrastructure, platform and application software which can be obtained from a single provider or from different service providers..

2. Background

In a study of the paper literature surrounding cloud computing, we found that there is a distinct focus on the needs of the scientific computing community. The Cloud is going to happen but which services should be offered on the cloud and for whom. What happens if smaller IT companies start to offer their services on the cloud and no one uses them?!We believe that everything eventually can move to the Cloud. The question is if users are ready for that and if it's the right move and this need must be addressed.

In Paper "A survey of mass data mining based on cloud-computing" author focused on solving the problems of distribution of mass data mining and efficient storage computing. The cloud computing model brings many benefits and convenience. This paper introduces the cloud computing and data mining and then simply introduce some existing parallel data mining algorithms based on cloud computing and data mining service platforms.

Cloud computing Impacts and issues paper focused on importance of cloud computing that it provides excellent support for amazing infrastructures, applications and services such as shared resource pool, broad network base, reduced IT cost or rapid elasticity of the cloud to handle varying customers demands as well as cloud computing various service and deployment models which is part of the main reason for adopting this computing system.

Data Mining for Data Cloud and Compute Cloud paper focused on cloud computing is the use of computing resources (hardware and software) that are delivered as a service over a network (typically the Internet). Cloud computing entrusts remote services with a user's data, software and computation. A storage cloud provides storage services, while a compute cloud provides computer services. Moreover in this paper we have reviewed the design and implementation of sector storage cloud and sphere compute cloud. Sector is the distributed file system, while sphere is the parallel in-storage data processing framework that can be used to process data stored in sector.

Review of Data Mining Techniques in Cloud Computing Database focused on data mining is the nontrivial extraction of implicit, previously unknown, and potentially useful information from data. It is the extraction of information from huge volume of data or set through the use of various data mining techniques. The data mining techniques like clustering, classification, neural network, genetic algorithms help in finding the hidden and previously unknown information from the database. Both Data Mining techniques and Cloud Computing helps the business organizations to achieve maximized profit and cut costs in different possible ways. The main aim of the work is to implement data mining technique in cloud computing using Google App Engine and Cloud SQL.

3. Objectives:

The knowledge mining in Cloud Computing permits organizations to alter the management of package and data storage. Exploitation data processing through Cloud Computing reduces the barriers that keep little corporations from benefiting of the information mining instruments. Mining in Cloud Computing permits organizations to alter the management of package and data storage, with assurance of economical, reliable and secure services for his or her users. The implementation of knowledge mining techniques through Cloud Computing can enable the users to retrieve meaning information from nearly integrated data warehouse that reduces the price of infrastructure and storage. CDM (Cloud knowledge Mining) offers tremendous potential for analyzing and extracting the (useful) info in varied fields of human activities: finance, banking, medicine, genetics, biology, pharmacy, marketing, etc. the applying of this technology ought to alter that with simply some clicks of the mouse one will reach the specified info regarding customers, their habits, interests, getting power, frequency of purchases of sure things, location then on.

4. Methods and Tools

We exist in a time of information that is the most important and most expensive resource. Huge amounts of data daily produce and in themselves hide potentially practical information. The data that is processed does not originate only from numerous information systems of companies; giant amount of it comes from "on-line" environment, with a variety of services that users use for both commercial and private purposes. This data contains significant potential, and out of it invaluable information about, for example, buying preferences, financial situation and clients (users) interests can be drawn.

The task of ICT is to create methods and tools for efficient data processing. Today, that is not an easy task, on the contrary; processing and storage of vast amounts of data that are multiplying daily, represents a significant problem and reveals the limitations of the traditional information and communication technologies and tools. For some time, and even presently, a significant problem represents a general lack of funds. Companies are no longer able to invest great funds in the development of their IT sectors. On the other hand, the need for treatment, demanding deep processing and analysis of data has never been greater.

The Microsoft suite of cloud-based services includes a new technical preview of Data Mining in the Cloud "DMCloud". DMCloud allows you to perform some basic data mining tasks leveraging a cloud-based Analysis Services connection. DMCloud is valuable capability for IWs that would like to begin considering SQL Server Data Mining without the added burden of needing a technology professional to first install Analysis Services. Additionally, IWs can use the DMCloud services no matter where they may physically be located as long as they have an Internet connection! The data mining tasks you can perform with DMCloud are the same Table Analysis Tools found in the traditional Excel Data Mining add-in. These data mining tasks include: Investigate Key Influencers Identify classes' Fill From Example Predict Highlight Exceptions Scenario Investigation Prediction Calculator Shopping Basket Investigation.

Data mining is used in a variety of applications such as Health care, Student organization, mathematics, Science, in a variety of website. The cloud model is composed of five necessary individuality, three service models, and four deployment models." The required individuality of cloud computing are on-demand self-service, broad network access, resource pooling, rapid softness and deliberate examine. The service models that make up cloud computing are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS). The deployment models of cloud computing are private cloud, community cloud, public cloud and hybrid cloud. The cloud computing companies and their products key features are:

1. **Sun Microsystems :** Sun Cloud More accessible application than any other open OS.
2. **IBM Dynamic Infrastructure:** Integrated power **management** to help you plan, predict, monitor and actively manage power expenditure of your BladeCenter servers.
3. **Amazon EC2:** Intended to make web scale **computing** easier for developers.
4. **Google App Engine:** No limit to the free trial period if you do not exceed the quota **allotted**.
5. **Microsoft Azure:** Currently offering a "development accelerator" discount plan. 15-**30** % discount off consumption charges for first 6 months.
6. **AT&T Synaptic Hosting:** Employ fully on-demand **infrastructure** or come together it with devoted mechanism to meet specific requirements.
7. **GoGrid Cloud Computing:** Free load **balancing** and free 24/7 support.
8. **Salesforce :** Offers cloud solutions for automation, customer service and platform, respectively. Transparency through real-time information on system performance and security at trust.salesforce.com.

5. Conclusion

The performance information of knowledge mining techniques throughout Cloud computing can allow the users to urge back important information from much incorporated data warehouse that reduces the prices of infrastructure and storage. Data processing is employed for extracting

probably valuable info from data. Knowledge mining in Cloud Computing permits organizations to mix the management of package and data storage, with guarantee of well organized, consistent and safe services for his or her users.

References / Bibliography

[1] L. Li and M. Zhang(2011), The strategy of mining association rule based on cloud computing. In IEEE Computer Society, pages 475–478.
[2] S. Ghemawat, H. Gobioff, and S.-T. Leung(2003), The google file system. SIGOPS Oper. Syst. Rev., 37(5):29–43.
[3] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker(2001). A scalable content-addressable network. SIGCOMM Comput. Commun. Rev., 31(4):161–172,
[4] W. Karim(2004), The Privacy Implications of Personal Locators: Why You Should Think Twice Before Voluntarily Availing Yourself to GPS Monitoring. Washington University Journal of Law and Policy,Vol-3, 14:485– 515.
[5] I. Roy, S. T. V. Setty, A. Kilzer, V. Shmatikov, and E. Witchel.Airavat(2010), security and privacy for mapreduce. In Proceedings of the 7th USENIX conference on Networked systems design and implementation, NSDI'10, Berkeley, CA, USA, USENIX Association,(vol-5)(120-150),pages 20–20.
[6] M. J. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge.(2001), Knowledge management and data mining for marketing. Decis. Support Syst., 31(1):127–137.
[7] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan. Chord(2001), A scalable peer-to-peer lookup service for internet applications. SIGCOMM Comput. Commun. Rev.,vol1, 31(4):149–160.
[8] L. Torgo(2010). Data Mining with R: Learning with Case Studies. Chapman & Hall/CRC,vol1,pg-180-198
[9] Alex Berson (Author), StephenJ.Smith (Author), Berson (Author), KurtThearling (Author)(2011) Building Data Mining Applications for CRM,vol1,pg-15-38.
[10] Todd Arias(2011),The Cloud Computing Handbook - Everything You Need to Know about Cloud Computing,vol1,pg-25-30.
[11] M. Kantardzic.(2002), Data Mining: Concepts, Models, Methods and Algorithms. John Wiley & Sons, Inc.,pg-18-23.
[12] P. Karl Rexer.(2010), 2010 data miner survey highlights the views of 735 data miners,pg-40-50.
[13] N. Santos, K. P. Gummadi, and R. Rodrigues.(2009). Towards trusted cloud computing. In HOTCLOUD. USENIX,vol1,pg-200-210.
[14] M. J. Shaw, C. Subramaniam, G. W. Tan, and M. E. Welge.(2001), Knowledge management and data mining for marketing. Decis. Support Syst., 31(1):127–137.
[15] D. J. Solove.(2007), 'I've Got Nothing to Hide' and Other Misunderstandings of Privacy. Social Science Paper Network Working Paper Series, 44.
[16] ByDinkarSitaram, GeethaManjunath(2010),Data-Driven Marketing,volume2,pg-4-5.
[17] ZebaQureshi, Jaya Bansal,SanjayBansal(2013),A Survey on Association Rule Mining in Cloud Computing,vol3,pg-318-321
[18] VikiPatil,Prof. V. B. Nikam(2013),Study of Data Mining algorithm in cloud computing using,vol2,pg-65-70
[19] NaskarAnkita,Mrs. Mishra Monika R(2013),Using Cloud Computing to provide Data Mining Services,vol2,pg-545-550
[20] Prof. Uzma Ali1,Prof. PunamKhandar(2013),Data Mining for Data Cloud and Compute Cloud,vol1,pg-1137-1141
[21] B. KAMALA(2013),A Study on Integrated approach of Data Mining and Cloud Mining,vol1,pg-35-38
[22] SrishtiSharma,Harshita Mehta(2014),Improving Cloud Security Using data Mining,vol16,pg-66-69
[23] C Mayanka(2014),Implementing Web Mining into Cloud Computing,vol5,pg-62-63
[24] Anthony T. Velte, Toby J. Velte, Robert Elsenpeter (2010), Cloud Computing: A Practical Approach, pg-173-191
[25] John W. Ritting house, James F. Ransome(2010),Cloud Computing - Implementation, Management, and Security,pg-29-55