A Study On Cloud Computing Data Mining

Mr.A.Srinivas, M. Kalyan Srinivas, A.V.R.K.Harsha Vardhan Varma

HOD and Associate Professor, Dept. Of CSE, Coastal Institute of Technology & Management, Vizianagaram, India

Students of Computer Science Engineering Coastal Institute of Technology & Management, Vizianagaram, India

Students of Computer Science Engineering Coastal Institute of Technology & Management, Vizianagaram, India

Abstarct: Cloud Computing has become a main source for the data processing, storage and distribution. The storage of the data is simple and free to use. In data mining the data which is used as data security in a parallel computing platform. The some of the key features are used for the distribution of the data in certain things for the user understandable language. As we implemented the cloud storage in different servers for the security reasons data mining concept is used for the efficiency of the each part of the data is in a secure state. According to this concept we use data effectiveness from some so the supports of the AnazonEC2 map reduce platform. This paper is an approach for the experimental result and documented for the cloud computing data mining.

Keywords: cloud computing, data mining, data mining in cloud computing

I.   INTRODUCTION

Data mining has been an effective tool to analyse data from different angles and getting useful information from data. Classification of data, categorization of data, and to find correlation of data patterns from the dataset. On the other hand, challenges as data storage and transfer approaches need to deal with prohibitive amount of data. The management of data resource and dataflow is becoming the main bottleneck. Large data set has become a major challenge and data intensive computing is now considered as the "fourth paradigm" in scientific discovery after theoretical, experimental, and computation science.

The internet is becoming an increasingly vital tool in everybody's life, both professional and personal, as its user and becoming more numerous. The most revolutionary concept of recent year is Cloud Computing. Many companies are choosing as an alternative to building their own IT infrastructure to host database or software, having a third party to host them on its large servers, so company's would have access to its data and software over the Internet.

The use of cloud computing is gaining popularity due to its mobility, huge availability and low cost. On the other hand it brings more threats to the security of the company's data and information. In recent years, data mining techniques have evolved and become more used, discovering knowledge in database becoming increasingly vital in various fields: business, medicine, science and engineering, spatial data etc.
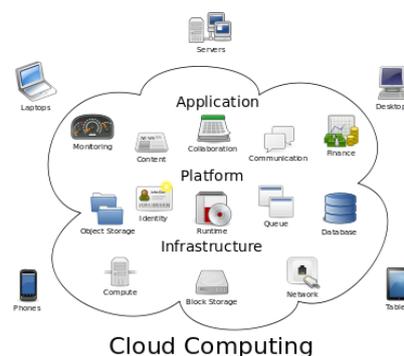
fig-1: Transferring data from one server to another server through the data mining.

II. ASPECTS REGARDING CLOUD COMPUTING

In 1960's-1990's **John McCarthy** who has opined the modern-day characteristics of cloud computing, later in 1990's telecommunication companies who previously offered primarily dedicated point to point data circuit, began offering virtual private network(VPN) services with comparable quality of service, but at a lower cost. By switching traffic as they saw fit to balance server use, they could use overall network bandwidth more effectively. They began to use the cloud symbols to donate the demarcation point between what the provider was responsible for cloud computing extends this boundary to cover server as well as the network infrastructure.

As computer become more prevalent, scientists and technologists explored ways to make large-scale computing power available to more users through time sharing, experimenting with algorithms to provide the optimal use of the infrastructure, platform and application with prioritized access to CPU and efficiency for the end users.

Since 2000 the dot-com bubble, Amazon played a key role in all the developing in cloud computing by modernizing their data center which like the most computer network, were using as little as 10% of their capacity at any one time, Just to leave room for occasional spikes having found that the new cloud architecture resulted in significant interval efficiency improvements by whereby small, Fast moving "two-pizza terms" cloud added new features faster and more easier, Amazon indented a new product development effort to provide cloud computing to external customers, and launched Amazon web services (AWS) on utility computing basis in 2006.

In early 2008, Eucalyptus became the first open source the federation of the clouds in the same year efforts where focused on providing quality of service guarantees to cloud based infrastructure, in the frame works of the IRMOS European commission funded project resulting to a real time cloud environment by mid 2008, Gartner saw an opportunity for cloud computing to shape the relationship among consumers of IT services , projected shift to computing will result in dramatic growth in IT products in some areas and significant reduction in other areas. On march 1, 2011, IBM announced the IBM Smart Cloud framework to support smarter Planet, Among the various components of the smarter computing foundation, cloud computing is a critical piece.

III. DIFFERENT CHARACTERISTICS FOR CLOUD COMPUTING

Client server model : Client server computing refers broadly to any distributed application that distinguishes between service providers and service requesters.
Grid computing: A form of all distributed and parallel computing, where by super and virtual composed by clusters of networked loosed coupled computers acting in concert to perform very large tasks.
Mainframe computer: Powerful computers where used mainly by large origination for critical applications ,typically bulk data processing such as censes, industry costumers statistics ,police and secret agent service. Enterprises resources planning and financial transaction processing.
Utility computing: The package of computing resource, such as computing and storage, as a mattered service similar to traditional public utility such as electricity.
Peer to peer: It means the distributed architecture without the need of central coordination participants are both suppliers and consumers of resources in contrast other model of client server model.

IV. CLOUD GAMING

It is also known as on demand gaming is a way to deliver games to the computers. Gaming data is stored in provider's servers, so that the gaming is independent of client computers used to play games.

Cloud computing exhibits the following key characteristics: Agility improves with users' ability to re-provision technological infrastructure resources.
Application programming interface: (api) accessibility to software that enables machines to interact with cloud software in the same way that a traditional user interface cloud computing systems typically use representational state transfer based (apis).
Cost: cost is claimed to reduced, and in a public delivery model capital expenditure is converted to operational expenditure as infrastructure is typically provided by a 3rd party and does not need to purchase for one-time or infrequent intensive computing tasks.
Virtualization: Virtualization Technology allows servers and storage device to be stored and utilization be increased applications can be easily migrated from one physical server to another. Centralization of infrastructure in location with lower cost such as electricity. Peak load capacity increases users need not engineer for highest possible load level.
Utilization and efficiency improvements for system that are often only 10-20% utilized Reliability: Reliability is

improved if multiple redundant sites are used, which makes well-designed cloud computing suitable for business continuity and disaster recovery.

Security: Security could improve due to centralization of data, increased security-focused resources. But concern can persist about loss of control over certain sensitive data, and the lack of security for stored kernels. Security is often as good as or better than other traditional system, in part because providers are able to devote resources to solving security issues is greatly increased when data is distributed over a wider area or greater number of devices and in multi-tenant system that are bring shared by unrelated user. Private Cloud installations are in part motivated by users for desire and control over the infrastructure and avoid losing control of information security.

Scalability and elasticity: Scalability and elasticity via dynamic provisioning of resource on fine=grained, self- service basis near real-time, without user having to engineer for peak loads.

Performance: Performance is monitored, and consistent and loosely coupled architecture are constructed using web services as the system interface.

Maintenance: Maintenance of Cloud Computing application is easier because they do need to be installed on each user's computer and can be accessed from different places.

Top Cloud Computing companies and Key Features:

Cloud Name	Key Feature
Sun Microsystems Sun Cloud	More available application than any other open os
IBM Dynamic Infrastructure	Integrated power management to help you plan, predict, monitor and activity manage power actively manage power consumption of your Blade Centre servers.
Amazon EC2	Designed to make wed-scale computing easier for developers
Google App Engine	No limit to free trial period if you do not exceed the quota allotted.
Microsoft Azure	Currently offering a "development accelerator" discount plan. 15-30 % discount off consumption changes for first 6 months.
AT&T Synaptic Hosting	Use fully on-demand infrastructure or combine it with dedicated components to meet specialized requirements.
Go Grid Cloud computing	Free load balancing and free 24/7 supports.
Salesforce	Offers cloud solutions for automation and platform system performance and security at trust.salesforce.com
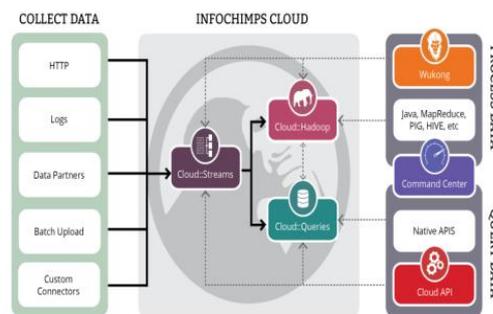
FIG-2: DATA FLOW IN THE CLOUD FOR HADOOP

Some Aspects Regarding Data mining: Data mining represents finding useful patterns or trends through large amounts of data. Data mining is defined as a "Type of Database Analysis that attempts to discover useful patterns or relationships in a group of data. The analysis uses advanced statistical methods, such as cluster analysis, and sometimes employs artificial intelligence or neural network techniques. A major goal of data mining is to discover previously

unknown   relationships   among   the   data,   especially   when   the   data   come   from   different   databases.

Some of the important for data mining technique:

Cloud name	Key feature
Clustering	Useful for exploring data and finding natural groupings. Members of cluster like more each other than they are like the members of other clusters common example including finding a new customer segments and life science discovery
Classification	Most commonly used technique for predicting a specific outcome such as response high/medium like to buy/not to buy
Association	Find rules associated frequently co-occurring items, used for market basket analysis, cross-sell, root cause analysis.
Regression	Technique for predicting a continuous numerical outcome such a customer process yield rates.
Attribute Importance	Ranks attributes according to strength of relationship with target attribute. Use cases include finding factors most associated with customers who respond to an offer
Anomaly Detection	Identifies unusual or suspicious cases based on deviation from the norm. Common examples include health care fraud, expense report fraud, and tax compliance.
Feature Extraction	Produces new attributes as linear combination of existing attributes. Applicable for text data decomposition and projection and pattern recognition

   Data mining is the extraction of hidden predictive information from large database, is a powerful new technology with great potential to help companies focus on the most important information in their data warehouse. Data mining tools predict future trends and behaviour, allowing businesses to make proactive, Knowledge- driven decisions. Businesses can make predictions about how well a product will sell or develop new advertising campaigns by using these new relationships reflected by the data mining algorithms. Data mining uses information from past data to analyse the outcome of a particular problem or situation that may arises. Data mining work to analyse data stored that data is bring analysed. That particular data may come from all parts of business from the production to the management. Managers also use data mining to compare and contrast among competitors. Data mining interrupt   its data into time analysis that can be used to increase sales, promote new product, or delete that is not value added to the company.

- Offices requiring analysis or dissemination of geo-referenced statistical data.
- Public health services searching for explanations of disease clustering.
- Environment agencies assessing the impact of changing land use patterns on climate change.
- Geo-marketing companies doing customer segmentation based on spatial location.
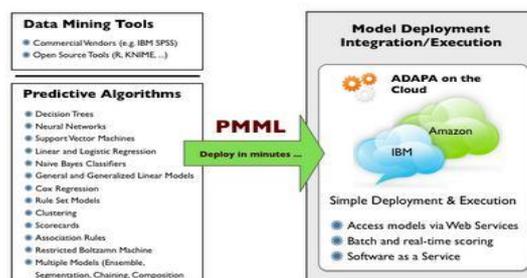
Fig-3: Data Mining in Cloud Computing

Data mining in a cloud computing: Data mining is one of the fastest growing fields in computer industry that deals with discovering patterns from large data sets. It is a part of knowledge discovery process and is used to extract human understandable information Mining is preferably used for a large amount of data and related algorithms often require large data sets to create quality models .The relationship between data mining and cloud is worth to discuss. Cloud

providers use data mining to provide clients better service. If clients are unaware of the information being collected ethical issues like privacy and individuality are violated. This can be serious data privacy issue if the cloud providers misuse the information. Again attackers outside cloud providers having unauthorized access to the cloud, also have the opportunity to mine cloud data. In both cases, attackers can use cheap and raw computing power provided by cloud computing to mine data and thus acquire useful information from data. The data mining in cloud computing allows organization to centralize the management of software and data storage with assurance of efficient reliable and secure service for their user.

The main effects of data mining tools being delivered by Cloud are: The customer only pays for the data mining that is need to reduce his cost for complex data mining. The customer do not have maintains of hardware infrastructure as he can apply for data mining through a browser.

Using cloud computing through data mining reduces the barriers that keep small companies from benefiting of the data mining instruments. The relationship between cloud computing through the Data Mining is the cloud uses to store the data in into the server and data mining is uses to provide clients server relationship as a service and the information being collected ethical issues like privacy and individuality and violated. According to the cloud computing the security reasons are less and the data can be loss for previous reasons they use data mining for the security reason in cloud from the attackers. The attackers can use cheap and raw computing for hacking the data base in the storage of the cloud so the data can be loss in server. Some mining algorithms are good enough to extract information up to the limit that violates client privacy.

For example: multivariate analysis identifies.

The relationship among variables and this technique can be used to determine the financial condition of an individual from his buy-sell records, clustering algorithms can be used to categorize people or entities and are suitable for finding behavioural patterns; association rule mining can be used to discover association relationships among large number of business transaction records etc. Analysis of GPS data is common nowadays and the results of such analysis can be used to create a comprehensive profile of a person covering his financial, health and social status. Thus analysis of data can reveal private information about a user and leaking this sort of information may do significant harm. As more research works are being done on mining, improved algorithms and tools are being developed. Thus, data mining is becoming more powerful and possessing more threat to cloud users. In Upcoming days, data mining based privacy attack can be a more regular weapon to be used against cloud users. In this approach we use the cloud computing through data mining in privacy reason.

V.   CONCLUSION

Cloud computing provides storage of data in a server by protecting data by using data mining concept. Actually we are discussing the cloud computing data mining for the advance use of security in data loss purpose. While the data we are storing in cloud is being separated in different servers for a security but the hackers using the cheap and raw cloud computing for the misuse of the software.

In Cloud computing the data is being shifted from one server to another server in a peer to peer transaction.

For example take as www.torrentz.ud   as a cloud computing data base in this torrent which is used as the data transaction from one to another from peer to peer and seed to seed and leachers to leachers because the data which is stored in each part of the server. As this cloud computing Data mining topic is used as a data encryption or data security from the data base.

REFERENCES

[1]   "CloudComputingNewWine.pdf."- http://www.cmlab.csie.ntu.edu.tw/~jimmychad/CN2011/Reading/CloudComputingNewWine.pdf.
[2]   "Top-Cloud-Computing-companies"-http:www.itstrategists.com/Top-Cloud-Computing-companies.aspx.
[3]   "Cloud Computing: Data-Intensive Computing and Scheduling"-By Frédéric Magoulès, Jie Pan, Fei Teng
      http://books.google.co.in/books?id=06c4jn6kV7AC&printsec=frontcover&dq=cloud+computing+based+on+data+mining+reference+books&hl=en&sa=X&ei=AWXaUdW4GMXprAfDq4DACA&ved=0CDcQ6AEwAA#v=onepage&q&f=false
[4]   "Data mining and Analytics"- http://cseweb.http://cseweb.ucsd.edu/~elkan/255/books.html
[5]   "Data mining concept by Doug Alexander"-http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/
[6]   "Concepts of cloud computing"-http://searchitchannel.techtarget.com/tip/Key-aspects-of-cloud-computing-services
[7]   "Different characteristic for the cloud computing"-deca.cuc.edu.cn/.../The-Characteristics-of-Cloud-Computing.pdf
[8]   "Grid Computing"-http://www.gridcafe.org/EN/what-is-the-grid.html
[9]   "Mainframe Computer"-http://www.businessdictionary.com/definition/mainframe-computer.html
[10] "Utility Computing"-http://www.techopedia.com/definition/14622/utility-computing
[11] Peer to Peer"-http://www.bleepingcomputer.com/glossary/definition125.html
[12] "Cloud Gaming"-http://www.techopedia.com/definition/26527/cloud-gaming
[13] "Key concepts for the cloud computing"-http://new.itstrategists.com/Top-Cloud-Computing-Companies.aspx
[14] "Cloud Storage Infrastructure"-http://www.ibm.com/developerworks/cloud/library/cl-cloudstorage/
[15] Key concepts of the data mining by Lijia Guo"-http://www.casact.org/pubs/forum/03wforum/03wf001.pdf

BIOGRAPHY

Srinivas Adapa S/O A Siva Rao was born in Visakhapatnam (Dist) Andhra Pradesh, India. He did his graduation, Post Graduation in Engineering with consistently good academic record. He did his post graduation degree from Charles Sturt University, Australia and having 12 years experience which includes more than 8 years of IT industry experience in India & Abroad. He is pursuing PhD in New World Mission Dunamis International University, South Africa.

Kalyan Srinivas M S/O M B A Rao was born in Visakhapatnam (dist) Andhra Pradesh India. He is pursuing B.Tech final year in Coastal Institute of Technology and Management, Vizianagaram India.

Harsha vardhan varma A V R K S/O A Lakshmipati Raju born in Vizianagaram (dist) Andhra Pradesh India. He is pursuing B.Tech final year in Coastal Institute of Technology and Management, Vizianagaram India.