International
Journal
Of Advanced Research in Engineering & Management (IJAREM)

Data Mining in Cloud Computing

Madhusudhan.R.Anegundi
Guest Lecturer, Dept of Computer Science,
Govt. Degree College, Sindhanoor

***Abstract*—** This paper describes how data mining is used in cloud computing. Data Mining is a process of extracting potentially useful information from raw data. The integration of data mining techniques into normal day-to-day activities has become common place. Every day people are confronted with targeted advertising, and data mining techniques help businesses to become more efficient by reducing costs. How (SaaS) and (PaaS) is very useful in cloud computing. Data mining applications can derive much demographic information concerning customers that was previously not known or hidden in the data. We have recently seen an increase in data mining techniques targeted to such applications as fraud detection, identifying criminal suspects, and prediction of potential terrorists. By and large, data mining systems that have been developed to data for clusters, distributed clusters and grids have assumed that the processors are the scarce resource, and hence shared.

Keywords— Cloud Computing, Data mining, Data Mining Trends, Data Mining Tools, How data mining are used in cloud computing.

I. INTRODUCTION

The Internet is becoming an increasingly vital tool in our everyday life, both professional and personal, as its users are becoming more numerous. It is not surprising that business is increasingly conducted over the Internet. Perhaps one of the most revolutionary concepts of recent years is Cloud Computing. The Cloud, as it is often referred to, involves using computing resources – hardware and software – that are delivered as a service over the Internet.

Cloud computing facilitates end-users or small companies to use computational resources such as software, storage, and processing capacities belonging to other companies (cloud service providers). Cloud services include Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [2]. Big corporate like Amazon, Google and Microsoft are providing cloud services in various forms. Amazon Web Services (AWS) provides cloud services that include Amazon Elastic Compute Cloud (EC2).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. At an equally significant extent in recent years, data mining techniques have evolved and became more used, discovering knowledge in databases becoming increasingly vital in various fields: business, medicine, science and engineering, spatial data etc. The emerging Cloud Computing trends provides for its users the unique benefit of unprecedented access to valuable data that can be turned into valuable insight that can help them achieve their business objectives.

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. The automated, prospective analyses offered by data mining move beyond the analyses of past events provided by retrospective tools typical of decision support systems. Data mining is sorting through data to identify patterns and establish relationships.

II. DATA MINING

Definition of Data Mining.

"Data Mining represents a process developed to examine large amounts of data routinely collected. The term also refers to a collection of tools used to perform the process. Data collected from various areas such as marketing, health, communication, etc., are used in data mining." "Data Mining is the extraction of hidden

predictive information from large databases; it is a powerful technology with great potential to help organizations focus on the most important information in their data warehouse."Questions those traditionally were too time consuming to resolve can be answered by the data mining tools in an effective manner. This helps to find the hidden patterns; predictive information that facilitates the experts with solution outside their expectations. The goal of data mining is to extract knowledge from dataset in human-understandable structures. In recent years data mining has been used widely in the areas of science and engineering, such as bioinformatics, genetics, medicine, education and engineering.

III. TRENDS IN DATA MINING

A. Historical Trends

Data mining application era was perceived in early 1980s principally focused on single tasks driven by research tools. Data mining is helpful in various disciplines like Data Base Management Systems (DBMS), Artificial Intelligence (AI), Machine Learning (ML) and Statistics. Historical trends of data mining are explained as follows:

1) *Data Trends:* Data mining algorithm work best with the numerical data especially collected from a single data base and various data mining techniques have developed for flat files, traditional and relational database where the data is mostly represented in the tabular form. Afterwards, with the convergence of Statistics and Machine Learning pave way to the evolution of various algorithms to mine the non numerical data and relational data bases.
2) *Computing Trends*: Development in fourth generation programming language influenced much in the field of data mining and various related computing techniques. Initially, most of the algorithms engaged to work only on statistical
techniques. Various computing techniques such as AI, ML and pattern reorganization evolved to do the data mining tasks in ease manner. Various data mining techniques like Induction, Compression, approximation and other algorithms developed to mine the large volume of heterogeneous data stored in the data warehouse.

B. Current Trends

Advancement in data mining with various integrations and implications of the methods and techniques have formed the present data mining applications to tackle the various challenges. The current trends of data mining application are described as follows:

C. Future Trends

Data mining has been acquiring noteworthy amount of importance in recent years and it has a strong industrial impact. Future of data mining companies would be promising in the coming years based on this observation. A huge amount of data gets agitate in the research, medical, corporate and media industries as it becomes great for anybody involves in gathering useful information. Increasing technology and future application areas always creates new challenges and opportunities for data mining. Advance data mining techniques can be developed and used by R& D and other information rich companies to discover useful patterns that can help in research or business development to ensure the growth and development of the companies.

TABLE I
CURRENT DATA MINING AREAS, TECHNIQUES AND VARIOUS DATA FORMAT

DATA MINING TYPE	APPLICATION AREAS	DATA FORMATS	DATA MINING TECHNIQUES/ALGORITHMS
TIME SERIES DATA MINING	BUSINESS AND FINANCIAL APPLICATIONS.	TIME SERIES DATA	RULE INDUCTION ALGORITHMS
HYPERMEDIA DATA MINING	INTERNET AND INTRANET APPLICATIONS.	HYPER TEXT DATA	CLASSIFICATION AND CLUSTERING TECHNIQUES
MULTIMEDIA DATA MINING	AUDIO/VIDEO APPLICATIONS	MULTIMEDIA DATA	RULE BASED DECISION TREE CLASSIFICATION ALGORITHMS
SPATIAL DATA MINING	NETWORK, REMOTE SENSING AND GIS APPLICATIONS.	SPATIAL DATA	SPATIAL CLUSTERING TECHNIQUES, SPATIAL OLAP

IV. CATEGORIES OF DATA MINING TOOLS

Most of the data mining tools can be classified into three categories: Traditional data mining tools, dash boards and text-mining tools. Description of each is as follows:

A. Traditional Data Mining Tools

Traditional mining programs help the companies to establish data patterns and trends by using various complex algorithms and techniques. Some of these tools are installed on the desktop computers to monitor the data and emphasize trends and others capture information residing outside a data base. Majority of these programs are supported by windows and UNIX versions. However, some software specializes in one operating system only. In addition to that some may work in only one database type. But, Most of the software will be able to handle any data using online analytical processing or a similar technology.

B. Dashboards

Dashboards reflect data changed and update on screen. Dashboards are normally installed in computers to monitor information in a database and it reflects data changes and updates the data in the form of a chart or table on the screen. It enables the user to see how the business is performing. Historical data can be referenced and checks against the current status in order to see the changes in the business. By this way, dashboards is very easy to use and helps the manager a lot with great appeal to have an overview of the company's performance.

C. Text-Mining Tools

The third type of data mining tools is called as a text-mining tool because of its ability to mine data from different kind of text starting from Microsoft Word, Acrobat PDF documents to simple text files. This provides facility of scanning the content and converts the selected into a format that is compatible with the tools database without opening different applications.

V. WHAT IS CLOUD COMPUTING

Cloud computing is a general term for anything that involves delivering hosted services over the Internet. These services are broadly divided into three categories: Infrastructure as a Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). The name cloud computing was inspired by the cloud symbol that's often used to represent the Internet in flowcharts and diagrams.

The term "cloud" is used as a metaphor for the Internet, based on the cloud drawing used in the past to represent the telephone network, The actual term "cloud" borrows from telephony in that telecommunications companies, who until the 1990s offered primarily dedicated point-to-point data circuits, began offering Virtual Private Network(VPN) services with comparable quality of service but at a much lower cost. In early 2008, Eucalyptus became the first open-source, AWS API-compatible platform for deploying private clouds. June 2, 2008 - Cloud computing is becoming one of the next industry buzz

words. It joins the ranks of terms including: grid computing, utility computing, virtualization, clustering, etc.

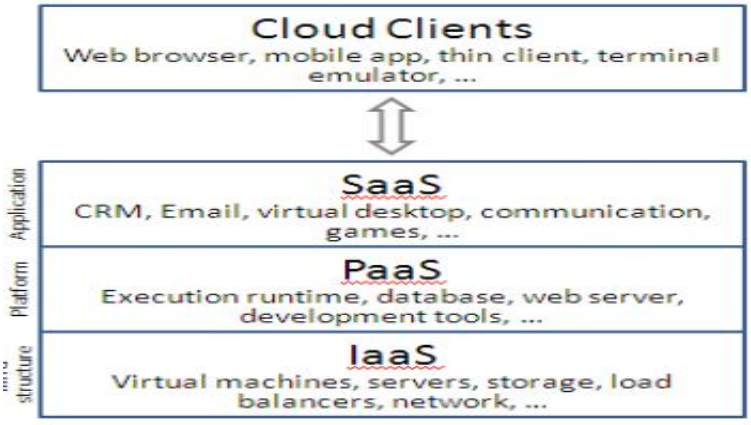
Cloud computing overlaps some of the concepts of distributed, grid and utility computing, however it does have its own meaning if contextually used correctly. The conceptual overlap is partly due to technology changes, usages and implementations over the years.
The cloud is a virtualization of resources that maintains and manages itself. There are of course people resources to keep hardware, operation systems and networking in proper order. But from the perspective of a user or application developer only the cloud is referenced.

VI. REGARDING CLOUD COMPUTING.

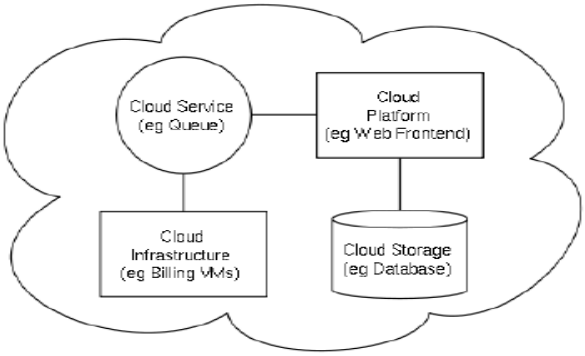
Cloud computing represents both the software and the hardware delivered as services over the Internet.
Cloud Computing is a new concept that defines the use of computing as a utility, that has recently attracted significant attention.
The computing paradigm shift on the last half century through six distinct phases:
• Phase 1: people used terminals to connect to powerful mainframes shared by many users.
• Phase 2: stand-alone personal computers became powerful enough to satisfy users' daily work.
• Phase 3: computer networks allowed multiple computers to connect to each other.
• Phase 4: local networks could connect to other local networks to establish a more global network.
• Phase 5: the electronic grid facilitated shared computing power and storage resources.
• Phase 6: Cloud Computing allows the exploitation of all available resources on the Internet in a scalable and simple way.

There are three types of cloud services infrastructure as a Service, platform as a Service, Software as a Service. In which SaaS is king of all the services.

- ✓ IaaS :
  - • Delivers computer infrastructure as a utility service, typically in a virtualized environment.
  - •Provides enormous potential for extensibility and scale.
- ✓ PaaS :
  - •Delivers a platform or solution stack on a cloud infrastructure.
  - •Sits on a top of the IaaS architecture and integrates with development and middleware capabilities as well as database, messaging and queuing functions.
- ✓ SaaS :
  - •Delivers the application over the Internet or Intranet via a cloud Infrastructure.
  - •Built on underlying IaaS and PaaS Layer.



Layers of Cloud Computing



Cloud Computing Sample Architecture

TABLE 2
TOP CLOUD COMPUTING COMPANIES AND KEY FEATURES

CLOUD NAME	KEY FEATURE
SUN MICROSYSTEMS SUN CLOUD.	MORE AVAILABLE APPLICATION THAN ANY OTHER OPEN OS.
GO GRID CLOUD COMPUTING	FREE LOAD BALANCING AND FREE 24/7 SUPPORT.
GOOGLE APP ENGINE	NO LIMIT TO THE FREE TRIAL PERIOD IF YOU DO NOT EXCEED THE QUOTA ALLOTTED.
AT&T SYNAPTIC HOSTING	USE FULLY ON-DEMAND INFRASTRUCTURE OR COMBINE IT WITH DEDICATED COMPONENTS TO MEET SPECIALIZED REQUIREMENTS
AMAZON EC2	DESIGNED TO MAKE WEB-SCALE COMPUTING EASIER FOR DEVELOPERS.

Cloud computing represents all possible resources on the Internet, offering infinite computing power. As cloud computing is becoming a more significant technology trend, it could reshape the IT sector and the IT marketplace.

VII. DATA MINING IN THE CLOUD

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 a
better service. If clients are unaware of the information being collected, ethical issues like privacy and individuality are violated. This can be a 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. According to the survey done by Rexer Analytics, 7% data miners use cloud to analyze data. As cloud is a massive source of centralized data, data mining gives attackers a great advantage in extracting valuable information and thus violating clients 'data privacy.

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:

- Analyze Key Influencers
- Detect Categories
- Fill From Example
- Forecast
- Highlight Exceptions
- Scenario Analysis
- Prediction Calculator
- Shopping Basket Analysis

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." "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."
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.

VIII. CONCLUSIONS

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. This paper provides an overview of the necessity and utility of data mining in cloud computing. As the need for data mining tools is growing every day, the ability of integrating them in cloud computing becomes more and more stringent.
Data mining is useful for both public and private sectors for finding patterns, forecasting, discovering knowledge in different domains such as finance, marketing, banking, insurance, health care and retailing. Data mining is commonly used in these domains to increase the sales, to reduce the cost and enhance research to reduce costs, enhance research.

REFERENCES

[1] Bhagyashree Ambulkar and Vaishali Borkar, "Data Mining in Cloud Computing", MPGI National Multi Conference 2012 (MPGINMC-2012), 7-8 April 2012.
[2] Mr. S. P. Deshpande and Dr. V. M. Thakare, "Data Mining System And Applications: A Review ," International Journal of Distributed and Parallel systems (IJDPS) Vol.1, No.1, September 2010, pp.32-44.
[3] Karimella Vikram and Niraj Upadhayaya, "Data Mining Tools and Techniques: a review," Computer Engineering and Intelligent Systems, Vol 2, No.8, 2011, pp.31-39.
[4] Mrs. Bharati M. Ramageri, "Data Mining Techniques And Applications," Indian Journal of Computer Science and Engineering, Vol. 1 No. 4, pp.301-305.
[5] Hemlata Sahu, Shalini Shrma and Seema Gondhalakar, "A Brief Overview on Data Mining Survey," International Journal of Computer Technology and Electronics Engineering (IJCTEE)., Vol.1, Issue 3,pp.114-121.
[6] IT Strategists, "Top Cloud Computing Companies and Key Features".
[7] M. Bramer. Principles of Data Mining. Springer, 2007.
[8] Introduction to Cloud Computing Architecture by Sun Microsystems,Inc., june 2009.