Data mining in Cloud Computing

Ruxandra-Ştefania PETRE
Bucharest Academy of Economic Studies
ruxandra_stefania.petre@yahoo.com

This paper describes how data mining is used in cloud computing. Data Mining is used for 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.
Data mining techniques and applications are very much needed in the cloud computing paradigm. 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.*
Keywords: *Cloud Computing, Data mining*

1 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 (shown as a cloud in most IT diagrams).
Many companies are choosing as an alternative to building their own IT infrastructure to host databases or software, having a third party to host them on its large servers, so the company 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.
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.

2 Some aspects 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.
In Figure 1 below it is illustrated the computing paradigm shift on the last half century through six distinct phases: [1]

- 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.

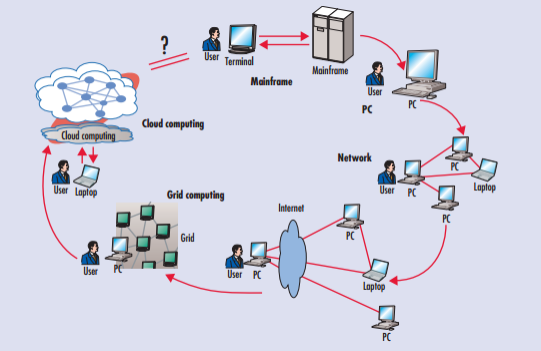

Figure 1. Computing paradigm shift of the last half century [1]

As it is defined by the National Institute of Standards and Technology, "Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.
This cloud model is composed of five essential characteristics, three service models, and four deployment models." [2]
The essential characteristics of cloud computing are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service.
The service models that compose 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.
Table 1 presents details on the top cloud computing companies and their products key features:

Table 1 Top Cloud Computing Companies and Key Features [3]

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 actively manage power consumption of your BladeCenter servers.
Amazon EC2	Designed to make web-scale computing easier for developers.
Google App Engine	No limit to the 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 charges for first 6 months.
AT&T Synaptic Hosting	Use fully on-demand infrastructure or combine it with dedicated components to meet specialized requirements.
GoGrid Cloud Computing	Free load balancing and free 24/7 support.
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.

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.

3 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." [4]

The most important data mining techniques and their description are presented in table 2 below:

Table 2 Data mining techniques [5]

Cloud Name	**Key Feature**
Clustering	Useful for exploring data and finding natural groupings. Members of a cluster are more like each other than they are like members of a different cluster. Common examples include finding new customer segments and life sciences discovery.
Classification	Most commonly used technique for predicting a specific outcome such as response / no-response, high / medium / low value customer, likely to buy / not buy.
Association	Find rules associated with frequently co-occurring items, used for market basket analysis, cross-sell, root cause analysis. Useful for product bundling, in-store placement, and defect analysis.
Regression	Technique for predicting a continuous numerical outcome such a customer lifetime value, house value, 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, factors most associated with healthy patients.
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, latent semantic analysis, data compression, data decomposition and projection, and pattern recognition.

Considering the varied data mining techniques and the great need for discovering patterns and trends in data that would lead to knowledge that could not be obtained otherwise, it's no wonder that data mining is used in the most varies field of activity.

"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." [6]
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.
The medical sector benefits from the data mining techniques, as well as the geographical data being better analyzed by using data mining.
Governments can discern illegal or embargoed activities done by individuals, associations or other governments with the implementation of the data mining techniques.
In short, data mining has developed uses in the majority of field of activity.

4 Data mining in Cloud Computing

Data mining techniques and applications are very much needed in the cloud computing paradigm.
As cloud computing is penetrating more and more in all ranges of business and scientific computing, it becomes a great area to be focused by data mining.
"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]
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:

- 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.

5 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.

References

[1] Jeffrey Voas and Jia Zhang, "Cloud Computing: New Wine or Just a New

Bottle?", IEEE Internet Computing Magazine, 2009, Link: http://www.cmlab.csie.ntu.edu.tw/~jimmychad/CN2011/Readings/CloudComputingNewWine.pdf.

[2] Peter Mell, and Timothy Grance, "The NIST Definition of Cloud Computing", The National Institute of Standards and Technology, USA, 2011, Link: http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

[3] IT Strategists, "Top Cloud Computing Companies and Key Features", Link: http://www.itstrategists.com/Top-Cloud-Computing-Companies.aspx.

[4] Merriam-Webster Dictionary, "Definition of data mining", Link: http://www.merriam-webster.com/dictionary/data%20mining.

[5] ORACLE, "Oracle Data Mining Mining Techniques and Algorithms", Link: http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/odm-techniques-algorithms-097163.html.

[6] Bhagyashree Ambulkar and Vaishali Borkar, "Data Mining in Cloud Computing", MPGI National Multi Conference 2012 (MPGINMC-2012), 7-8 April 2012, Link: http://research.ijcaonline.org/ncrtc/number6/mpginmc1047.pdf.

Ruxandra-Ştefania PETRE graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 2010. She graduated from the Business Support Databases Master of the Academy of Economic Studies in 2012. At present she is a Junior System Architect at LOXON Solutions since November 2011. She is developing and implementing Business Intelligence and Data Warehousing solutions for the banking system.