

Impact of Cloud Computing Datamining in Digitalization

Dr. Rajeev Sharma¹, Dr. Rupak Sharma² and Jai Prakash Bhati³

^{1,2}Assistant Professor Department of Computer Applications SRM University NCR campus Modinagar-201204, India.

³Assistant Professor Department of Computer Science and Engineering, Noida International University Grater Noida, India.

¹Orcid: 0000-0002-1882-7036

Abstract

Digital India is an initiative of Government of India to integrate the government departments and the people of India. The aims of digital India is to minimize the paper work and provide the government services digitally or electronically to the citizen of India. Cloud computing data mining play an important job to extract the useful information. Cloud computing is a kind of computing that concentrate on *sharing computing resources* rather than having local servers or personal devices to handle applications. Data mining is basically a process to retrieve a data from the data store or data ware house. This paper will elaborate how data mining is useful in cloud or cloud computing. Data mining is a process that analyzes a huge amount of data to find new and hidden information that improves business effectiveness in various fields.

Data mining is adopted by various industries to gain the business competitive benefits and that improves the business growth. The mixing of data mining techniques into normal day-to-day activities has become common place. Every day people are faced with targeted advertising, and data mining techniques help businesses to become more efficient by reducing the costs. This paper will provide an overview of the essential and utility of data mining in cloud computing. The accomplishment of data mining techniques in Cloud computing will permit the users to retrieve significant information from practically integrated data warehouse that reduces the costs of storage and infrastructure in digital India.

Keywords: Cloud Computing, Data mining, Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).

INTRODUCTION

The importance of Internet technology is huge in our lives both professional and personal, as Users are becoming more abundant. The business is conducting over the internet now. The one of the most concept in this regard is cloud computing. To store, retrieve the data and programs over the Internet instead of our computer's hard drive is done through Cloud computing. The cloud is just a metaphor for the Internet. It goes back to the days of flowcharts and presentations that

would represent the gigantic server-farm infrastructure of the Internet as nothing but a puffy, white cumulonimbus cloud, accepting connections and doling out information as it floats.

There is an entirely different "cloud" when it comes to business. Some businesses choose to implement Software-as-a-Service (SaaS), where the business subscribes to an application it accesses over the Internet. There's also Platform-as-a-Service (PaaS), where a business can create its own custom applications for use by all in the company. And don't forget the mighty Infrastructure-as-a-Service (IaaS), where players like Amazon, Microsoft, Google, and Rackspace provide a backbone that can be "rented out" by other companies. (For example, Netflix provides services to you because it's a customer of the cloud-services at Amazon.)

The use of Cloud Computing is increasing because of its mobility, huge availability of data and low cost. On the other hand it brings more threats to the security of the company's data and information.

In the recent years, data mining techniques have extensively used to discover knowledge in database and used in various fields: science and engineering business, medicine, spatial data etc.

The rising Cloud Computing trends provides for its users the unique benefit of extraordinary access to valuable data that can be turned into valuable insight that can help them to achieve their business objectives.

CLOUD COMPUTING ASPECTS

Basically cloud computing represents both the hardware and the software delivered as services over the Internet .The cloud Computing is an innovative notion that defines the use of computing as a utility, that has recently attracted significant attention. Cloud computing enables companies to consume and compute the resources as a utility -- just like electricity -- rather than having to build and maintain computing infrastructures in-house.

There are so many striking benefits for businesses and end users provided by Cloud computing. Three main benefits of cloud computing includes:

- Self-service provisioning: End users can turn up computing

resources for almost any type of workload on-demand.

- Elasticity: Companies can extend (scale up) and scale down as the computing needs increase and decrease respectively
- Pay per use: It allows users to pay only for the resources and workloads what they use as the computing resources are measured at a granular level. Cloud computing services can be private, public or hybrid.

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.

CHARACTERISTIC ABOUT DATA MINING

Data mining is widely used in different areas. There are several commercial data mining system available today and yet there are so many challenges in this field. Data mining represents finding useful patterns or trends through large amounts of data.

Data mining is defined as a “type or kind of database analysis that attempts to find out 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 find out previously unknown relationships among the data, especially when the data come from diverse databases.” [4]

(i) Data Mining Applications

Data mining is widely used in the following areas –

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Biological Data Analysis
- Other Scientific Applications
- Intrusion Detection

(ii) Financial Data Analysis

The financial data in banking and financial industry is generally reliable and of high quality which facilitates systematic data analysis and data mining. Some of the typical cases are as follows –

- Design and construction of data warehouses for multidimensional data analysis and data mining.
- The Loan payment forecast and customer credit policy analysis.
- Categorization and clustering of customers for targeted marketing.
- Detection of money laundering and other financial crimes.

(iii) Retail Industry

Data Mining is extensively used in Retail Industry because it collects huge amount of data from on sales, customer purchasing history, goods transportation, consumption and services. It is natural that the quantity of data collected will continue to expand rapidly because of the increasing ease,

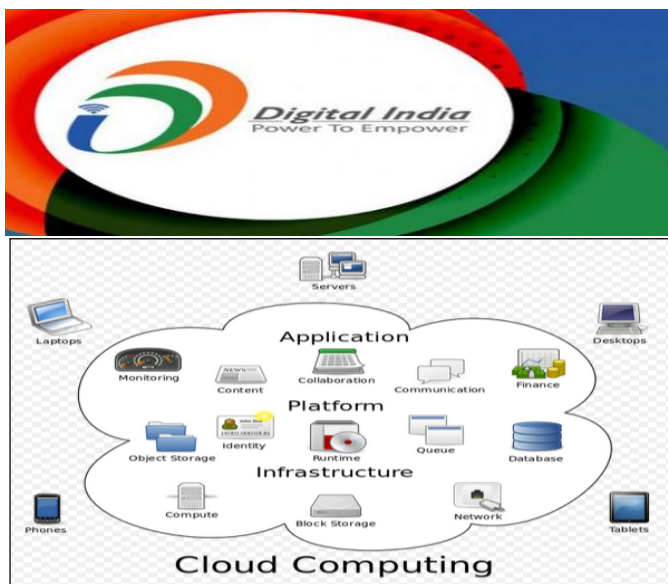


Figure 1: Computing the Cloud with Digital India

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

Company Name	Key Feature
Sun Microsystems Sun Cloud	More Available application than anyother open OS
IBM Dynamic Infrastructure	Integrated power management to help you plan, predict, Monitor and actively manage power consumption of Your Blade Center 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

Cloud computing represents all possible resources on the

availability and popularity of the web. Data mining in retail industry helps in identifying customer buying patterns and trends that lead to improved quality of customer service and good customer retention and satisfaction. Here is the list of examples of data mining in the retail industry –

- Design and Construction of data warehouses based on the benefits of data mining.
- Multidimensional analysis of sales, customers, products, time and region.
- Analysis of effectiveness of sales campaigns.
- Customer Retention.
- Product recommendation and cross-referencing of items.

(iv) Telecommunication Industry

Now a day's the telecommunication industry is one of the most rising industries providing various services such as fax, pager, cellular phone, internet messenger, images, e-mail, web data transmission, etc. Due to the development of new computer and communication technologies, the telecommunication industry is rapidly expanding. This is the reason why data mining is become very important to help and understand the business. Data mining in telecommunication industry helps in identifying the telecommunication patterns, catch fraudulent activities, make better use of resource, and improve quality of service. Here is the list of examples for which data mining improves telecommunication services –

- Multidimensional Analysis of Telecommunication data.
- Fraudulent pattern analysis.
- Identification of unusual patterns.
- Multidimensional association and sequential patterns analysis.
- Mobile Telecommunication services.
- Use of visualization tools in telecommunication data analysis.

(v) Biological Data Analysis

In recent times, we have seen a tremendous growth in the field of biology such as genomics, proteomics, functional Genomics and biomedical research. Biological data mining is a very important part of Bioinformatics. Following are the aspects in which data mining contributes for biological data analysis –

- Semantic integration of heterogeneous, distributed genomic and proteomic databases.
- Alignment, indexing, similarity search and comparative analysis multiple nucleotide sequences.

- Discovery of structural patterns and analysis of genetic networks and protein pathways.
- Association and path analysis.
- Visualization tools in genetic data analysis.

(vi) Other Scientific Applications

The applications discussed above tend to handle relatively small and homogeneous data sets for which the statistical techniques are appropriate. Huge amount of data have been collected from scientific domains such as geosciences, astronomy, etc. A large amount of data sets is being generated because of the fast numerical simulations in various fields such as climate and ecosystem modeling, chemical engineering, fluid dynamics, etc. Following are the applications of data mining in the field of Scientific Applications –

- Data Warehouses and data preprocessing.
- Graph-based mining.
- Visualization and domain specific knowledge.

(vii) Intrusion Detection

Intrusion refers to any type of action that threatens integrity or reliability, confidentiality, or the availability of network resources. In this world of connectivity, security has become the major issue. With increased usage of internet and availability of the tools and tricks for intruding and attacking network prompted intrusion detection to become a critical component of network administration. Here is the list of areas in which data mining technology may be applied for intrusion detection –

- To Develop the data mining algorithm for intrusion detection.
- Association and correlation analysis, aggregation to help select and build discriminating attributes.
- Analysis of Stream data.
- Distributed data mining.
- Visualization and query tools.

(viii) Data Mining System Products

There are so many data mining system products and domain specific data mining applications. The new data mining systems and applications are being added to the previous systems. Also, efforts are being made to standardize data mining languages.

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

Role of 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.[7]

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.

The use of data mining through Cloud computing reduces the difficulty that keep small companies from benefiting of the data mining instruments.

“Cloud Computing denotes the new development 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 permits the organizations to centralize software management and data storage, with guarantee of efficient, reliable and secure services for their users.” [6] By the accomplishment of data mining techniques through Cloud computing will permit the users to retrieve significant information from virtually integrated data warehouse that reduces the costs of infrastructure and storage.

CONCLUSIONS

Today the 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. 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. Cloud computing provides storage of data in a server by protecting data by using data mining concept. In Cloud computing the data is being shifted from one server to another server in a peer to peer transaction. 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 .this will help the people that government services are made available to citizens electronically by reducing paperwork.

REFERENCES

- [1] GH Forman, J Zahorjan - Computer, 1994 - doi.ieeecomputersociety.org
- [2] David P. Helmbold, "A dynamic disk spin-down technique for mobile computing", citeseer.ist.psu.edu, 1996
- [3] 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>.
- [4] Merriam-Webster Dictionary, "Definition of data mining", Link: <http://www.merriam-webster.com/dictionary/data%20mining>.
- [5] Landay, J.A. Kaufmann, T.R., "user interface issues in mobile computing", Workstation Operating Systems, 1993.
- [6] T Imielinski, BR Badrinath "mobile wireless computing, challenges in data management-Communications of the ACM, 1994 - portal.acm.org
- [7] Goss, S. (2013, April 10). Data-mining and our personal privacy. Retrieved from The Telegraph: <http://www.macon.com/2013/04/10/2429775/data-mining-and-our-personal-privacy.html>