

# Chapter 10

## A Study on the Relationship Between Cloud Computing and Data Mining in Business Organizations



Dilip Kumar Sharma, A. Dharmaraj, Alim Al Ayub Ahmed, K. Suresh Kumar, Khongdet Phasinam , and Mohd Naved

**Abstract** In order to enhance information services quality, data extraction is a method through which potentially helpful data may be extracted. Integrated with cloud computing, data mining technology enables customers to obtain valuable information from a data warehouse that lowers infrastructure and storage cost. In recent years the data mining sector and society as a whole have become quite attractive large data availability and urgent need for the transformation of such information into valuable expertise and data Market research, fraud prevention and retention of customers, manufacturing control and scientific research. The natural development of information technology may be seen as a consequence of data mining. Security and privacy of user data is of major issue in the usage of cloud computing in data mining. In the context of discovering new, legitimate, usable, and comprehensible data formats, data mining is regarded to be an essential activity. The integration of data mining techniques with cloud computing offers a flexible and scalable architecture that can be utilized to efficiently extract large quantities of information from

---

D. K. Sharma

Department of Mathematics, Jaypee University of Engineering and Technology, Guna, Madhya Pradesh, India

A. Dharmaraj

Karpagam Academy of Higher Education, Coimbatore, India

A. Al Ayub Ahmed

School of Accounting, Jiujiang University, 551 Qianjindonglu, Jiujiang, Jiangxi, China

K. Suresh Kumar (✉)

MBA Department, Panimalar Engineering College, Varadarajapuram, Nasarathpettai, Poonamallee, Chennai, India

e-mail: [pecmba19@gmail.com](mailto:pecmba19@gmail.com)

K. Phasinam

School of Agricultural and Food Engineering, Faculty of Food and Agricultural Technology, Pibulsongkram Rajabhat University, Phitsanulok, Thailand

M. Naved

Department of Business Analytics, Jagannath University, Delhi-NCR, India

near-connection data sources in order to provide valuable, decision-making knowledge. This article offers a general explanation of the necessity for data mining to be integrated into cloud computing, to provide its customers effective and secured services and to decrease infrastructure and operating requirements in business.

## 10.1 Introduction

The Internet has become an essential resource in our everyday activities in recent years, because the quantity of information generated by people who utilize online services is huge. In this data, hidden data may be utilized to make successful choices. Cloud infrastructure is utilized to substantially find valuable information in data mining integration techniques. Cloud computing is designed to alter the conventional computing method by delivering both physical assets and programming. These services are offered via the internet [1]. Its cheap cost, portability, and enormous availability make it more attractive. It offers limitless storage and processing power, resulting in huge numbers of data. For finding knowledge on databases, data mining techniques are employed. It is utilized for the analysis of data from various origins and the provision of relevant information. The Data Mining and the Cloud Computing Relationship Approach offers quick access to technology and an information finding system that consists of several decentralized data analysis services.

We are frequently called the information era in age. We think that information leads to power and success in that information era and that we have gathered enormous quantities of information due to advanced technology such as computers, satellites, and so on [2]. Initially, we began gathering and storing all kinds of data with the introduction of computers and methods for mass data knowledge relying on the computer's ability to assist sort this fusion of information. Sadly, these huge data sets, housed on separate structures, became overpowering very quickly. The early turmoil resulted in organized systems and database management systems being created (DBMS). We have now far more knowledge than we can deal with: from transactions, scientific knowledge, satellite photos, documentary evidence, and military intelligence. Data mining was an efficient technique for analysing data from many perspectives and obtaining valuable data insights. Classifying data, categorizing data and finding connection between datasets data patterns. Many companies are now beginning to use data mining as a technique to cope with the aggressive data analysis environment. The cloud allows you to access your information at all times from everywhere. The cloud eliminates the requirement being in the similar physical place as your gear [3].

## 10.2 Concept for Cloud Computing

The term “cloud computing” is a catchall term for a new sort of Internet-based computing. It's a unique concept that explains the use of computers as a service, and

it's gotten a lot of attention recently. Cloud computing, according to the National NIST is a concept that gives on-demand access to a shared pool of customizable computing assets (e.g. networks, servers, storage, applications, and services) that can be swiftly delivered and discharged with minimum administration effort or service provider engagement.

Cloud computing is an emerging concept change where computers are transferred from particular computers or servers to a "cloud" of computers. Cloud users must only consider the computer service requested, since the fundamental facts on how this is done are concealed. The high-performance computing technique is carried out by bringing together all computer resources and by software instead of a human being [4].

### 10.3 Cloud Computing Advantages

- Less computer charge: no high performance and affordable computer is needed for the use of web-based cloud computing apps.
- Enhanced Performance: Computers start and operate quicker in a cloud computing environment because they have less memory loaded processes and applications.
- Lower software expenses: you can obtain much of what you need free of charge instead of buying costly software programmes.
- Another benefit of cloud computing is that you are no longer confronted with the choice between outdated software and expensive upgrade fees. Updates occur automatically if the programme is on the Internet
- Cloud computing provides practically infinite storage capacity.
- Greater information reliability: In contrast to desktop computing, where computer accidents in the cloud do not affect data storage if the hard drive fails or loses all important data.

### 10.4 Cloud Computing Disadvantages

- It does not function well with low-speed connections.
- It needs a continuous internet connection.
- Information or data stored in cloud computing may not be secure.

### 10.5 Clouds Are Classified into Some Categories

Many firms share the services in a community cloud, and only those companies have access. Companies or a cloud service provider could own and control the technology [5].

**Fig. 10.1** Fundamental models for the provision of cloud services

Cloud Clients		
Web browser, mobile app, thin client, etc.		
SAAS	PaaS	IaaS
CRM, Virtual desktop, Email,	Webserver, Database, Development tool,....	Virtual machines, Servers, Storage,

A cloud provider owns and maintains a public cloud that is accessible through the Internet. Public services such as online photo storage, e-mail services, and social networking sites are examples. The term "hybrid cloud" refers to a system that includes different methods for sharing resources (for example, combining public and community clouds). A private cloud is a cloud infrastructure that is reserved for a single company and is managed by the business or a third party.

### 10.6 Models for Cloud Computing

- (i) **Service as a software (SaaS)** The client is given a whole application as a service over the internet under this approach. A single service instance is running on the cloud and many end users are being served [6]. On the client side, an initial investment in servers or software licenses is not necessary, whilst expenses for providers are reduced since just one application has to be hosted and maintained. Headquarters like Google, Sales force, Microsoft, Zoho, and others are offering nowadays SaaS (Fig. 10.1).
- (ii) **Platform as a service (PaaS)** Here is wrapped and provided a service that allows other higher levels of support to be built on software layers or development environments. The client may develop its own apps on the infrastructure of the provider.
- (iii) **Infrastructure as a services (IaaS)** As a standardized network service, IaaS offers basic storage and computer capacity. The workloads are pooled and made accessible via servers, storage systems, networking systems, network infrastructure space etc.

### 10.7 Companies in Cloud Top and Key Characteristics

Table 10.1 shows the cloud names with their key components.

**Table 10.1** Companies in cloud top and key characteristics

Cloud name	Key component
Sun microsystems sun cloud	More applications are available than on any other open operating system
Amazon EC2	The goal is to make web scale computing more accessible to developers
Microsoft	A discount plan for development accelerators is now available. Discounts ranging from 15 to 30% off consumption costs are available for the first six months
GoGrid cloud computing	Automatic load levelling as well as free, round-the-clock assistance
IBM dynamic infrastructure	Processing power in the context of relationships may assist you in planning, predicting, monitoring, and actively managing the power usage of your Blade Center servers
Google app engine	If you do not surpass the quota given, there is no time restriction on the testing period
AT&T synaptic hosting	You may either use completely on-demand architecture or mix it with specific components to fulfill particular needs

## 10.8 Overview of Data Mining

Data mining is the process of detecting undiscovered, meaningful patterns and correlations in massive data sets using sophisticated data and technology. These tools include statistical models, mathematical algorithms, and machine learning techniques (algorithms that improve their performance automatically over time, such as neural networks or decision trees) [7]. As a result, data mining comprises more than simply data collection and management; it also includes estimation and forecasting. Data mining is becoming more common in the business and government sectors. Data mining is commonly utilized in industries such as banking, insurance, medicine, and retailing to save money, improve research, and increase sales.

Data mining applications in the public organizations were originally employed to identify fraud and waste and were increasingly being utilized for measurement and improvement of programmed performance. Data Mining's primary purpose is to automatically discover patterns with little human input and effort. Data mining is a significant tool for managing future market trends and judgement. Data mining equipment and methods in many areas in different forms may effectively be used [8].

Data mining, also known as Knowledge Discovery in Databases, is the time-consuming process of extracting implicit, previously unknown, and potentially important information from databases (KDD). In databases, the terms "data extraction" and "knowledge discovery" are commonly interchanged (or KDDs), data mining is an important aspect of the knowledge discovery process. The diagram below depicts data mining as part of an iterative knowledge discovery process.

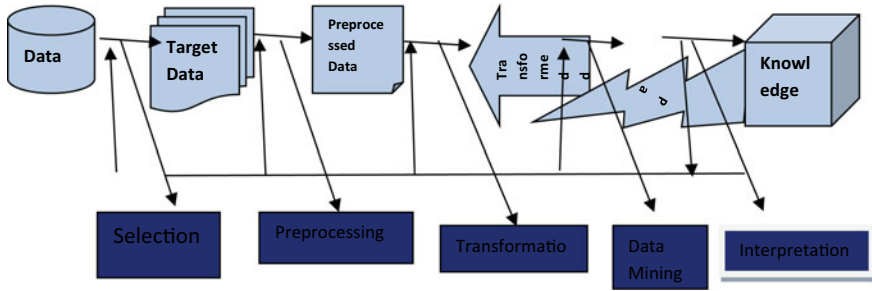


Fig. 10.2 Cloud computing and data mining in business organizations

### 10.9 Knowledge Discovery Process (KDD)

The several stages in the process of KDD.

- Data integration—Relationship information from various data sources.
- Data Selection and cleaning—Analytical data are collected from the database; noise is eliminated and inconsequential data is deleted.
- Data transformation—This phase includes consolidating and converting data to mining formats e.g. by aggregating data summaries.
- Mining of data—That’s the most essential phase and is achieved via smart training datasets.

7Evaluate patterns the most popular method for forecasting a particular result, such as a response, high/medium/low value client, that is susceptible to buying/not buying.

- On-assessment consists on identifying intriguing trends.
- Presentation of knowledge—different visualization and presentation methods are used to show the end user the gathered or mined information (Fig. 10.2).

### 10.10 Techniques for Data Mining

Table 10.2 discusses the Data Mining Techniques along with their key components.

### 10.11 Data Mining Applications

The following are the main areas of applications for data mining:

- (i) **Fraud** utilized for monitoring fraud by credit cards, monitoring millions of accounts. Fraud detection: They are used to detect financial transactions that may suggest money laundering.

**Table 10.2** Techniques for data mining

Data mining	Key component
Removal of feature	It develops new qualities as a continuous combination of current features. This technique's uses include text data, latent semantic analysis, data compression, data decomposition and projection, and pattern recognition
Clustering	This programme may be used to analyse data and find natural groupings of data. Inhabitants of one cluster are more similar to one another than residents of another cluster, which is a positive thing. The discovery of new consumer categories and the discovery of new life sciences are two examples
Identification of anomalies	The way occurrences deviate from the norm is used to describe them as strange or suspicious. Health-care fraud, expense-report fraud, and tax evasion are all instances of common fraud
Association	Criteria linked with commonly encountered products may be discovered and utilized for market basket analysis, cross-sell opportunities and root cause analysis. Product packaging, in-store positioning, and fault analysis are all possible with this tool
Regression	Lifetime value, home worth, and process yield rates are all examples of continuous numerical outcomes that may be predicted using this method
Significance of attribute	The characteristics are ranked in order of their strength of connection with the target attribute. Examples of applications include identifying the variables that are most closely linked with consumers who react to an offer and the factors that are most closely associated with healthy patients
Categories	When forecasting a particular result such as a response or no response, high, medium, or low value client, or whether someone will buy or not will purchase anything. The most frequently utilized method

- (ii) **Investment** Many businesses utilize investment data mining, although most do not disclose their methods. Investment: LBS Capital Management is one example. It utilizes portfolio management experts, neural networks, and genetic algorithms.
- (iii) **Marketing** Database marketing systems are the main application in marketing, analysing customer databases in order to identify and forecast various consumer groupings.
- (iv) **Telecommunications** the TASA provides methods for pruning, grouping, and sorting to improve the outcomes of a simple brute-force rule search. With versatile information extraction methods, interaction and iteration may be explored with several sets of found rules.

## 10.12 Data Mining and Cloud Computing Relationship in Business

Data mining techniques and applications are crucial in the cloud computing sector. The practice of organizing information from unstructured or semi-structured internet data sources is known as data mining [9]. Cloud computing data mining allows organizations to centralize software and data storage management, assuring trustworthy, efficient, and reliable services for their consumers. The use of data mining technologies such as SaaS, PaaS, and IaaS to collect data in cloud computing is examined. Cloud data mining is used to analyse and extract important data in a variety of industries, including banking, medical, and marketing [10]. With only a few mouse clicks, you can acquire all the information you need about customer behaviour, habits, interests, and location. Smaller businesses can use the cloud to employ a cloud service for efficient analysis of all data in the company that was previously solely reserved for big data [11].

When dealing with large volumes of data, Data Mining is preferred, and related approaches usually require large data sets to achieve classification accuracy. Data mining is used by cloud operators to improve client experience [12]. Customers may extract important information from virtually any data source using cloud computing data mining techniques, which reduces equipment and storage costs. Cloud Computing is a new trend in internet services that focuses on job-processing servers' clouds [13]. Data mining is the technique of obtaining unstructured or semi-structured data from internet data sources in cloud computing. Cloud computing is the delivery of software and hardware as a service via the Internet, and cloud computing data mining technology is no exception [14].

The following are the advantages for data mining and cloud computing.

- The client pays just for the data mining tools he requires.
- The client does not have to keep a physical server since he may use the browser to do data mining.
- Inefficient storage capacity.
- Virtual computers can be launched shortly.
- No relational database queries.
- Communication channel queue.

## 10.13 Conclusion

This study provides a high-level overview of data mining. Relationships with cloud computing service providers are critical for companies in order to make effective decisions when predicting current performance and patterns. Computing is the component that is given, and data mining is the component that is offered as well. Data mining cannot be completed without the use of a Data Mining software, as well as without the



use of cloud computing or cloud computing services. They are all too tasty and effective when combined, just like cake and ice cream are when served separately. Cloud computing is dependent on computer systems that are capable of accepting tasks from a distance. Obtaining organized data from unappropriated or semi-structured Web data sources is accomplished via the process of data mining. Cloud Computing data mining allows businesses to centralize software and data storage management, thus providing customers with more cost-effective and reliable services.

## References

1. Ambulkar, B., Borkar, V.: Data mining in cloud computing. In: MPGI National Multi Conference, vol. 2012 (2012)
2. Dillon, T., Wu, C., Chang, E.: Cloud computing: issues and challenges. In: 2010 24th IEEE International Conference on Advanced Information Networking and Applications, pp. 27–33. IEEE (2010)
3. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. *AI Mag.* **17**(3), 37–37 (1996)
4. Feyyad, U.M.: Data mining and knowledge discovery: making sense out of data. *IEEE Expert* **11**(5), 20–25 (1996)
5. Jain, A., Pandey, A.K.: Multiple quality optimizations in electrical discharge drilling of mild steel sheet. *Mater. Today Proc.* **4**(8), 7252–7261 (2017)
6. Jain, A., Pandey, A.K.: Modeling and optimizing of different quality characteristics in electrical discharge drilling of titanium alloy (grade-5) sheet. *Mater. Today Proc.* **18**, 182–191 (2019)
7. Jain, A., Yadav, A.K., Shrivastava, Y.: Modelling and optimization of different quality characteristics in electric discharge drilling of titanium alloy sheet. *Mater. Today Proc.* **21**, 1680–1684 (2020)
8. Kamala, B.: A study on integrated approach of data mining and cloud mining. *Int. J. Adv. Comput. Sci. Cloud Comput.* ISSN 2321-4058 (2013)
9. Mell, P., Grance, T.: The NIST definition of cloud computing (2011)
10. Panwar, V., Sharma, D.K., Kumar, K.P., Jain, A., Thakar, C.: Experimental investigations and optimization of surface roughness in turning of en 36 alloy steel using response surface methodology and genetic algorithm. *Mater. Today Proc.* (2021)
11. Patil, V., Nikam, V.B.: Study of data mining algorithm in cloud computing using MapReduce framework. *J. Eng. Comput. Appl. Sci. (JEC&AS)* **2**(7), 65–70 (2013)
12. Petre, R.S.: Data mining in cloud computing. *Database Syst. J.* **3**(3), 67–71 (2012)
13. Simmon, E.: Evaluation of cloud computing services based on NIST SP 800–145. *NIST Spec. Publ.* **500**, 322 (2018)
14. Viloría, A., Acuña, G.C., Franco, D.J.A., Hernández-Palma, H., Fuentes, J.P., Rambal, E.P.: Integration of data mining techniques to PostgreSQL database manager system. *Procedia Comput. Sci.* **155**, 575–580 (2019)