Starring role of Data Mining in Cloud Computing Paradigm

C.Edward Jaya Singh
Assistant Professor, Nesamony Memorial Christian College, Marthandam, Tamilnadu, 629161, India

Dr.E.Baburaj
Professor, Narayanaguru Siddhartha College of Engg. & Technology, Padanthalumoodu, Tamilnadu, India

Abstract

The ideal support of Information and communication technology leads to the enlargement of Big Data processing mechanisms like Data Mining. It is an exercise of extracting concealed as well as valuable information from raw Data. Today, with the rapid growth of the Information Technology the size of the data has been increased from KB level PB level. The objective of data mining process is also additional and more problematical, so the data mining algorithms are needed to be more competent. Cloud computing paradigm can provide the infrastructure to gigantic and multifaceted data of data mining, as well as innovative challenging issues for data mining. The cloud computing researches are materialized. This Script deals with the study of how data mining key features are used in cloud computing and also converse about the basic concept of cloud computing services and the role of data mining algorithms for the effectiveness and sketches out how data mining is recycled in cloud computing paradigm.

1. Introduction

With the rapid growth of processing and storage technologies and the accomplishment of the Internet, computing resources have become cost effective, more authoritative and more collectively available than ever before. This technological propensity has enabled the awareness of a new and innovative computing model called Cloud Computing. NIST definition of cloud computing: Cloud computing is a model for enabling 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. Today, the cloud computing plays a vital role and undertaking broad changes in the way IT services are designed, delivered, consumed, and managed. The boom in cloud computing over the past few years has led to a situation that is common to many innovations and new technologies: “Cloud computing” was coined for what happens when applications and facilities are moved into the “cloud” World. Popularity of cloud computing is increasing day by day in distributed computing environment. There is a growing trend of using cloud environments for storage and data processing needs. To use the full potential of cloud computing, data is transferred, processed, retrieved and stored by external cloud providers. Data owners are very skeptical to place their data outside their own control sphere. Their main concerns are the confidentiality, integrity, security and methods of mining the data from the cloud.

2. Data Mining Services

Data Mining refers to extracting or “Mining” knowledge from huge volumes of scattered data. It is a dynamic process where intelligent methods are applied in order to extract Data Patterns. Many other terms carry a similar or slightly different meaning to data mining, such as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. The KDD as a process consists of an iterative sequence of the following steps:

- Data cleaning (to remove noise and inconsistent data)
- Data integration (where multiple data sources may be combined)
- Data selection (where data relevant to the analysis task are retrieved from the database)
- Data transformation (where data are transformed or consolidated into forms appropriate for mining by performing summary or aggregation operations, for instance)
- Data mining (intelligent methods are applied in order to extract data patterns)
- Pattern evaluation (to identify the truly interesting patterns representing knowledge)
3. Few Aspects Regarding Data Mining

Data mining characterizes finding useful hidden patterns or trends through bulky amounts of data. Data mining is defined as a type of distributed, relational or object oriented database analysis that attempts to discover useful patterns or relationships in a group of data. Data mining in Cloud Computing: Data mining techniques and applications are very much required in the cloud computing paradigm. The data mining role in Cloud Computing allows the organizations to centralize the management of software and data storage, with declaration of resourceful, reliable and more protected.

4. Primary Cloud Model

A layered model of cloud computing environment can be divided into 4 layers such as Hardware or Datacenter layer, Infrastructure layer, Platform layer and Application layer.

There are different types of clouds, each with its own benefits and drawbacks. Public clouds: A cloud in which service providers offer their resources as services to the general public. It offers several key benefits to service providers, including no initial capital investment on infrastructure and shifting of risks to infrastructure providers. However, public clouds lack fine-grained control over data, network and security settings, which hampers their effectiveness in many business scenarios. Private clouds: Private clouds are designed for exclusive use by a single organization. A private cloud offers the highest degree of control over performance, reliability and security. Hybrid clouds: A hybrid cloud is a combination of public and private cloud models that tries to address the limitations of each approach. In a hybrid cloud, part of the service infrastructure runs in private clouds while the remaining part runs in public clouds. Hybrid clouds offer more flexibility than both public and private clouds. Virtual Private Cloud: An alternative solution to addressing the limitations of both public and private clouds is called Virtual Private Cloud. A VPC is essentially a platform running on top of public clouds. The main difference is that a VPC leverages virtual private network (VPN) technology that allows service providers to design their own topology and security settings such as firewall rules. VPC provides seamless transition from a proprietary service infrastructure to a cloud-based infrastructure, owing to the virtualized network layer.
5. The Responsibility of Data Mining In Cloud

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. 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. These data mining tasks include: Analyze Key Influencers, Detect Categories, Fill From example, Forecast, Highlight Exceptions, Scenario Analysis, Prediction Calculator, and Shopping Basket Analysis.

The implementation of data mining techniques through Cloud computing will allow the users to retrieve carrying great weight in order to virtually integrated data warehouse that reduces the costs of infrastructure and storage.

6. Current Research Challenges In Cloud

Countless obtainable issues have not been entirely addressed, while new challenges keep emerging from various applications. This section describes some of the challenging research issues in cloud computing for researchers who are much interested in this field.

6.1 Automated service provisioning

One of the primary key features of cloud computing is the capability of acquiring and releasing resources on-demand. The objective of a service provider in this case is to allocate and de-allocate resources from the cloud to satisfy its service level objectives while minimizing its operational cost.

6.2 Virtual machine migration

Virtualization can provide significant benefits in cloud computing by enabling virtual machine migration to balance load across the data center. In addition, virtual machine migration enables robust and highly responsive provisioning in data centers. Now, detecting workload hotspots and initiating a migration lacks the agility to respond to rapid workload changes.

6.3 Server consolidation

Server consolidation is an effective innovative approach to take full advantage of resource utilization while minimizing energy consumption in a cloud computing environment. Existing virtual machine migration technology residing on multiple servers onto a single server at that time the remaining servers can be set to an energy-saving state. The problem of optimally consolidating servers in a data center is often formulated as a variant of the vector bin-packing problem which is an NP-hard optimization problem.

6.4. Energy management

Improving energy efficiency is another major issue in cloud computing. Infrastructure providers are under enormous pressure to reduce energy consumption. Designing energy-efficient data centers has recently received considerable attention. This problem can be from several directions such as Energy efficient hardware architecture, Energy-aware job scheduling and server consolidation. In this admiration, the minority researchers have recently started to investigate coordinated solutions for performance and power management in a active cloud environment.

6.5. Traffic management and analysis

Examination of data traffic is significant for today’s data centers in cloud environment. Network operators also need to know how traffic flows through the network in order to make many of the management and planning decisions. Currently, there is not much work on measurement and analysis of data center traffic.

6.6. Data security

Data security is another important research topic in cloud computing. Since service providers typically do not have access to the physical security system of data centers, they must rely on the infrastructure provider to achieve full data security. The hardware layer must be trusted using hardware TPM. Secondly, the virtualization platform must be trusted using secure virtual machine monitors. VM migration should only be allowed if both source and destination servers are trusted. Recent work has been devoted to designing efficient protocols for trust establishment and management.

www.ijsret.org
6.7. Storage technologies and data management

These file systems are different from traditional distributed file systems in their storage structure, access pattern and application programming interface. In particular, they do not implement the standard POSIX interface, and therefore introduce compatibility issues with legacy file systems and applications. Several research efforts have studied this problem but not yet optimum.

6.8. Novel cloud architectures

Now, most of the commercial clouds are implemented in large data centers and operated in a centralized fashion. Although this design achieves economy-of-scale and high manageability, it also comes with its limitations such high energy expense and high initial investment for constructing data centers. Recent researchers suggest that small size data centers can be more advantageous than big data centers in many cases: a small data center does not consume so much power, hence it does not require a powerful and yet expensive cooling system; small data centers are cheaper to build and better geographically distributed than large data centers. The researchers think to build Nano-Data centers with full pledged manner.

The data mining technologies provided throughout Cloud computing is an extremely essential trait for today’s businesses to make proactive, knowledge driven decisions, as it helps them have future trends and behaviors predicted. This chapter 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 aptitude of integrating them in cloud computing becomes progressively stringent. The current technologies are not matured enough to realize its full potential. In future the researchers highly deliberate the data mining technologies to implement in fog computing.

7. Conclusion

Popularity of cloud computing is increasing day by day in distributed computing environment. There is a growing trend of using cloud environments for storage and data processing needs. To use the full potential of cloud computing, data is transferred, processed, retrieved and stored by external cloud providers. Data owners are very skeptical to place their data outside their own control sphere. Their main concerns are the confidentiality, integrity, security and methods of mining the data from the cloud.

References


