An Efficient Approach For Integrating Data Mining Into Cloud Computing

[1]N. Janardhan   [2]T. Sree Pravallika   [3]Sowjanya Gorantla
[1,2,3]*Department of Computer Science and Engineering,*
K L University, Guntur, Andhra Pradesh, India

ABSTRACT

Cloud Computing and Data Mining are the two emerging trends now a days in the world of information technology and computing. This paper deals with the study of how data mining is used in cloud computing. Data Mining is a process of extracting potentially useful information from raw data, and this paper also gives a glimpse on how SaaS is very useful in cloud computing and how can we apply the Saas for the Data Mining. The integration of data mining techniques into normal day-to-day activities has become common place. We are confronted daily with targeted advertising, and businesses have become more efficient through the use of data mining activities to reduce costs. 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. When processors become available, the data is moved to the processors.

Keywords

Data Mining, what is cloud Computing, how data mining are used in cloud computing, SaaS, key Characteristics of SaaS.

1. INTRODUCTION

Data mining, the extraction of hidden predictive information from large databases, is a powerful new Data 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. As data sets have grown in size and complexity, direct hands on data analysis has increasingly been augmented with indirect, automatic data processing. This has been aided by other discoveries in computer science, such as neural networks, cluster analysis, genetic algorithms (1950s), decision trees (1960s) and support vector machines (1990s). Data mining is

the process of applying these methods to data with the intention of uncovering hidden patterns in large data sets. Data mining is sorting through data to identify patterns and establish relationships.

1.1 DATA MINING PARAMETERS

1..*Association* - Looking for patterns where one event is connected to another event.
2.*Sequence or path analysis* - Looking for patterns where one event leads to another later event
3.*Classification* - Looking for new patterns
4.*Clustering* - Finding and visually documenting groups of facts not previously known
5. *Forecasting* - Discovering patterns in data that can lead to reasonable predictions about the future This area of data mining is known as predictive analytics.
E.g. Visual Numeric's has been providing advanced forecasting and data mining solutions across a wide range of industries such as aerospace, government, telecommunications, financial services and healthcare. Visual Numeric's' forecasting solutions combine technical expertise, decades of hands-on experience and powerful products to create the highest quality solutions possible for your visual data analysis needs. Same as, so there are various applications of data mining in real world. So there are many applications of Data mining in real world As, Hospital, Student Management, Airline Reservation, Forecasting, Biometrics, Mathematics, Geographical, Web Mining, Parallel Processing, Space Organization, Data Integrity, etc. So there are many application in which the data mining term is very useful. So from many application we will be discussing on cloud computing, how data mining is used in cloud computing.

2. 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. In early 2008, OpenNebula, enhanced in the RESERVOIR European Commission-funded project, became the first open-source software for deploying private and hybrid clouds, and for the federation of 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. Cloud computing really is accessing resources and services needed to perform functions with dynamically changing needs. An application or service developer requests access from the cloud rather than a specific endpoint or named resource.
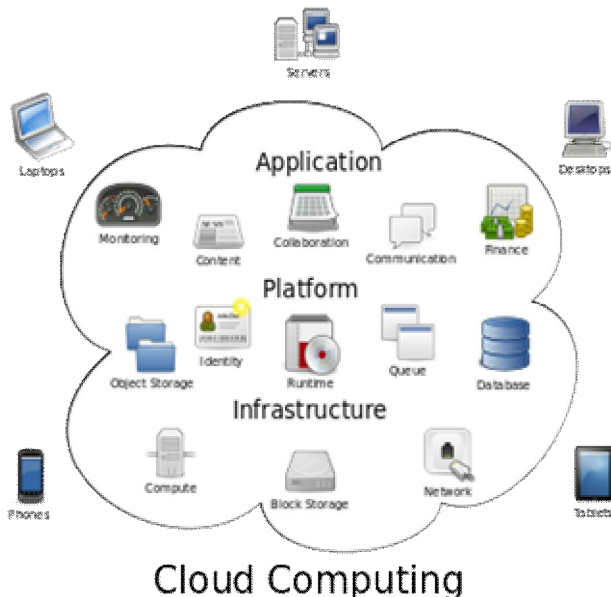

Fig.1. Cloud Computing Sample Architecture

2.1 Frameworks provide mechanisms

- ⚔ Self-healing
- ⚔ Self monitoring
- ⚔ Resource registration and discovery
- ⚔ Service level agreement definitions
- ⚔ Automatic reconfiguration

3. A GLANCE ON CLOUD SERVICES

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. Major players in this filed are Amazon's EC2, Google App Engine etc.

- ⚔ **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. Examples are Forece.com offered by Salesforce.com

- ⚔ **SaaS** : Delivers the application over the Internet or Intranet via a cloud Infrastructure. Built on underlying IaaS and PaaS Layer. Examples are the electronic mails that we are using today.

4. REDUCE DATA MINING COSTS BY SaaS – CLOUD MINING IS BORN

The SaaS Distribution model (Software-as-a-Service) helps to reduce costs by providing flexible license options and outsourcing the hardware effort. At SaaS, the software is not applied in the company, it lies at a software service providers server. That means the provider deals with the hardware, looks after software updates and maintains technically everything.

In Cloud Mining, the servers that provide the software are the Cloud. This can be the public cloud from Google, Amazon.com etc, or a private cloud on the servers of a single provider. That has two main effects; on one hand the customer only pays for the tools of Data Mining he needs. That makes him save a lot compared to complex Data Mining suites that he is not using exhaustive.

And on the other hand he just pays for the costs that are generated by using the Cloud. He does not have to maintain a hardware infrastructure, he can apply data mining just via his browser. This reduces the barriers that keep small companies from benefiting of Data Mining.

4.1 KEY CHARACTERISTICS OF SaaS

1. *Centralized feature updates :* This obviates the need for downloadable patches and upgrades typical of an on-premise software installation.
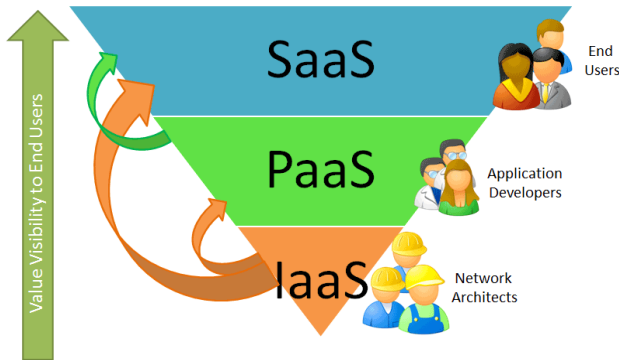

Fig.2. Cloud Services and Users

2. Single-instance, multi-tenant architecture: A one-to-many model implies a single physical instance with customers hosted in separate logical spaces. There can be multiple variations of how a single instance really gets implemented and how multi-tenancy really gets achieved.

3. Managed centrally and accessed over the Internet: There is no software component installed at the customer site. All applications can be accessed remotely over the web.

4. Generally priced on a per-user basis: The minimum number of users that companies can sign up for varies from one SaaS vendor to another and also depends on what stage the SaaS vendor is in their evolution path as a company. Some do charge additional fees for extra bandwidth and storage.

5. Mostly subscription-based, no upfront license

costs: This implies that functional leaders (from sales, marketing, HR and manufacturing) do not have to go through their IT department to get them approved.

Fig.3. Cloud computing Characteristics

4.2 KEY DRIVERS BEHIND ADOPTON

- ⚑ Reliability and popularity of web applications

- ⚑ Security is sufficiently well-trusted

- ⚑ Improved network bandwidth.

- ⚑ Low cost of ownership

5. CLOUD MINING BY LAYERED TECHNOLOGIES

Layered Technologies is a leading global provider of managed dedicated hosting, on-demand virtualization/cloud computing and Web services. By providing high-quality technology, infrastructure and support, Layered Tech enables customers to eliminate capital expenses and save on operating costs while focusing on core business issues. Layered Tech's scalable infrastructure powers millions of sites and Internet-enabled applications including e-commerce, software as a service (SaaS)[5] , content distribution and more. Our clients range from leading-edge Web 2.0 startups, successful mid-sized enterprises and some of the world's largest consultancy and integration firms. Jan. 11, 2010 - Layered Technologies, Inc. (Layered Tech), a leading worldwide provider of on-demand IT infrastructure, has developed a new virtual private data center (VPDC) platform with levels of managed services, security and flexibility via a proprietary API that were previously unavailable in an integrated offering. The new platform is a hybrid cloud computing infrastructure that gives customers a virtualized environment on dedicated servers within Layered Tech data centers, as well as levels of flexibility on how to securely access their VPDC, whether by dedicated lines, VPN or Internet. Layered Tech's new integrated virtualized platform gives customers the reliability of dedicated servers with the high availability, processing power and scalability of virtual machines to meet constantly changing business needs. The customer API, created by Layered Tech and based on industry standard protocols (SOAP and XML-RPC), provides easy connectivity and enables customers to perform activities such as customizing proprietary applications, monitoring and managing resources, reviewing analytics, and more via computer or mobile device. A new Virtual Private Data Center (VPDC) platform from Layered Technologies. The provider of on-demand IT infrastructure tries to overcome users concerns if their data is safe by having a enterprise level security standart. It will be possible to "design, order and deploy a

secure virtualized environment within an hour". Customers can choose whatever plattform they need, including 3Tera's AppLogic, VMware and Microsoft Hyper-V.

6. DATA MINING IN THE CLOUD

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. [1] 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

7. CONCLUSION AND FUTURE WORK

Data mining is used in various applications such as Health care, Student management, mathematics, Science, in various website. 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. Here we explore the how the data mining tools like SAS, PAS and IaaS are used in cloud computing to extract the information.

A cloud provider for a data mining and natural language processing system. Leading cloud computing providers Amazon Web Services, Windows Azure, OpenStack.

People use this feature to build information listing, get information about different topics by searching in forums etc. Companies use this service to see what kind of information is floating in the world wide web for their products or services and take actions based on the data presented

The information retrieval practical model through the multi-agent system with data mining in a cloud computing environment has been proposed. It is however, recommended that users should ensure that the request made to the IaaS is within the scope of integrated data warehouse and is clear and

simple. Thus, making the work for the multi-agent system easier through application of the data mining algorithms to retrieve meaningful information from the data warehouse. cloud computing allows the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage.

8. REFERENCES

[1] Building Data Mining Applications for CRM [Paperback] by Alex Berson (Author), Stephen J.Smith (Author), Berson (Author), Kurt Thearling (Author).

[2] Data-Driven Marketing: The 15 Metrics Everyone in... by Mark Jeffery.

[3] Cloud Computing with the Platform By Roger Jennings Windows Azure.

[4] Moving To The Cloud: Developing Apps in the New World of Cloud Computing, By Dinkar Sitaram, Geetha Manjunath

[5] The Cloud Computing Handbook - Everything You Need to Know about Cloud Computing, By Todd Arias.

[6]http://searchsqlserver.techtarget.com/definition/data-mining

[7]http://www.ijcaonline.org/volume15/number7/p   c387 2623.pdf

[8]http://www.waset.org/journals/waset/v39/v39-72.pdf

[9]http://www.estard.com/data_mining_marketing/data mining_campaign.asp

[10]http://dssresources.com/books/contents/berry97.html

[11]http://www.marketingprofs.com/articles/2010/3567/the -nine-most-common-data-mining-techniques-used-   in-predictive-analytics