International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 4 | ISSN : 2456-3307

Review of Clustering Technique using SaaS on the Cloud

Kanagalakshmi V, Ruby Gnanaselvam
Bishop Appasamy College of Arts and Science, 129,Race Course Road,Coimbatore, Tamil Nadu, India

ABSTRACT

Cloud computing is a general term for the delivery of hosted services over the internet. Software as a service (saas) it one of the most important in cloud computing . Then it can be used for different types of business solution. In real world variety organizations had successfully adapted the SaaS concept. In order to analyze the benefits' of SaaS. It has to be estimated properly. The existing SaaS evaluation models are only focusing on quality attributes of software. similar to conventional software services. But the SaaS on the cloud needs to be considered related characteristics of cloud. For this purpose the new estimate model is proposed based on Data mining techniques of clustering. This paper focus on review of the clustering technique service provider and service users in estimate SaaS on the cloud.

Keywords: Cloud Computing, Data Mining, Clustering Methods, Saas.

I. INTRODUCTION

CLOUD COMPUTING is a type of Internet-based computing that provides shared computer processing resources and data to computers and other devices on demand. It means the ability to run many computer at this same time ,is a model for enabling ubiquitous, on-demand access to a shared pool of configurable computing resources (e.g., computer networks, servers, storage, applications and services), which can be rapidly provisioned and released with minimal management effort. Cloud computing some basic characteristics are On-Demand Usage , Ubiquitous Access, Multi-tenancy (Resourcing Pooling) Elasticity (and Scalability) ,Measured Usage, Resiliency, Reliability , Scalability ,Cost. Cloud providers and cloud consumers can assess these characteristics individually and collectively to measure the value offering of a given cloud platform. Although cloud-based services will inherit and exhibit individual characteristics to varying extents, usually the greater the degree to which they are supported and utilized, the greater the resulting value proposition. The important of this that it removes the normal bass lessand costs of using business software. Software is synonymous with complexity and and cost development. SaaS is a software delivery method to provide access to software and its function remotely as web based services. Software as a service (saas)allow organization to access business functionality at typically less than paying for license application since SaaS pricing is based a monthly fee. Also the because the software is hosted remotely , user don't need to investing additional hardware. The characteristics of SaaS are as follows:

- Network-based access to, and management of, commercially available software.
- Activities managed from central locations rather than at each customer's site, enabling customers to access applications remotely via the Web.
- Application delivery typically closer to a one-to-many model (single instance, multi-tenant architecture) than to a one-to-one model, including architecture, pricing, partnering, and management characteristics.
- Centralized feature updating, which obviates the need for end-users to download patches and upgrades.

II. LITERATURE REVIEW

1. Data Mining

Data Mining is defined as the procedure of extracting information from huge sets of data. In other words, we can say that data mining is mining knowledge from data data mining as an essential step in the process of knowledge discovery. Here is the list of steps involved in the knowledge discovery process

- Data Cleaning
- Data Integration .
- Data Selection
- Data Transformation .
- Data Mining
- Pattern Evaluation
- Knowledge Presentation .

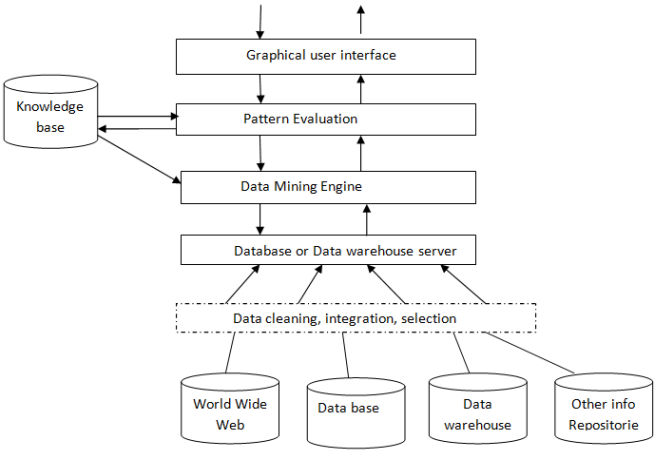

Figure 1: Architecture of data mining

KDD is known as a simplified process such as pre-processing, data mining and result of validation. Pre-processing is necessary to analyze multivariate data sets before data mining. The target set is then cleaned. Data Cleaning to remove noise and inconsistent data. Data mining involves seven common classes are include Anomaly detection, Association rules, Clustering, Classification, Regression, Summarization, Sequential pattern mining. The first step of Knowledge discovery from data is to verify that the patterns produced by data mining algorithms occur in the wide data set. A commonly used data mining technique is clustering. With classification of objects into different groups by portioning sets of data into a series of subset(cluster)

2. Clustering;

Cluster analysis or **clustering** is the task of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense or another) to each other than to those in other groups (clusters). Cluster is also called data segmentation in some applications because clustering partitions large data set into group according to their similarity. The following typical requirements of clustering in data mining:

- Scalability
- Dealing with different types of attributes
- Discovery of cluster with arbitrary shape
- Minimal requirements for domain knowledge to determine input parameters
- Ability to deal with noise and outliers
- High dimensionality
- Constraint based clustering
- Interpretability and usability

2. Types of Clustering

Clustering methods can be classified into the following categories

- Partitioning Method
- Hierarchical Method
- Density-based Method
- Grid-Based Method
- Model-Based Method
- Constraint-based Method

1) Partitioning Method:

Suppose we are given a database of 'n' objects and the partitioning method constructs 'k' partition of data. Each partition will represent a cluster and $k \leq n$. It means that it will classify the data into k groups, which satisfy the following requirements

- Each group contains at least one object.
- Each object must belong to exactly one group.

2) Hierarchical Method:

This method creates a hierarchical decomposition of the given set of data objects. We can classify hierarchical methods on the basis of how the hierarchical decomposition is formed. There are two approaches here

- Agglomerative Approach
- Divisive Approach

3) Agglomerative Approach:

This approach is also known as the bottom-up approach. In this, we start with each object forming a separate group. It keeps on merging the objects or groups that are close to one another. It keep on doing so until all of the groups are merged into one or until the termination condition holds.

4) Divisive Approach:

This approach is also known as the top-down approach. In this, we start with all of the objects in the same cluster. In the continuous iteration, a cluster is split up into smaller clusters. It is down until each object in one cluster or the termination condition holds. This method is rigid, i.e., once a merging or splitting is done, it can never be undone.

5) Density-based Method:

Clustering based on density (local cluster creation) . Such as density –connected points. It consist of major features like discover clusters of arbitrary shape then handle the noise and own scan.

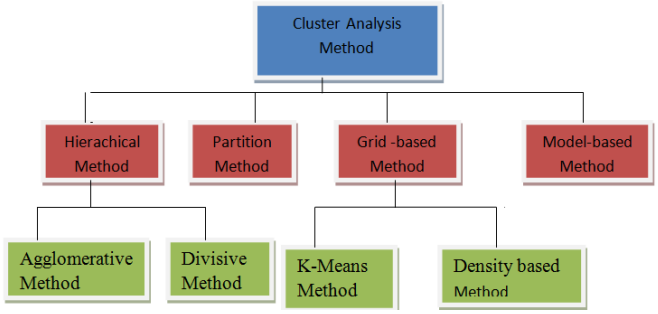

Figure 2: Methods of Clustering

6) Grid-based Method:

In this, the objects together form a grid. The object space is quantized into finite number of cells that form a grid structure. The major advantage of this method is fast processing time. It is dependent only on the number of cells in each dimension in the quantized space.

7) Model-based methods:

In this method, a model is hypothesized for each cluster to find the best fit of data for a given model. This method locates the clusters by clustering the density function. It reflects spatial distribution of the data points.This method also provides a way to automatically determine the number of clusters based on standard statistics, taking outlier or noise into account. It therefore yields robust clustering methods.

8) Constraint-based Method

In this method, the clustering is performed by the incorporation of user or application-oriented constraints. A constraint refers to the user expectation or the properties of desired clustering results. Constraints provide us with an interactive way of communication with the clustering process. Constraints can be specified by the user or the application requirement.

4. Trends of cloud computing

Every day cloud computing develop in most of the industry . They are some current cloud computing trends in change in cloud computing industry like Cmaas (Cloud monitoring as a service),Security, Hybrid Cloud Computing, consumption

5. Cloud service models

Cloud computing providers to perform some service are include: IaaS(infrastructure as a service), PaaS(Platform as a service),and SaaS(software as a service

Infrastructure as a service(IaaS)

Is a form of cloud computing that providers virtualized computing resources to over the internet (IAAS). TaaS provider the underlying operating systems, security , networking, and servers. For developing such applications, services, and for deploying development tools, database.

Platform as a service (PaaS)

It is one of the service in cloud computing , The providers as a platform allowing customer to develop and run manage the applications without complexity. Such as data base and web servers. PaaS platform are software development and management include Amazon web services, Google app engine.

Software as a service (SaaS)

The third service is one of the cloud computing. SaaS removes the need for organizations to install and run the application in own computers. SaaS benfits are include **Flexible payments**: Rather than purchasing software to install, or additional hardware to support it, customers subscribe to a SaaS offering. Generally, they pay for this service on a monthly basis using a pay-as-you-go model. Transitioning costs to a recurring operating expense allows many businesses to exercise better and more predictable budgeting. Users can also terminate SaaS offerings at any time to stop those recurring costs. **Scalable usage** : Cloud services like SaaS offer high scalability, which gives customers the option to access more, or fewer, services or features on-demand. SaaS is closely related to the ASP (application service provider) and on demand computing software delivery models. The hosted application management model of SaaS is similar to ASP: the provider hosts the customer's software and delivers it to approved end users over the internet. In the *software on demand* SaaS model, the provider gives customers network-based access to a single copy of an application that the provider created specifically for SaaS distribution. The application's source code is the same for all customers and when new features are functionalities are rolled out, they are rolled out to all customers. Depending upon the service level agreement (SLA), the customer's data for each model may be stored locally, in the cloud or both locally and in the cloud.

6. Importance of SaaS Evaluation:

SaaS providers it can be some of the evaluation are include. Security, Flexibility, Quality of service ,SLA , Global search . Security protecting the SaaS solution needs to be holistic. Another factor to consider is that many SaaS solutions involve multiple providers. There may be an Internet provider, a firewall provider and four or five others in the mix. When a problem occurs, there will be finger-pointing. It's the classical issue that on-premise IT has always had. There is no easy way to address this up front. The solution is to limit the number of vendors so they take more responsibility over performance rather than having the opportunity to point a finger in any direction when there are many vendors. Alternatively, buyers can work with companies like Savvis that manage these issues for them.

III. III.PROPSED MODEL

In the proposed model following pre-processing steps need to performed before applying model. Pre – processing is an important issue for both data warehousing and data mining as a real world data tend to to be complete .noisy and unconfident or irrelevant. The proposed model contains two steps to be followed SaaS data.

1) Extract SaaS data
2) Transform SaaS data
3) Load in RDBMS/MDBMS format
4) Apply clustering algorithm
5) Obtain appropriate cluster

1. Exatract SaaS data:

SaaS data are normally is not stored on any kind of data repository , ctive and efficient management then analysis of stream data poses a great challenge. So that only SaaS data from source dada needs to be extracted and separated from actual from source data.

2. Transform SaaS data:

Transformation is the method of dealing with inconsistencies and irrelevant and noisy with SaaS data after extracting SaaS data, SaaS data to transfor4med to common format and removed noisy, irrelevant, inconsistency and missing values.

3. Load in MDBMS format

Stream data are generated continuously in dynamic environment , with huge volumes, infinite flow and fast changing behavior. Most stream data represent low level information. Consisting of various kinds of detailed temporal and other features also used,

4. Appling clustering algorithm

The proposed model needs is applied after loading data stream in the multidimensional format and the model has to follow some systematic steps. Saas data are stored in the multidimensional formats, Then the MDBMS data is applied as input to various clustering

methods, each cluster key attribute to be compared as per given Saas data. Find out the gap,Such as a regional gaps, dollar gaps etc ,Then apply the clustering algorithm on given the SaaS data. After applying clustering algorithm cluster output will be generated in performance. Finally clustering analysis can be done in performance of response.

5. Obtain appropriate cluster

They applying clustering algorithms, clusters to formed based on software services provided on the cloud by service providers, These clusters will help to service user to make a decision suitable in requirement of software services.

IV.CONCLUSION

Clustering technique for estimating SaaS to will help to estimate potential software services on the cloud computing by using Data mining technique in clustering algorithms. The cluster model would be very helpful to software service providers to evaluate their own services to cloud users. they helps service provider to increase availability and scalability of software services on the cloud computing environment . They suitable for cloud users demand and requirement. It also help cloud users to evaluate. Potential software services available on the cloud computing environment.

V. REFERENCES

[1]. Clustering composite SaaS components in Cloud computing using a Grouping Genetic Algorithm conference Paper • June 2012 with 88 ReadsDOI : 10.1109/CEC.2012.6256562 Conference: Evolutionary Computation (CEC), 2012 IEEE Congress on

[2]. Ruxandra Stefania PETRE,"Data Mining in Cloud Computing" published in"Database Systems Journal vol.III, no.3/2012"

[3]. Xian Chen, Abhishek Srivastava and Paul Sorenson,"Towards a QoS-Focused SaaS Evaluation Model", published in CASCON2008.

[4]. Renu Asnani, A distributed k-mean clustering algorithm for cloud data mining, Dec 2015.

[5]. Anjali J Rathod, Daneshwari I Hatti, Analysis of Clustering Algorithms for Data Management in Sensor-Cloud , July 2015 .

[6]. Hao Yuan, Changbing Li, Research on Fuzzy Clustering Method for cloud Computing Task Scheduling, 2016.

[7]. Esha Sarkar, C.H Sekhar, Organizing Data in Cloud using Clustering Approach, May 2014 . 8 Dibya Jyoti Bora,D r. Anil Kumar, A Comparative Study Between Fuzzy Clustering Algorithm

[8]. Jae Yoo Lee, Jung Woo Lee, Du Wan Cheunt, and Soo Dong Kim,"A Quality Model for Evaluating Software-as-aService in Cloud Computing", published in 2009 Seventh ACIS International Conference on Software Engineering Research, Management and Applications.

[9]. S. Kanimozhi,"Effective Constraint based Clustering Approach for Collaborative Filtering Recommendation using Social Network Analysis", in Bonfring International Journal of Data Mining, Vol. 1, Special Issue, December 2011.

[10]. P. Botella, X. Burgues, J.P. Carvallo, X. Franch, G. Grau, J. Marco, C. Quer,"ISO/IEC 9126 in practice: what do we need to know?"

[11]. Soumita Seth, Debnath Bhattacharyya, Tai-hoon Kim. ,"Constraint-Based Clustering to Find Communities in a Social Network".

[12].