A PREMEDITATED CDM ALGORITHM IN CLOUD COMPUTING ENVIRONMENT FOR FPM

Kuldeep Mishra
M Tech scholar (CSE), SRGI, Jhansi, INDIA,

Ravi Rai Chaudhary
Asst. Prof (CSE), SRGI, Jhansi, INDIA,

Dheresh Soni
Assistant Professor (CSE), PCST, Bhopal, INDIA,

ABSTRACT

Cloud computing is a new concept with a broad definition. It has been acknowledged as one of the prevailing models for providing IT capacities. Cloud computing is a new network computing paradigm based on IP architecture, and its potential lies in new business applications. The computing paradigm that comes with cloud computing has incurred great concerns on the security of data, especially the integrity and confidentiality of data, as cloud service providers may have complete control on the computing infrastructure that underpins the services. The main task associated with cloud computing is next generation data center transformation. In this paper we generalized the formulation of data mining technique with cloud computing environment and generate the result. In data mining we want to find useful patterns with different methodology. The main issue with data mining techniques is that the space required for the item set and there operations are very huge. If we combine data mining techniques with cloud computing environment, then we can rent the space from the cloud providers on demand. This solution can solve the problem of huge space with better application usage in low cost. We can apply data mining techniques without taking any consideration of space. This paper basically survey and analyze the utility for solving the above situation.

Keywords: Cloud Computing, cloud storage, data mining, frequent pattern

I. INTRODUCTION

Cloud Computing [1,2] is a new business model. The term “Cloud computing” describes it as a system platform or a kind of software application. First, a system platform means, based on real
time, it can dynamically proviso, configure, re-configure and de-proviso a system. In a cloud computing platform, server is a physical server or a virtual server. High end cloud computing generally includes other computation resources.

Cloud computing is a new concept with a broad definition. Cloud computing is a new network computing paradigm based on IP architecture, and its potential lies in new business applications. For the majority of operators and enterprises, the main task associated with cloud computing is next generation data center transformation. "Computing" generally refers to computing application; that is, any IT application in industry or in the market. Because network technologies are being converged, all applications in information, communication, and video are integrated on a unified platform. Likewise, computing in cloud computing refers to any integrated application. The key characteristic of cloud computing is not "computing" but "cloud." It distributes the computing tasks to the resource pool constituted of a large number of computers, so that a variety of application systems can obtain computing power, storage space and a variety of software services on demand. The novelty of the Cloud Computing is that it almost provides unlimited cheap storage and computing power. This provides a platform for the storage and mining of mass data.

The role of data analytics increases in several request domains to cope with the big amount of captured data. Cloud computing adopt virtualization, service-oriented architecture, autonomic computing, and utility computing. Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility over a network. This will ensure cloud computing becomes more widespread among enterprises, institutions, organizations, and operators. Cloud computing not only provides traditional IT resource usage and application services, but also supports full resource usage and application services such as IT, communications, video, mobile, and Internet of Things using a converged network infrastructure. Cloud computing technologies include Key feature of unified fabric, unified virtualization, and unified computing system.

Cloud computing has become one of the key considerations both in academic world and industry. Cheap, apparently infinite computing resources that can be allocated approximately right away and pay-as-you-go pricing schemes are some of the reasons for the success of cloud computing. We discuss few aspects of cloud computing and also there area. We propose a novel approach which is cloud computing mapping and management through class and object hierarchy. In this approach we first design a cloud environment where we can analyze several object oriented aspects based on some assumptions. Then we deduce message passing behavior through a backup files based on the properties of object orient like class and object.

Association rule mining is an important research topic of data mining; its task is to find all subsets of items which frequently occur, and the relationship between them. Association rule mining has two main steps: the establishment of frequent item sets and the establishment of association rules. Apriori algorithm [3] is the most classic and most widely used algorithm for mining frequent item sets which generate Boolean association rules. The algorithm uses an iterative method called layer search to generate \((k + 1)\) item sets from the \(k\) item sets. In this paper we describe a new algorithm which provides the way for data mining or data mining association on cloud environment so that we can achieve a better way to handle a large amount of data.

In this paper we discuss several technical issues related to security concern. We provide here an overview of executing data mining services on grid. The rest of this paper is arranged as follows: Section 2 introduces Cloud Computing; Section 3 describes about data mining; Section 4 shows the Recent Scenario; Section 5 describes the proposed method. Section 6 describes Conclusion and future prospect.
II. CLOUD COMPUTING

Cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet). A Cloud is a type of parallel and distributed system consisting of a collection of interconnected and virtualized computers that are dynamically provisioned and presented as one or more unified computing resources based on service-level agreements established through negotiation between the service provider and consumers. The increased degree of connectivity and the increasing amount of data has led many providers and in particular data centers to employ larger infrastructures with dynamic load and access balancing.

A cloud service has three distinct characteristics that differentiate it from traditional hosting. 1. It is sold on demand, typically by the minute or the hour. 2. It is elastic - a user can have as much or as little of a service as they want at any given time 3. Service is fully managed by the provider (the consumer needs nothing but a personal computer and Internet access). There are several reasons to adopt cloud computing like cost, scalability, business agility, and disaster recovery. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or service provider interaction.

This cloud model promotes availability and is composed of:

1. Four deployment models:
   1.1 Private cloud
   1.2 Community cloud
   1.3 Public cloud
   1.4 Hybrid cloud

2. Three service models:
   2.1 Cloud Software as a Service (SaaS)
   2.2 Cloud Platform as a Service (PaaS)
   2.3 Cloud Infrastructure as a Service (IaaS)

3. Five essential characteristics:
   3.1 On-demand self-service
   3.2 Broad network access
   3.3 Resource pooling
   3.4 Rapid elasticity
   3.5 Measured Service

4. Key enabling technologies include:
   4.1 Fast wide-area networks
   4.2 Powerful, inexpensive server computers
   4.3 High-performance virtualization commodity hardware
In general, a public (external) cloud is an environment that exists outside a company’s firewall. It can be a service offered by a third-party vendor. It could also be referred to as a shared or multi-tenanted, virtualized infrastructure managed by means of a self-service portal.

A private (Internal) cloud reproduces the delivery models of a public cloud and does so behind a firewall for the exclusive benefit of an organization and its customers. The self-service administration interface is still in place while the IT infrastructure resources being collected are internal. In a hybrid cloud environment, external services are leveraged to extend or supplement an internal cloud.
III. DATA MINING TECHNIQUES

Here are several major data mining techniques have been developed and used in data mining projects recently including association, classification, clustering, prediction and sequential patterns. We will briefly examine those data mining techniques with example to have a good overview of them.

Association

Association is one of the best known data mining technique. In association, a pattern is discovered based on a relationship of a particular item on other items in the same transaction. For example, the association technique is used in market basket analysis to identify what products that customers frequently purchase together. Based on this data businesses can have corresponding marketing campaign to sell more products to make more profit.

Classification

Classification is a classic data mining technique based on machine learning. Basically classification is used to classify each item in a set of data into one of predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we make the software that can learn how to classify the data items into groups. For example, we can apply classification in application that “given all past records of employees who left the company, predict which current employees are probably to leave in the future.” In this case, we divide the employee’s records into two groups that are “leave” and “stay”. And then we can ask our data mining software to classify the employees into each group.

Clustering

Clustering is a data mining technique that makes meaningful or useful cluster of objects that have similar characteristic using automatic technique. Different from classification, clustering technique also defines the classes and put objects in them, while in classification objects are assigned into predefined classes. To make the concept clearer, we can take library as an example. In a library, books have a wide range of topics available. The challenge is how to keep those books in a way that readers can take several books in a specific topic without hassle. By using clustering technique, we can keep books that have some kind of similarities in one cluster or one shelf and label it with a meaningful name. If readers want to grab books in a topic, he or she would only go to that shelf instead of looking the whole in the whole library.

Prediction

The prediction as it name implied is one of a data mining techniques that discovers relationship between independent variables and relationship between dependent and independent variables. For instance, prediction analysis technique can be used in sale to predict profit for the future if we consider sale is an independent variable, profit could be a dependent variable. Then based on the historical sale and profit data, we can draw a fitted regression curve that is used for profit prediction.

Sequential Patterns

Sequential patterns analysis in one of data mining technique that seeks to discover similar patterns in data transaction over a business period. The uncover patterns are used for further business analysis to recognize relationships among data.
IV. RECENT SCENARIO

In 2010, Kawuu W. Lin et al. [4] proposed a set of strategies for many-task frequent pattern mining. Through empirical evaluations on various simulation conditions, the proposed strategies deliver excellent performance in terms of execution time.

In 2010, Yang Lai et al. [5] proposed a data mining framework on Hadoop using the Java Persistence API (JPA) and MySQL Cluster. The framework is elaborated in the implementation of a decision tree algorithm on Hadoop. We compare the data indexing algorithm with Hadoop MapFile indexing, which performs a binary search, in a modest cloud environment. The results show the algorithm is more efficient than naïve MapFile indexing. They compare the JDBC and JPA implementations of the data mining framework. The performance shows the framework is efficient for data mining on Hadoop.

In 2010, Jiabin Deng et al. [6] propose about the use of Power-law Distributions and Improved Cubic Spline Interpolation for multi-perspective analysis of shareware download frequency. The tasks include data mining the usage patterns and to build a mathematical model. Through analysis and checks, in accordance with changes to usage requirements, our proposed methods will intelligently adjust the data redundancy of cloud storage. Thus, storage resources are fine tuned and storage efficiency is greatly enhanced.

In 2011, Lingjuan Li et al. [7] proposed a strategy of mining association rules in cloud computing environment is focused on. Firstly, cloud computing, Hadoop, MapReduce programming model, Apriori algorithm and parallel association rule mining algorithm are introduced. Then, a parallel association rule mining strategy adapting to the cloud computing environment is designed. It includes data set division method, data set allocation method, improved Apriori algorithm, and the implementation procedure of the improved Apriori algorithm on MapReduce. Finally, the Hadoop platform is built and the experiment for testing performance of the strategy as well as the improved algorithm has been done.

In 2011, T.R. Gopalakrishnan Nair et al. [8] presents a specific method of implementing k-means approach for data mining in such scenarios. In this approach data is geographically distributed in multiple regions formed under several virtual machines. The results show that hierarchical virtual k-means approach is an efficient mining scheme for cloud databases.

In 2011, Lingjuan Li et al. [9] Focus on the strategy of mining association rules in cloud computing environment. Firstly, cloud computing, Hadoop, Map Reduce programming model, Apriori algorithm and parallel association rule mining algorithm are introduced. Then, a parallel association rule mining strategy adapting to the cloud computing environment is designed. It includes data set division method, data set allocation method, improved Apriori algorithm, and the implementation procedure of the improved Apriori algorithm on Map Reduce. Finally, the Hadoop platform is built and the experiment for testing performance of the strategy as well as the improved algorithm has been done.

In 2011, Fabrizio Marozzo et al. [10] present a Data Mining Cloud App framework that supports the execution of parameter sweeping data mining applications on a Cloud. The framework has been implemented using the Windows Azure platform, and evaluated through a set of parameter sweeping clustering and classification applications. The experimental results demonstrate the effectiveness of the proposed framework, as well as the scalability that can be achieved through the parallel execution of parameter sweeping applications on a pool of virtual servers.

V. PROPOSED ALGORITHM (CDM)

In this section we describe our proposed methodology that is cloud data mining (CDM). Our approach is easily understandable with the help of the flowchart [Fig. 3].
By the below flowchart we can easily understand the whole phenomena. In our approach we can enter in the cloud or non cloud environment. Then we can perform data mining technique to find frequent patterns and relevant associations. After the completion of data mining technique we can deduce the cost in cloud and non cloud environment. Our Proposed algorithm is shown below:

**Algorithm: CDM**

1. Create a new database acting as non cloud environment
2. Again create a new database acting as a cloud environment.
3. We can add file from non cloud or from the main database to the cloud environment when applying for the operation.
4. We can enter in the data mining environment either through the cloud environment or from the non cloud environment.
5. Read the database.
6. While (object.read()!=-1)
   
   {  
      [Start Reading]
   }
   
   [compute the occurrences]
   
   For i=1 to n iterations do
   
   {  
      Itemset[i]=count;
      Count++;
   }

7. Make different associations based on the conditions.
8. Enter the minimum support.

   [Enter the minimum support]
   
   If(min-sup= = msdb)
   
   {  
      Prune(db, key)
   }  
   
   Else  
   
   {  
      [Empty]  
   }

   Prune(db, key)
   
   enter the min-sup
   
   For i=1 to n do
   
   If(count[i]>min-sup)
   
   List=itemset[i];

   Else
   
   Remove from the list

9. Add the final result
10. Finish
VI. RESULT

When cloud computing introduced usage of cloud and its application is not so high. As the time passed and people comes to know the advantage of this technology the usage increases gradually and lot of application were also introduced in the cloud environment. Fig 4 Graph 1 shows the result and scenario of increasing application of different category in cloud environment which make cloud and its application easy for the users. The application which is developed in this dissertation when merged with cloud environment it also increases the application category in cloud environment. Fig. 5 Graph 2 shows the result in term of cost, application, and space in cloud environment. When the user need to develop some application he has to purchases all his required items like application software (required), storage, platform and database etc. This increases the cost but if the user use cloud environment and take services (IAAS, PAAS, SAAS) from cloud service provider the cost get reduced for the same application. Graph 2 shows that with the use of cloud user can use application in low cost, better on demand application usage and full storage space provide by cloud provider. Low cost means that user has to pay for the application which he use as per agreement. Application usage get better using this application means, generally 50 percent of cloud is need by provider from where he provide its service and remaining 50 percent is used by user. It is better in term of performance. In this dissertation we had also compared the cost in the cloud and non cloud environment considering only platform as a service as the role module in cloud computing.
We took the calculation of cloud computing in the following manner:

<table>
<thead>
<tr>
<th>Metric (on rent)</th>
<th>No. of resource</th>
<th>Cost in Rs per month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application Software</td>
<td>1</td>
<td>500 Rs per month</td>
</tr>
<tr>
<td>API for Application</td>
<td>1</td>
<td>100 Rs per month</td>
</tr>
<tr>
<td>Query (Load) for application</td>
<td>1</td>
<td>100 Rs per month</td>
</tr>
<tr>
<td>Storage</td>
<td>1GB</td>
<td>100 Rs per month</td>
</tr>
</tbody>
</table>

The final output is deduced by the below formula

Cloud Cost = (Application Cost + API + Query Load + Storage) * month

In the non cloud environment we consider the computation in the following manner:

<table>
<thead>
<tr>
<th>Metric</th>
<th>No. of resource</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Application Software</td>
<td>1</td>
<td>10 thousand Rs.</td>
</tr>
<tr>
<td>Database (Oracle)</td>
<td>1</td>
<td>20 thousand Rs.</td>
</tr>
<tr>
<td>Database (MS SQL Server)</td>
<td>1</td>
<td>20 thousand Rs.</td>
</tr>
<tr>
<td>Storage</td>
<td>100 GB</td>
<td>1 thousand Rs.</td>
</tr>
</tbody>
</table>
The final output is deduced by the below formula

\[
\text{Non Cloud Cost} = (\text{Application Software} + \text{Storage} + \text{Database}) \times (\text{man} \times \text{average salary}) \times \text{month}
\]

```
VII. CONCLUSION AND FUTURE PROSPECT

In this paper we want to generalize the formulation of data mining techniques with cloud computing environment. In data mining we want to find useful patterns with different methodology. The main issue with data mining techniques is that the space required for the item set and there operations are very huge. If we combine data mining techniques with cloud computing environment, then we can rent the space from the cloud providers on demand. This solution can solve the problem of huge space and we can apply data mining techniques without taking any consideration of space. This paper basically survey and analyze the utility for solving the above situation.
In future we concentrate on the real time scenario with their implementation and stress can be also given over profit maximization using data mining.

VIII. REFERENCES

```


[15] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns Without Candidate Generation. Proc. of ACM Int. Conf. on Management of Data (SIGMOD), 2000, pp. 1-12.

