International Journal of Research
Available at https://edupediapublications.org/journals
p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 03 Issue 18
December 2016

Analysis of Data in Cloud Computing Environments

Raghvendra Kumar[1], Dr. Prasant Kumar Pattnaik[2], Dr. Yogesh Sharma[3]
[1], *Ph.D Scholor, Jodhpur National University, Jodhpur, Rajsthan, India*
2, *School of Computer Engineering, KIIT University,Bhubaneswar, India,*
[3],*Jodhpur National University, Jodhpur, Rajsthan, India*
patnaikprasant@gmail.com, raghvendraagrawal7@gmail.com, yogeshsharma@gmail.com

***Abstract*—Data mining is extracting the information from large amount of data which stores in multiple horizontally and vertically partitioned databases. The information conveying the message direct or indirect. In this paper we combined the concept of association rule mining (Data Mining Techniques) with the cloud computing for analyzing the large amount of data base that is horizontally distributed in cloud environment. As we know that the cloud computing is upcoming techniques that offers the tremendous advantages in economical aspect, such as reduce the time of market, flexible computing capabilities and limitless computing power. To use the full potential of cloud computing the data is transferred, processed and stored by external cloud providers. Cloud computing has benefited the IT industries with less infrastructure investments and maintenance. As cloud provides the services like infrastructure as a service (IAAS), platform as a service (PAAS) and software as a service (SAAS) to its clients. The privacy is an essential service to provide in private and public cloud environments where the data can be easily hacked or tempered. In this paper aims to analyze the large amount by using the three main association rule mining techniques support, confidence and lift/importance and also provide the high privacy to all the cloud owner or provider using the hash function in cloud computing environments with zero percentage of data leakage and last in this paper we shows the comparison result of different privacy preserving techniques secure sum, secure subtraction, secure multiplication, secure union and secure intersection on the same horizontally partitioned database in cloud environments.**

Keywords— Data Mining, Cloud computing, Association rule mining, Data mining techniques, Association rule mining in cloud, Privacy preserving techniques.

I. Introduction

The process of extracting the information from the large amount of data is known as data mining [1]. Most of the people think that data mining is a synonym of knowledge discovery. But actually the data mining is one of the steps of knowledge discovery in the database. Knowledge discovery process includes the data cleaning, data selection, data integration, data transformation, data mining, pattern evaluation and knowledge presentation. Data mining has attracted the great deal of attention in information industry as well as business area of the need of turning large amount of data into the useful information [2]. Data mining is useful in an explanatory scenario in which there are no predefined notions about what will constitute an interesting outcome [3]. The database system industry has witnessed an evolutionary path in the development of the following functionality like data collection and data creation, data management and advanced a data analysis technique that includes the data warehousing and data mining. Prediction and description are considered as two main goals of data mining. Predictive data mining which produces the model of system described by the given data sets and descriptive data mining, which produces new, non trivial information based on the available data set. The main goal of prediction and description is to achieved through data mining takes such as classification, discovering association rules and clustering [4]. Cloud computing can be defined as the use of the computing resources that are delivered as a service over the network [5]. With traditional computing techniques, we run the software and store data on our computer systems. These files or data could be shared over the network. The importance of cloud computing lies in the fact that the software are not from our own computer but rather stored on the server and accessed through internet. Even if computer crashes, the software is still available for other user presents in the cloud environment. The concept of cloud computing comes from the clouds; a cloud can be considered a large group of interconnected computers which can be personal computers or network servers. They can be public cloud or private cloud. The cloud computing has spread rapidly through information technology industry and organizations. The ability of organizations to tap into computer application and other software via the cloud and thus free themselves from building and managing their own technology infrastructure seems potentially irresistible. In fact some companies providing cloud services have been growing at double digit rates despite the recent economic downfalls.

Cloud computing can be considered as a new technology to apply data mining. There are lots of data and unfortunately this huge amount of data is difficult to mine or analyze in term of computational resources. With the cloud computing concept the data mining and analysis can be more accessible and easy due to cost effective computational resources. Here in this paper we have discussed the usage of cloud computing platforms as a possible solution for mining and analysis of large amount of data with the high privacy. There are some advantages and disadvantages of using data mining with cloud computing. The main advantage is that the cloud computing combined with data mining can provided the powerful capacities of storage and computing and an excellent resources managements[6]. Due the explosive data grouth and amount of

computation involved in data mining, an efficient and high performance computation is very necessary for successful data mining application. Data mining in the cloud computing environments can be considered as the future of data mining because of the advantage of cloud computing concepts. Cloud computing provides the greater capabilities in data mining and data analytics [7]. The major concern about data mining is that the space required by the operations and item sets is very large. But if be merging the concept of data mining and cloud computing we can save a considerable amount of space [8]. And disadvantage of data mining with cloud computing, there are certain issue associated with data mining in the cloud computing. The major issue of data mining with cloud computing is privacy as the cloud provider has the complete control on the underlying computing infrastructure [8]. In this paper, special care has been taken so as to ensure the privacy of data under the cloud computing environment, when the cloud admin wants to analyze the huge amount of database.
Section1 introduce the brief introduction of the data mining in cloud computing environment and also explain the techniques of data mining and privacy preservation concept in cloud environment. Section2 introduces what are problem occurs when the data is distributed in cloud environment and data analyst want to analyze the distributed data. Section3 gives how the author is motivated to this research work from this work author explain the different privacy concepts in cloud environment. Section 4 describes the goal of this research work. Section 5 survey of analysis of data using association rule mining in cloud computing environment with privacy concept. Section6 shows the implementation and comparison result after using the different data analysis and privacy preserving techniques. Section7 a case study of market basket analysis in cloud environment. Section8 conclude and provides the direction for future research.

II. Problem Definition

Let i≥3 be the number of cloud nodes or cloud owner. Each cloud owner has their own private database DBi with the Ti transactions. We are considering three main data analysis techniques support (Si), confidence (Ci) and lift/ importance (Li) as percentages. The goal is to calculate the global support, global confidence and global lift with satisfying the thresholds values and without disclosing their private data or information to the entire cloud owner presented in the environments.

III. Motivation

The motivation behind dividing the input data into the portioned based database is to reduce the computation space by dealing with the smaller database that need to operated simulationasouly in the cloud environment. However we need to combine the intermediate result to calculating the global result. Using this concept (algorithm1) we can decreases the data analysis complexity, knowing that the time complexity of data mining processes is proportional to the size of the database.

IV. Objective

Cloud computing users work with data and applications that are often located in off-premises. How every many organization are uncomfortable with the having data and application on systems that they are not controlling. There are lack of knowledge on how cloud computing impact the privacy of data stored, proceeded and transmitted in cloud computing environment.

The main goal of this paper is to design an algorithm that clarify the impact of cloud computing privacy preservation.

By making a step by step recommendation

1. How data can be partitioned in the cloud environment
2. How data partitioning relate to privacy controls needed to preserve the privacy of the data
3. How to analyze the data with calculating the global support, confidence and lift without disclosing their own data to the other cloud nodes in the cloud environments.

V. Literature Survey

The association rule mining is one of the most important data mining techniques due to its wide applications. Most of the association rule mining algorithms are made for centralized database, where there are no external communications [9]. With the increased the size of data the computation time and memory requirements increase to great extent [10]. These difficulties have lead to parallel and distributed algorithms [11]. To accomplish this concept of dividing the database into the distributed database it to different nodes is used. There two main distributed database are discussed, vertical partitioned distributed database and horizontally portioned distributed database. Vertical partitioned distributed database using the idea of secure sum for secure calculation of inter-site, the sum of support degree of every sub items which are distributed in different sites is calculated. The item set is determined as global frequent item set if its support is greater than the threshold value. The approach of vertically portioned data mining has been extended to variety of data mining applications such as decision trees, SVM classification, Naïve Bayes classifier and K means clustering. And horizontally partitioned distributed database, the key idea is to find global frequent item sets, while ensuring no leakage of inter site information. It only calculates the secure sum of support degree inter sites. Thus the overall item sets support degree is acquired. The item sets with support degree greater than the threshold are the global frequent item sets. The association rule mining from the cloud can be done using sector/sphere framework [12]. Due to development of network and distributed paradigm the implementation of association rule algorithm (Apriori Algorithm) have been possible in cloud computing. The famous IT corporations such as google, Amazon and IBM have their different cloud computing architectures. Google has its google app engine which is composed the google file system, big table and map reduce

[13]. Amazon provides its cloud services by amazon web services, which contains the simple storage services, simple DB and elastic computing service [14]. A cloud based infrastructure to support data mining applications was developed. Which consists of storage cloud called sector and a compute cloud called sphere [15]. A improved apriori algorithm has been developed on map reduce which can handle vast amount of data [16]. Here large number of nodes are used based on the Hadoop platform. The various parallel algorithm for association rule mining faces problems for large inter site communication cost. This is due to large number of space required to maintain local count when the candidate's sets is large. To overcome this problem in the distributed association rule mining algorithms a de clustering approach is used to eliminates the inter communication cost between sites [17].

VI. Implementation and Result

There are Three physically distributed but logically connected Cloud nodes/ Resources

1. The Cloud Admin/ Analyst/Programmer, who wishes to perform aggregate data analytics over distributed datasets. Our main goal is to make easy to use for an average programmer who is not privacy or security expert.
2. The data owner, who owns more than one datasets, and would like to allow analysts to perform data analytics over the distributed datasets without compromising the privacy of the user in the data sets.
3. The service provider, who hosts the cloud services

We assume that the data owner, services provider and cloud admin are trusted, and every cloud nodes have their own data owner, data owner don't want to disclose their private information to other data owner presents in the environment, so that every cloud owner uses different privacy protocols to protect their private data sets. Algorithm1 shows the different privacy preserving protocols Secure Sum, Secure Subtraction, Secure Multiplication, Secure Union and Secure Intersection. And in this paper for data analysis in the cloud environment we used three main data mining techniques support, confidence and lift/importance. Figure 1 shows the arrangements of all the cloud nodes in the environment. Here in this paper the cloud admin are analyzing the global values of support, confidence and lift by using the different techniques and also concentrate on the privacy preservation. In figure shows how the cloud owner or cloud nodes are connected in the cloud environment. The cloud owners are connected with each other through the ring network. They are physically distributed but they are logically connected with each other. The support, confidence and lift/importance are calculated by using the following lemmas:

Lemma1: Shows the technique for calculating of support

Support=probe (X∪Y)/ Number of transaction≥ Minimum support

Probe (X∪Y) ≥ Minimum support*Number of Transaction

Probe (X∪Y) - Minimum support*Number of Transaction≥0

Lemma2: Shows the techniques for calculating of confidence

Confidence=probe(X∪Y)/ Probe(X)≥ Minimum confidence

Probe(X∪Y)≥ Minimum support*Probe(X)*Number of Transaction

Probe(X∪Y)- Minimum support*Probe(X) *Number of Transaction≥0

Lemma3: Shows the techniques for calculating of lift/importance

Lift/Importance=probe(X∪Y)/ Probe(X)*Probe(Y)≥ Minimum Lift

Probe(X∪Y)≥ Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|

Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|≥0

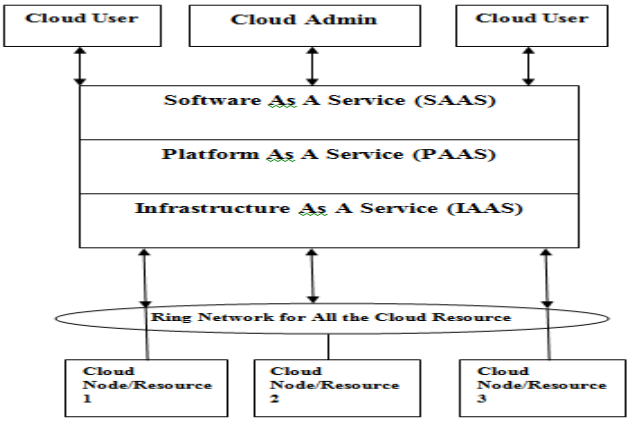

Figure1: Shows the arrangements of cloud nodes in the cloud environment

Algorithm1. (Calculation of Global support, confidence and Lift) using the list of protocols secure union, Intersection, Scalar Product, Secure Sum, Secure Subtraction in distributed cloud environment

Input: Distributed horizontal partitioned database in cloud computing environment, each cloud node have their own private database

Output: Secure global support, global confidence and global lift/importance

Step1: Consider the number of cloud nodes N≥3 from 1 to N

Step2: Segmented/fragmented the centralized database into the distributed cloud database (CD_1, CD_2….. CD_n)

Step3: For each cloud nodes

For I=1 to N-1 do /*Partitioned cloud database*/

Step4: Arrange all the cloud nodes in the ring format and consider the cloud node CD_1 as the protocol initiator node/* Consider the Protocol initiator node as trusted node*/

Step5: Converting the row wise database into the column wise database /*Reducing the number of transaction so reduced the space & time complexity in database*/

Step6: All the Cloud nodes now calculating the number of frequent item and non frequent item set from their own database

Step7: Select the number of frequent item set whose value greater than the minimum support value, minimum confidence and minimum lift

Calculated Support Value (CSV) ≥ Minimum Threshold Value (MT)

Calculated Confidence Value (CCV) ≥ Minimum Threshold Value (MT)

Calculated Lift Value (CLV) ≥ Minimum Threshold Value (MT)

Step8: All the Cloud Node has their own private Hash function (H mod N) for encrypting the data

Step9: Sanded the private key to the protocol initiator node

For I=2 to N do

Step10:

Let K=1 to 3 does

Switch (K)

Case1:

/*Calculation of the partial support by using the following formula*/

Partial Support=X. Support-Minimum Support*| Database Size |

For K=1 to 3 do;

/* Support*/

If (K=1)

/*Calculation of Partial Support Value*/

Enter Choice from J=1 to 5 does

If (J=1)

{

/*Now every cloud node takes union with the encryption key*/

For I=1 to N do;

Begin

PSi= (Xi. Support-Minimum Support*| Database Size |) ∪ Hi mod Ni

For each 1 to N-1 do;

Counter ++;

Func1 ( );

End;

}

Else If (J=2)

{

/*Now every cloud node takes intersection with the encryption key*/

For I=1 to N do;

Begin

PSi=(Xi. Support-Minimum Support*| Database Size |) ∩ Hi mod Ni

For each 1 to N-1 do;

Counter ++;

Func2 ( );

End

}

Else If (J=3)

{

/*Now every cloud node takes multiplication with the encryption key*/

For I=1 to N do;

Begin

PSi= (Xi. Support-Minimum Support*| Database Size |) *Hi mod Ni

For each 1 to N-1 do;

Counter ++;

Func3 ( );

End

}

Else If (J=4)

{

/*Now every cloud node takes sum with the encryption key*/

For I=1 to N do;

Begin

PSi= (Xi. Support-Minimum Support*| Database Size |) + Hi mod Ni

For each 1 to N-1 do;

Counter ++;

Func4 ( );

End

}

Else If (J=5)

{

/*Now every cloud node takes Secure Subtraction with the encryption key*/

For I=1 to N do;

Begin

PS= (X. Support-Minimum Support*| Database Size |) - H mod N

For each 1 to N-1 do;

Counter ++;

Func5 ( );

End

}

Else

{

Print “Does Not Occur”;

}

Func1 ( )

/*Now Cloud node CD_1 calculate the global support value by using the Secure Intersection formula*/

Global support count = Sum of all partial support -(H_1 mod N_1 ∩H_2 mod N_2∩….. ∩H_n mod N_n)

Func2 ( )

/*Now Cloud node CD_1 calculate the global support value by using the Secure Union formula*/

Global support count = Sum of all partial support -(H_1 mod N_1∪H_2 mod N_2∪….. ∪H_n mod N_n)

Func3 ( )

/* Now Cloud node CD_1 calculate the global support value by using the Secure Sum formula*/

Global support count = Sum of all partial support -(H_1 mod N_1+H_2 mod N_2+….. +H_n mod N_n)

Func4 ( )

/* Now Cloud node CD_1 calculate the global support value by using the Secure Multiplication formula*/
Global support count = Sum of all partial support -(H_1 mod N_1*H_2 mod N_2….. *H_n mod N_n)
Func5 ( )
/* Now Cloud node CD_1 calculate the global support value by using the Secure Subtraction formula*/
Global support count = Sum of all partial support -(-(H_1 mod N_1+H_2 mod N_2….. +H_n mod $N_{n)}$)
}
Break;
Case2:
{
/*Calculation of the partial confidence by using the following formula*/
Partial Confidence=Support (X∪Y)- Minimum confidence*Probe(X) *| Database Size |
If (K=2)
Enter Choice from J=6 to 10 does
If (J=6)
{
/*Now every cloud node takes union with the encryption key*/
For I=1 to N do;
Begin
PCi= (Probe(X∪Y)- Minimum confidence*Probe(X) *| Database Size |) ∪ Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func6 ( );
End;
}
Else If (J=7)
{
/*Now every cloud node takes intersection with the encryption key*/
For I=1 to N do;
Begin
PCi=(Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) ∩ Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func7 ( );
End
}
Else If (J=8)
{
/*Now every cloud node takes multiplication with the encryption key*/
For I=1 to N do;
Begin
PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) *Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func8 ( );
End
}
Else If (J=9)
{
/*Now every cloud node takes sum with the encryption key*/
For I=1 to N do;
Begin
PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) + Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func9 ( );
End
}
Else If (J=10)
{
/*Now every cloud node takes Secure Subtraction with the encryption key*/
For I=1 to N do;
Begin
PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |)) - H mod N
For each 1 to N-1 do;
Counter ++;
Func10 ( );
End
}
Else
{
Print “Does Not Occur”;
}

Func6 ( )
/* Now Cloud node CD_1 calculate the global confidence value by using the following formula*/
Global confidence count = Sum of all partial confidence -(H_1 mod N_1 ∩H_2 mod N_2∩….. ∩H_n mod N_n)
Func7 ( )
/* Now Cloud node CD_1 calculate the global confidence value by using the following formula*/
Global confidence count = Sum of all partial confidence -(H_1 mod N_1∪H_2 mod N_2∪….. ∪H_n mod N_n)
Func8 ( )
/* Now Cloud node CD_1 calculate the global confidence value by using the following formula*/
Global confidence count = Sum of all partial confidence -(H_1 mod N_1+H_2 mod N_2+….. +H_n mod N_n)
Func9 ( )
/* Now Cloud node CD_1 calculate the global confidence value by using the following formula*/
Global confidence count = Sum of all partial confidence -(H_1 mod N_1*H_2 mod N_2….. *H_n mod N_n)
Func10 ( )
/* Now Cloud node CD_1 calculate the global confidence value by using the following formula*/
Global confidence count = Sum of all partial confidence -(-(H_1 mod N_1+H_2 mod N_2….. +H_n mod $N_{n)}$)
}
Break;

Case3:
/*Calculation of the partial lift by using the following formula*/
Partial Lift = Support (X∪Y) - Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|
If (K=3)
Enter Choice from J=11 to 15 does
If (J=11)
{
/*Now every cloud node takes union with the encryption key*/
For I=1 to N do;
Begin
PLi= (Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∪ Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func11 ( );
End;
}
Else If (J=12)
{
/*Now every cloud node takes intersection with the encryption key*/
For I=1 to N do;
Begin
PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∩ Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func12 ( );
End
}
Else If (J=13)
{
/*Now every cloud node takes multiplication with the encryption key*/
For I=1 to N do;
Begin
PLi= (Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) *Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func13 ( );
End
}
Else If (J=14)
{
/*Now every cloud node takes sum with the encryption key*/
For I=1 to N do;
Begin
PLi= (Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) + Hi mod Ni
For each 1 to N-1 do;
Counter ++;
Func14 ( );
End
}
Else If (J=15)
{
/*Now every cloud node takes Secure Subtraction with the encryption key*/
For I=1 to N do;
Begin
PLi= (Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) - H mod N
For each 1 to N-1 do;
Counter ++;
Func15 ( );
End
}
Else
{
Print “Does Not Occur”;
}

Func11 ( )
/* Now Cloud node CD_1 calculate the global lift value by using the following formula*/
Global lift count = Sum of all partial lift -(H_1 mod N_1 ∩H_2 mod N_2∩….. ∩H_n mod N_n)
Func12 ( )
/* Now Cloud node CD_1 calculate the global lift value by using the following formula*/
Global lift count = Sum of all partial lift -(H_1 mod N_1∪H_2 mod N_2∪….. ∪H_n mod N_n)
Func13 ( )
/* Now Cloud node CD_1 calculate the global lift value by using the following formula*/
Global lift count = Sum of all partial lift -(H_1 mod N_1+H_2 mod N_2+….. +H_n mod N_n)
Func14 ( )
/* Now Cloud node CD_1 calculate the global lift value by using the following formula*/
Global lift count = Sum of all partial lift -(H_1 mod N_1*H_2 mod N_2….. *H_n mod N_n)
Func15 ( )
/* Now Cloud node CD_1 calculate the global lift value by using the following formula*/
Global lift count = Sum of all partial lift -(-(H_1 mod N_1+H_2 mod N_2….. +H_n mod N_n))
}
Break;
Default:
{
Print “does not match”;
}

Step11: Now Cloud node CD_1 broadcast the final global support/confidence/lift value to all the cloud nodes present in the cloud distributed database environments

In this paper, we have taken the market database for analyzing the database into cloud environment as centralized database, first divided the centralized database (Table1) into the distributed database(Table, Table Table) by using the different partitioning concept horizontal partitioned and vertical partitioned but in this paper we divided into the horizontally partitioned database ($DB=DB_1$, DB_2, ……DB_n), and convert the market database into the binary database shown in Table2, so all the nodes have their equal number of attributes. Every database have their own cloud owner and the main work of cloud owner or cloud nodes that don't want to disclose their private information to all the data owner presented in the cloud environments. Now the entire data owner calculated the number of frequent and infrequent item set by using the apriori algorithm for considering the minimum support 40%, minimum confidence 40% and minimum lift 10% for all the data set. And convert the row wise database into the column wise database (Table 3 horizontal partitioned database & Table 4 shows the converted data from row wise to column wise for data owner1) to reduce the number of transaction and also reduce the time and space complexity for analyzing the database. And every data owner have their own private hash key so all the data owner or cloud nodes sanded that key to the protocol initiator node. In this we consider cloud owner1 as the protocol initiator node. So all the calculation for global support, global confidence and global lift is done by the cloud owner 1. And arrange all the cloud nodes in the ring format, the architecture of our proposed is shown in figure1. The data owner1 calculated the frequent item sets are, So now consider only those attribute whose value is greater than the minimum support values= {Support1, Support2, Support3, Support4, Support5, Support6} and Select S3 as attribute for calculating the global support value and consider the hash function for providing the privacy to the database, H=1 and N=2 for all the cases of data owner 1, Calculate the hash key=H mod N, we used different protocols for calculating the partial support value for data owner 1, appendex2 shows the calculation of support, confidence and lift for rule designing

Secure Sum

Partial Support=X. Support-Minimum Support*|DB|+H mod N

Partial Support=2-0.4*4+1 mod 2=1.4

Secure Subtraction

Partial Support=X. Support-Minimum Support*|DB|-H mod N

Partial Support=2-0.4*4-1 mod 2=-0.6

Secure Multiplication

Partial Support= (X. Support-Minimum Support*|DB|)*H mod N

Partial Support= (2-0.4*4)*1 mod 2=0.4

Secure Union

Partial Support=(X. Support-Minimum Support*|DB|)∪H mod N

Partial Support=(2-0.4*4) ∪1 mod 2=(0.4) ∪(1)=1

Secure Intersection

Partial Support=(X. Support-Minimum Support*|DB|)∩H mod N

Partial Support=(2-0.4*4) ∩1 mod 2=(0.4) ∩ (1)=0.4

Calculation of Partial Confidence (S3→S4)

Secure Multiplication

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) *Hi mod Ni

PCi= (1-0.4*2*4)*1mod2=-2.2

Secure Addition

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) +Hi mod Ni

PCi= (1-0.4*2*4)+1mod2=-1.2

Secure Subtraction

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |)-Hi mod Ni

PCi= (1-0.4*2*4)-1mod2=-3.2

Secure union

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) ∪ Hi mod Ni

PCi= (1-0.4*2*4) ∪ 1mod2=(-2.2) ∪ (1)=1

Secure Intersection

PCi= (Probe(X∪Y) - Minimum confidence *Probe(X) *| Database Size |) ∩Hi mod Ni

PCi= (1-0.4*2*4) ∩1mod2=(-2.2) ∩(1)=-2.2

Calculation of Partial lift (S3→S4) at Minimum Lift=10%

Secure Intersection

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∩ Hi mod Ni

PLi=(1-0.1*2*2*4*4) ∩(1mod2)=(-5.4) ∩(1)=-5.4

Secure union

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∪ Hi mod Ni

PLi=(1-0.1*2*2*4*4) ∪ (1mod2)=(-5.4) ∪ (1)=1

Secure Sum

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) + Hi mod Ni

PLi=(1-0.1*2*2*4*4) +(1mod2)=-4.4

Secure Subtraction

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) - Hi mod Ni

PLi=(1-0.1*2*2*4*4)-(1mod2)=(-5.4)-(1)=-6.4

Secure Multiplication

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) * Hi mod Ni

PLi=(1-0.1*2*2*4*4)*(1mod2)=(-5.4)*(1)=-5.4

Data owner2 have the database that shown in Table 5 and Table 6 shows the converted database from row wise to column wise database. Then the data owner 2 calculated the frequent and infrequent item sets, calculation is shown in appendex2, so data owner2 have the following frequent item sets {Support3= 3/4=0.75, Support4= 2/4=0.50, Support6= 2/4=0.50} whose support value greater than the minimum support value, Select S3 as attribute for calculating the global values and consider the hash function for providing the high

privacy to the database so data owner 2 consider the H=10 and N=20 for all the cases of dataset that is available in data owner2, Calculate the hash key=H mod N

Secure Sum

Partial Support=X. Support-Minimum Support*|DB|+H mod N

Partial Support=3-0.4*4+10 mod 20=11.4

Secure Subtraction

Partial Support=X. Support-Minimum Support*|DB|-H mod N

Partial Support=3-0.4*4-10 mod 20=-8.6

Secure Multiplication

Partial Support= (X. Support-Minimum Support*|DB|)*H mod N

Partial Support= (3-0.4*4)*10 mod 20=14

Secure Union

Partial Support=(X. Support-Minimum Support*|DB|)∪H mod N

Partial Support=(3-0.4*4) ∪10 mod 20=(1.4) ∪(10)=10

Secure Intersection

Partial Support=(X. Support-Minimum Support*|DB|)∩H mod N

Partial Support=(3-0.4*4) ∩10 mod 20=(1.4) ∩ (10)=1.4

Now calculate the confidence of all the rules

Confidence (3→4)=Support (3→4)/ Support(3)=1/3=0.33, Confidence (3→6)=Support (3→6)/ Support(3)=1/3=0.33, Confidence (4→6)=Support (4→6)/ Support(4)=2/2=1.0

Calculation of Partial Confidence (S3→S4)

Secure Multiplication

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) *Hi mod Ni

PCi= (1-0.4*3*4)*10mod20=-38

Secure Addition

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) +Hi mod Ni

PCi= (1-0.4*3*4)+10mod20=6.2

Secure Subtraction

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |)-Hi mod Ni

PCi= (1-0.4*3*4)-10mod20=-13.8

Secure union

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) ∪ Hi mod Ni

PCi= (1-0.4*3*4) ∪ 10mod20=(-3.8) ∪ (10)=10

Secure Intersection

PCi= (Probe(X∪Y) - Minimum confidence *Probe(X) *| Database Size |) ∩Hi mod Ni

PCi= (1-0.4*3*4) ∩10mod20=(-3.8) ∩(10)=-3.8

Calculation of Partial lift for the rule (S3→S4)

Secure Intersection

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∩ Hi mod Ni

PLi=(1-0.1*3*2*4*4) ∩(10mod20)=(-8.6) ∩(10)=-8.6

Secure union

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∪ Hi mod Ni

PLi=(1-0.1*3*2*4*4) ∪ (10mod20)=(-8.6) ∪ (10)=10

Secure Sum

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) + Hi mod Ni

PLi=(1-0.1*3*2*4*4) +(10mod20)=1.4

Secure Subtraction

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) - Hi mod Ni

PLi=(1-0.1*3*2*4*4)-(10mod20)=-18.6

Secure Multiplication

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) * Hi mod Ni

PLi=(1-0.1*2*3*4*4)*(10mod20)=(-8.6)*(10)=-86

Data owner3 have the database that shown in Table 7 and Table 8 shows the converted database from row wise to column wise database. Then the data owner 3 calculated the frequent and infrequent item sets, calculation is shown in appendex2, so data owner3 have the following frequent item sets {support1, Support3, Support4, Support5, Support6} whose support value greater than the minimum support value, Select S3 as attribute for calculating the global values and consider the hash function for providing the high privacy to the database so data owner 3 consider the H=10 and N=10 for all the cases of dataset that is available in data owner3, Calculate the hash key=H mod N, Select S3 as attributes for calculating the global values

Secure Sum

Partial Support=X. Support-Minimum Support*|DB|+H mod N

Partial Support=3-0.4*4+10 mod 10=1.4

Secure Subtraction

Partial Support=X. Support-Minimum Support*|DB|-H mod N

Partial Support=3-0.4*4-10 mod 10=1.6

Secure Multiplication

Partial Support= (X. Support-Minimum Support*|DB|)*H mod N

Partial Support= (3-0.4*4)*10 mod 10=0

Secure Union

Partial Support=(X. Support-Minimum Support*|DB|)∪H mod N

Partial Support=(3-0.4*4) ∪10 mod 10=(1.4) ∪(0)=1.4

Secure Intersection

Partial Support=(X. Support-Minimum Support*|DB|)∩H mod N

Partial Support=(3-0.4*4) ∩10 mod 10=(1.4) ∩ (0)=0

Calculation of Partial Confidence (S3→S4)

Secure Multiplication

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) *Hi mod Ni

PCi= (2-0.4*3*4)*10mod10=0

Secure Addition

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) +Hi mod Ni

PCi= (2-0.4*3*4)+10mod10=-2.8

Secure Subtraction

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |)-Hi mod Ni

PCi= (2-0.4*3*4)-10mod10=-2.8

Secure union

PCi= (Probe(X∪Y)- Minimum confidence *Probe(X) *| Database Size |) ∪ Hi mod Ni

PCi= (2-0.4*3*4) ∪ 10mod10=(-2.8) ∪ (0)=0

Secure Intersection

PCi= (Probe(X∪Y) - Minimum confidence *Probe(X) *| Database Size |) ∩Hi mod Ni

PCi= (2-0.4*3*4) ∩10mod10=(-2.8) ∩(0)=-2.8

Calculation of Partial lift (S3→S4)

Secure Intersection

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∩ Hi mod Ni

PLi=(2-0.1*3*3*4*4) ∩(10mod10)=(-12.4) ∩(0)=-12.4

Secure union

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) ∪ Hi mod Ni

PLi=(2-0.1*3*3*4*4) ∪ (10mod10)=(-12.4) ∪ (0)=0

Secure Sum

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) + Hi mod Ni

PLi=(2-0.1*3*3*4*4) +(10mod10)=-12.4

Secure Subtraction

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) - Hi mod Ni

PLi=(2-0.1*3*3*4*4)-(10mod10)=-12.4

Secure Multiplication

PLi=(Probe(X∪Y)- Minimum Lift*Probe(X)*|Number of Transaction|*Probe(Y)*|Number of Transaction|) * Hi mod Ni

PLi=(2-0.1*3*3*4*4)*(10mod10)=(-12.4)*(0)=0

Now the cloud admin calculating the global support, global confidence and global lift values by using the following formulas

Global support count = Sum of all partial support -(H_1 mod N_1 ∩H_2 mod N_2∩….. ∩H_n mod N_n)

Global support count =0.4+1.4+0-(1∩10∩0)=1.8-0=1.8

Global support count = Sum of all partial support -(H_1 mod N_1∪H_2 mod N_2∪….. ∪H_n mod N_n)

Global support count=1+10+1.4-(1∪10∪0)=12.4-10=2.4

Global support count = Sum of all partial support -(H_1 mod N_1+H_2 mod N_2+….. +H_n mod N_n)

Global support count=1.4+11.4+1.4-(1+10+0)=3.2

Global support count = Sum of all partial support -(H_1 mod N_1*H_2 mod N_2….. *H_n mod N_n)

Global support count=0.4+14+0-(1*10*0)=14.4

Global support count = Sum of all partial support -(-(H_1 mod N_1+H_2 mod N_2….. +H_n mod N_n))

Global support count=-0.6-8.6+1.6-(-(1+10+0))=3.4

Now calculating the calculating the global confidence values

Global confidence count = Sum of all partial confidence -(H_1 mod N_1 ∩H_2 mod N_2∩….. ∩H_n mod N_n)

Global confidence count=-2.2-3.8-2.8-(1∩10∩0)=-8.8

Global confidence count = Sum of all partial confidence -(H_1 mod N_1∪H_2 mod N_2∪….. ∪H_n mod N_n)

Global confidence count=1+10+0-(1∪10∪0)=1

Global confidence count = Sum of all partial confidence -(H_1 mod N_1+H_2 mod N_2+….. +H_n mod N_n)

Global confidence count=-1.2+6.2-2.8-(1+10+0)=-8.8

Global confidence count = Sum of all partial confidence -(H_1 mod N_1*H_2 mod N_2….. *H_n mod N_n)

Global confidence count=-2.8-38+0-(1*10*0)=-40.8

Global confidence count = Sum of all partial confidence -(-(H_1 mod N_1+H_2 mod N_2….. +H_n mod N_n))

Global confidence count=-3.2-13.8-2.8-(-(1+10+0))=-8.8

Now calculating global lift values by using the following formulas

Global lift count = Sum of all partial lift -(H_1 mod N_1 ∩H_2 mod N_2∩….. ∩H_n mod N_n)

Global lift count=-5.4-8.6-12.4-(1∩ 10∩ 0)=-26.4

Global lift count = Sum of all partial lift -(H_1 mod N_1∪H_2 mod N_2∪….. ∪H_n mod N_n)

Global lift count=1+10+0-(1∪10∪0)=1

Global lift count = Sum of all partial lift -(H_1 mod N_1+H_2 mod N_2+….. +H_n mod N_n)

Global lift count=-4.4+1.4-12.4-(1+10+0)=-26.4

Global lift count = Sum of all partial lift -(H_1 mod N_1*H_2 mod N_2….. *H_n mod N_n)

Global lift count=-5.4-86+0-(1*10*0)=-91.4

Global lift count = Sum of all partial lift -(-(H_1 mod N_1+H_2 mod N_2….. +H_n mod N_n))

Global lift count=-6.4-18.6-12.4-(-(1+10+0))=-26.4

Consider only those global support, confidence and lift values whose values greater than zero, so after applying all concept of data mining and privacy we analyze the comparison results, so we conclude that the global support is 14.4 after applying the secure multiplication with the hash key so the cloud admin broadcast the results to all the cloud nodes presented in the cloud environments. Calculated global support 1 after applying the secure intersection and hash key to all the cloud data for every cloud owner, now cloud admin broadcast the global confidence values to all the cloud owner or cloud nodes presented in the cloud environments. Now finally calculated the global lift/importance values1 by applying the secure union with hash key. Now the cloud admin broadcast the result to all the cloud nodes presented in the cloud environments. And Table 9 shows the comparison result after applying all the protocols and hash key encryption method.

VII. Case Study: Market Basket Analysis in Cloud Environment

Market basket analysis is well known techniques for uncovering the relationship between the item sets which are purchased together. Also known as scanner panel data, the techniques identifies the association between the items or categories of item that purchase together or between items that the customer purchases together. Despite the research how to best extract and optimize the best associations and nothing has been written about the overall ability of the market basket analysis or increase the revenue of the market or on the magnitude such existence of the business. More recently market basket analysis has been combined with additional consumer behaviors. Market basket analysis first proposed in supermarket in Sweden. The techniques typically used for evaluating products affinity for retails stores but it can be also used to evaluate affinity for other types of choices: purchases of services, menu choices at restaurant etc. more recently the Market basket analysis has been combined with additional consumer information including in store behavior, visual effects arising from merchandise positioning in the store and choice experiments via mailed surveys. The market basket analysis is nonetheless widely used in part because of its conceptual simplicity. Given a set of transactions involving two or more items, the Market basket analysis starts by identifying those transactions that involves pairs of items A and B, the number of transactions involving item A is NA the number of transaction involving item B is NB and the number of transactions involving both A and B is NAB. These numbers can be obtained via simple queries in a database of transactions. The market basket analysis based on the three metrics support, confidence and lift. All three metrics are derived from the transaction record for the business. The first techniques defined for market basket analysis is support, which is the probability of an association. Given the number of items A and B occurs together in the same transaction and the total number of transaction as above the support is PAB=NAB/N. the second techniques defined is the market basket analysis is the conditional probability of an items to be purchased, given that another one has already been purchased. For example a customer who has already placed item A in her shopping cart will have a different probability of purchasing item B than if she had not decided to purchase A. mathematically confidence is given by CAB=PAB/PA. The third techniques in the market basket analysis is the lift LAB=SAB/SA*SB. Because the lift is a ratio of probabilities, it mitigates some of the problems with the earlier example of milk and bread. We now are applying the market basket analysis in the cloud environment. Where the centralized database shown in table1 is divided first into the distributed database so each cloud node have their own private database. and the cloud admin want to analyze the database without any data leakage. In this paper we assume the each cloud node have the four number of transaction, and consider the minimum support=40%, minimum confidence=40% and minimum lift=10% for all the database table presented in annexure 1. And cloud admin want to analyze the all database so the admin calculated the global support, global confidence and global lift value by using the algorithm1. Calculated global support=14.4 after applying the secure multiplication protocols. Global confidence=1 after applying the secure intersection protocols and global lift=1 after applying the secure union protocol. Now the cloud admin broadcast the analyzed result to all the nodes presented in the cloud environment.

VIII. Conclusion

As an important development direction of the next generation of cloud computing, cloud computing attracted many attention by industry world and academic circle. Cloud data has many characteristics such as distributed storage, mass related and position related data. Cloud computing is an architecture which is also known for its powerful capability of computation and storage resource sharing. These features make cloud computing favorable to data mining service in network environment. In this paper we have implement the algorithm to analyze the cloud database with high privacy, for analysis of cloud database we used the different association mining techniques support, confidence and lift, for privacy we used the different privacy preserving techniques to preserve the privacy to the cloud database, finally we compare all the result with the help of market distributed database in cloud environments. The integration of association rule mining and cloud computing for analyzing the cloud database with high privacy, in this paper is at the initial stage of our research work of data analysis we require the further improvement in cloud computing environment.

References

[1]. Jaiwei H., Kamber M. Pei J., Data mining concepts and techniques, 3rd edition 2011.
[2]. Katadri V., et. al. A review on data mining from past and future, International journal of computer applications, Feb. 2011, 15(7),
[3]. Gordan J., Hayesi C., Elron D, Huang L, Niel R. Exploring the future of cloud computing, World economic forum 2010.
[4].Arun K Pujari, Data mining Tecniques, 3rd edition, 15 October 2013.
[5]. Velte A. T., Elsenpeter R, Velte T.J. Cloud computing: A Practical Approach, TATA McGraw Hill Edition, 1st edition, 2009.
[6]. Ling Jaun Li, Jhang M. The stategy of mining association rule based cloud computing, International conference on business computing and global information, IEEE Explore, 2011.
[7]. Ambulkar B., Borkar V. Data mining in cloud computing, Proceedings published by international journal of computer application, ISSN 0975-8887, April 7-8, 2012, pp.23-26.
[8]. Pareek A., Gupta M., Review of data mining techniques in cloud computing database, International journal of advance computer research, June 2012, 2 (2), pp.52-55.
[9]. Ashrafi M. Z., Tanir D., Smith K.. ODAM: An optimized distributed association rule mining algorithm, IEEE distributed system online, March 2004, 5(3), pp 2-18.
[10] Lin, K.W, Luo Y. C. A fast parallel algorithm for discovering frequent pattern, IEEE international conference on granular computing, Aug 17-19, 2009, pp 398-403.
[11]Liu B., Cao S.G., He W. Distributed data mining for e business, Information technology and management, June 2011, 12(2), pp.67-79.
[12]. T.V. Mahendra, et. al., Data mining for high performance data cloud using association rule mining, international journal of advanced research in computer science and software engineering, January-2012, 2(1).
[13]. Bedra, A. Getting started with Google App Engine and Closure", Internet computing, 2010, 14, pp.85-88.

[14]. Wang,L, et. al. Scientific cloud computing early definition and experience, In procceding of the 10th IEEE international conference on high performance computing and communications.2008, pp 825-830.
[15]. Grossman, RL, Nerode A., Ravan AP., Rischel H. Compute and storage clouds wide area high performance network, Future generation computer system, 2009, 25(2), pp.179-183.
[16]. Yue Y.X., Zhen L., Yan F. Map reduce as a programming model for association rules algorithm on Hadoop, 3rd international conference on information science and interaction science, 2010, pp 2486-2494.
[17] Frank S.C., Kuo Y.H., Huang Y.M., Towards boosting distributed association rule mining by data de-clustering , Information science, Nov-2010, 180(22), pp.4263-4289.

Appendex1: Shows all the database tables for all the cloud nodes presented in the cloud computing environments.

Table1: Market database (Attributes and Transactions)

Tid	Attribute1	Attribute2	Attribute3	Attribute4
1	Bread	Butter		
2	Bread	Butter	Egg	
3			Egg	Jam
4	Bread			Jam
5	Bread	Butter	Egg	
6		Butter	Egg	
7	Bread			
8		Butter		
9	Bread	Butter	Egg	
10			Egg	Jam
11		Butter		
12			Egg	Jam
13		Butter	Egg	Jam
14	Bread			
15	Bread	Butter		Jam
16		Butter	Egg	Jam
17	Bread	Butter	Egg	
18		Butter	Egg	

Table2: Conversion of sensitive database into the binary database

Tid	Attribute1	Attribute2	Attribute3	Attribute4
1	1	1	0	0
2	1	1	1	0
3	0	0	1	1
4	1	0	0	1
5	1	1	1	0
6	0	1	1	0
7	1	0	0	0
8	0	1	0	0
9	1	1	1	0
10	0	0	1	1
11	0	1	0	0
12	0	0	1	1
13	0	1	1	1
14	1	0	0	0
15	1	1	0	1
16	0	1	1	1
17	1	1	1	0
18	0	1	1	0

Table3: Cloud Node 1 or Data Owner1 Data base

Tid	Attribute1	Attribute2	Attribute3	Attribute4
1	1	1	0	0
2	1	1	1	0
3	0	0	1	1
4	1	0	0	1
5	1	1	1	0
6	0	1	1	0

Table 4: Transfer row into column and column into row by using the matrix transpose method for cloud node1 or data owner1

Tid	1	2	3	4	5	6
Attribute1	1	1	0	1	1	0
Attribute2	1	1	0	0	1	1
Attribute3	0	1	1	0	1	1
Attribute4	0	0	1	1	0	0

Table5 : Cloud Node 2 or Data Owner2 Data base

Tid	Attribute1	Attribute2	Attribute3	Attribute4
1	1	0	0	0
2	0	1	0	0
3	1	1	1	0
4	0	0	1	1
5	0	1	0	0
6	0	0	1	1

Table6: Transfer row into column and column into row by using the matrix transpose method for cloud node2 or data owner2

Tid	1	2	3	4	5	6
Attribute1	1	0	1	0	0	0
Attribute2	0	1	1	0	1	0
Attribute3	0	0	1	1	0	1
Attribute4	0	0	0	1	0	1

Table7: Cloud Node 3 or Data Owner3 Data base

Tid	Attribute1	Attribute2	Attribute3	Attribute4
1	0	1	1	1
2	1	0	0	0
3	1	1	0	1
4	0	1	1	1
5	1	1	1	0
6	0	1	1	0

Table8: Transfer row into column and column into row by using the matrix transpose method for cloud node3 or data owner3

Tid	1	2	3	4	5	6
Attribute1	0	1	1	0	1	0
Attribute2	1	0	1	1	1	1
Attribute3	1	0	0	1	1	1
Attribute4	1	0	1	1	0	0

Table9: Shows the Compression result after applying the all data mining protocols and hash function

S/No	Secure Sum	Secure Subtraction	Secure Multiplication	Secure Union	Secure Intersection
Global Support	3.2	3.4	14.4	2.4	1.8
Global Confidence	-8.8	-8.8	-40.8	-8.8	1
Global Lift/ Importance	-26.4	-26.4	-91.4	1	-26.4

Appendex2: Shows the calculation of support, confidence and lift values for all the data owner presented in the cloud computing environments

1. Cloud Node1 or Data Owner1

Support count1=2, Support count2=3, Support count3=2, Support count4=2, Support count for attribute 5=3, Support count for attribute 6=2,

Support1= 2/4=0.50, Support2= 3/4=0.75, Support3= 2/4=0.50, Support4= 2/4=0.50, Support5= 3/4=0.75, Support6= 2/4=0.50.

Minimum support Value=40%=40/100=0.40

So now consider only those attribute whose value is greater than the minimum support values= {Support1, Support2, Support3, Support4, Support5, Support6}

Rule designing

{Support1→Support2, Support1→Support3, Support1→Support4, Support1→Support5, Support1→Support6, Support2→Support3, Support2→Support4, Support2→Support5, Support2→Support6, Support3→Support4, Support3→Support5, Support3→Support6, Support4→Support5, Support4→Support6, Support5→Support6}

Now calculate the confidence of all the rules

Confidence(1→2)=Support(1→2)/Support(1)=2/2=1,Confidence(1→3)=Support(1→2)/Support(1)=0/2=0,

Confidence(1→4)=Support (1→4)/ Support(1)=1/2=0.50,

Confidence(1→5)=Support (1→5)/ Support(1)=2/2=1,

Confidence(1→6)=Support (1→6)/ Support(1)=1/2=0.50,

Confidence (2→3)=Support(2→3)/Support(2)=1/2=0.5,Confidence(2→4)=Support(2→4)/Support(2)=1/2=0.5,Confidence(2→5)=Support (2→5)/ Support(2)=3/2=1.5, Confidence(2→6)=Support (2→6)/ Support(2)=2/2=1.0, Confidence (3→4)=Support (3→4)/ Support(3)=1/2=0.5, Confidence (3→5)=Support (3→5)/ Support(3)=1/2=0.5, Confidence (3→6)=Support (3→6)/ Support(3)=1/2=0.5, Confidence (4→5)=Support (4→5)/ Support(4)=1/2=0.5, Confidence (4→6)=Support (4→6)/ Support(4)=0/2=0.0, Confidence (5→6)=Support (5→6)/ Support(5)=2/3=0.66

lift(1→2)=Support(1→2)/Support(1)*Support(2)=2/2*3=0.33, Lift(1→3)=Support(1→3)/Support(1)*support(3)=0/2*2=0,Lift(1→4)=Support(1→4)/Support(1)*support(4)=1/2*2=0.25,Lift(1→5)=Support(1→5)/Support(1)*support(5)=2/2*3=0.33,

Lift(1→6)=Support(1→6)/Support(1)*support(6)=1/2*2=0.25 , Lift(2→3)=Support (2→3)/ Support(2)*support(3)=1/2*2=0.25,
Lift(2→4)=Support(2→4)/Support(2)*support(4)=1/2*2=0.25 ,Lift(2→5)=Support(2→5)/Support(2)*support(5)=3/2*3=0.5 0,Lift(2→6)=Support(2→6)/Support(2)*support(6)=2/2*2=0. 50,Lift(3→4)=Support(3→4)/Support(3)*support(4)=1/2*2=0 .25,Lift(3→5)=Support(3→5)/Support(3)*support(5)=1/2*3= 0.16,Lift(3→6)=Support(3→6)/Support(3)*support(6)=1/2*2 =0.25,lift(4→5)=Support(4→5)/Support(4)*support(5)=1/2*3 =0.16,Lift(4→6)=Support(4→6)/Support(4)*support(6)=0/2* 2=0.0,Lift(5→6)=Support(5→6)/Support(5)*support(6)=2/2

2. Cloud node2 or Data owner2

Support count1=1, Support count2=1, Support count3=3, Support count4=2, Support count for attribute 5=1, Support count for attribute 6=2,
Support1= 1/4=0.25, Support2= 1/4=0.25, Support3= 3/4=0.75, Support4= 2/4=0.50, Support5= 1/4=0.25, Support6= 2/4=0.50.
Minimum support Value=40%=40/100=0.40
So now consider only those attribute whose value is greater than the minimum support values= {Support3, Support4, Support6}
Rule designing
{Support3→Support4, Support3→Support6, Support4→Support6}
Now calculate the confidence of all the rules
Confidence (3→4)=Support (3→4)/ Support(3)=1/3=0.33,
Confidence (3→6)=Support (3→6)/ Support(3)=1/3=0.33,
Confidence (4→6)=Support (4→6)/ Support(4)=2/2=1.0
Lift(3→4)=Support(3→4)/Support(3)*support(4)=1/3*2=0.16 ,Lift(3→6)=Support(3→6)/Support(3)*support(6)=1/3*2=0.1 6,
Lift(4→6)=Support(4→6)/Support(4)*support(6)=2/2*2=0.50

3. Cloud node3 or Data owner3

Support count1=3, Support count2=1, Support count3=3, Support count4=3, Support count for attribute 5=3, Support count for attribute 6=2,
Support1= 3/4=0.75, Support2= 1/4=0.25, Support3= 3/4=0.75, Support4= 3/4=0.75, Support5= 3/4=0.75, Support6= 2/4=0.50.
Minimum support Value=40%=40/100=0.40
So now consider only those attribute whose value is greater than the minimum support values= { Support1, Support3, Support4, Support5, Support6}
Rule designing
{Support1→Support3, Support1→Support4, Support1→Support5, Support1→Support6, Support3→Support4, Support3→Support5, Support3→Support6, Support4→Support5, Support4→Support6, Support5→Support6}
Now calculate the confidence of all the rules
Confidence(1→3)=Support(1→3)/Support(1)=2/3=0.66,
Confidence(1→4)=Support (1→4)/ Support(1)=3/3=1.00,
Confidence(1→5)=Support (1→5)/ Support(1)=3/3=1.0,
Confidence(1→6)=Support (1→6)/ Support(1)=2/3=0.66,
Confidence (3→4)=Support (3→4)/ Support(3)=2/3=0.66,
Confidence (3→5)=Support (3→5)/ Support(3)=1/3=0.33,
Confidence (3→6)=Support (3→6)/ Support(3)=1/3=0.33,
Confidence (4→5)=Support (4→5)/ Support(4)=2/3=0.66,
Confidence (4→6)=Support (4→6)/ Support(4)=2/3=0.66,
Confidence (5→6)=Support (5→6)/ Support(5)=2/3=0.66
Lift(1→3)=Support(1→3)/Support(1)*support(3)=2/3*3=0.22 ,Lift(1→4)=Support(1→4)/Support(1)*support(4)=3/3*3=0.3 3,Lift(1→5)=Support(1→5)/Support(1)*support(5)=3/3*3=0. 33,
Lift(1→6)=Support(1→6)/Support(1)*support(6)=2/3*2=0.33 ,Lift(3→4)=Support(3→4)/Support(3)*support(4)=2/3*3=0.2 2,Lift(3→5)=Support(3→5)/Support(3)*support(5)=1/3*3=0. 11,Lift(3→6)=Support(3→6)/Support(3)*support(6)=1/3*2=0 .16,lift(4→5)=Support(4→5)/Support(4)*support(5)=2/3*3=0. 22,Lift(4→6)=Support(4→6)/Support(4)*support(6)=2/3*2=0 .33,Lift(5→6)=Support(5→6)/Support(5)*support(6)=2/2*3= 0.33