Discovery of frequent itemsets using Graph and Cluster Technique

D.Senthil Kumar[1], Dr. N. Jayaveeran[2]
[1]*Research Scholar, PG and Research dept. of Computer Science, Khadir Mohideen College, Adirampattinam, India.*
[2] *Head and Associate Professor PG and Research dept. of Computer Science, Khadir Mohideen College, Adirampattinam, India.*
(E-mail: ds_kum77*@yahoo.com)*

***Abstract*—** There is a common belief that the association rule mining has a major stake in data mining research domain. Numerous algorithms are proposed for discovering frequent itemsets. This is the key process in association rule mining. This proposed paper introduces two concepts. First leads to analyze the ability to explore the demographic parameters of the underlying entities and their inter relations using the traditional graph theory approach (vertices and edges). Second is the improved algorithm based on graph and clustering based mining association rules. This improved algorithm is named Discovery frequent itemsets using Graph and Cluster Technique (DFIGS). The DFIGS algorithm scans the database of transaction only once to generate a cluster table and then clusters the transactions into cluster according to their length. The DFIGS algorithm is used to find frequent itemsets and will be extracted directly by scanning the cluster table. This method reduces memory requirement and time to retrieve the datasets and hence it is scalable for any large size of the database.

***Keywords*—** *Association rule mining, Data Mining, Relational, cluster, graph, DFIGS.)*

I. INTRODUCTION

In the present educational environment, providing accommodation to students in the college hostel with necessary facilities, safeguarding and inducing them to concentrate on their studies hour become an arduous. Students staying in hostel differ in state, race, religion, language, food, clothe, culture and habits. For example NRI students and north Indian students especially from Manipur, Nagaland, Assam etc. Studying in tamilnadu are provided separate accommodation and facilities are given to take the food which they desire. Some college provide separate single room with attached bath room. Without minding the money, a few are ready to pay a very high amount for leading a sophisticated life.

Here carries are going to be analyzed for students' hesitation to stay in hostel in general. In a few hostels, students may be restricted to use adequate water for taking bath and washing their clothes.

In some hostels limited quantity of food may be supplied. They may not like more usage of spice, oil, butter, ghee etc for the preparation of food in the mess. Sometimes hostlers may be harshly treated by assistants or supervisors or hostel office clerks. The cleanliness maintained in the kitchen may be disliked by some.

In fact, if the Institution hostel is able to gather data coming from different students (existing databases called students databases) and to analyze them, the hostel can supply its foods the requested quality in due time and can react faster when students have problems. Then hostel could become more efforts and be able to develop more and more its quality accommodation.

Frequently, data are shared throughout different Institution hostel areas and applications and are not usually centralized in a centralized unique view. Consequently, the process of decision-making from data analysis is uneasy. Decisional information systems (DIS) have been recognized as an appropriate approach to facilitate access of data stored in different applications. The aim of DIS is to discover links between data in order to check a set of intuitive hypothesis based on experience. This way, DIS allow to understand and to evaluate the quality within environment.

Successfully supporting decision-making is based on the availability of integrated high-quality data organized and presented in a timely and easily understandable manner. Data warehouses emerged to meet this requirement. With the rapid development in size and number of available databases in commercial, organizational, administrative and other applications [1], it is the urgent need for new technologies and automated tools to change this wealth of data resources into useful information. The most challenges in database mining is developing fast and efficient algorithms that can deal with large volume of data because several data mining algorithm computation to solve diverse data mining problem. They are mainly classified as associations, classifications, sequential patterns and clustering [2].

A. Association Rule

Association rule mining is a very important research topic in the data mining field, it's problem in large databases is to generate all association rules [17], of the form X =>Y, that will produce strong association rules which satisfy both minimum support degree (min_sup) and the minimum

confidence degree (min_conf) greater then the user defined minimum support and minimum confidence [2, 3.4].

Definition 1: let X={x1, x2 ….. xn} be a set of items, then $D = \{< T_{id}, T > | T \subseteq X\}$ is a transaction database, where T_{id} is an identifier which be associated with each transaction.

Definition 2: Let $A \subseteq X, B \subseteq X, and A \cap B = \phi$, we called $A \Rightarrow B$ as association rule.

Most of the algorithm generally in two steps execution. First, to finding all sets of items that have support above the given minimum, and then generating the desired rules from these item sets. The apriori algorithm has in same first, to find all frequent itemsets, the other is generate strong association rules from frequent item sets during pruning [18].

B. Apriori Algorithm

The apriori algorithm is a fast algorithm for mining association rules and is based on [5] algorithms for mining association rules, the problem of this algorithm is number of data scans n, where n is the size of large nonempty itemset and number of discovering rules is huge while most of the rules are no interesting. Therefore several improved algorithms were proposed after apriori for efficiency and scalability.

- This modelling environment should allow the hostel to optimize the better students quality and like to stay.
- The DIS will provide the international company, decision aid-tools for decision makers in order to obtain information and knowledge for operational, tactical but also strategic decisions. These tools will allow the firm to:
- Reach the goal between the needs required for taken into facility (inmate, food, water, health care, etc.) in order to make an sufficient services to their students,
- Increasing the students care at right timed (the moment is identified with an analysis of evolution of performance rate), and
- Gain students good feedback.

This paper is organized as follows. At first, the basic concept of data warehouse and the state of art in the studied field is introduced. Section 2 discuss the literature review on association rule. Section 3 introduces the modelling environment built with the extends the case to UML design for the fact table developed for this case. At the end, the finding are concluded in research perspectives.

II. LITERATURE REVIEW ON ASSOCIATION RULE

Existing studies in data mining have presented efficient algorithm for discovering association rules. But the main drawback of the first algorithm is the need to do multiple passes over the datasets to generate frequent itemsets. The apriori association rule algorithm proposed by Agrawal and Srikant [6] can discover meaningful itemsets, but a large number of the candidate itemsets are generated from single itemsets and level by level in the process of creating association rules. Performance is severely affected because the database is scanned repeatedly to each candidate itemset with the database.

FP – Growth [7] out performs all candidate set generation and test algorithms as it mines frequent patterns without candidate generation. The main problem is no common prefixes within the data items.

Sample algorithm reduces the scanning to the datasets, it's scans single scan, but wastes considerable time on candidate itemsets [8].

The column wise apriori algorithm [9] and the tree based association rule algorithm [10] transformed the storage structure of the data, to reduce the time of scans of the transaction database.

The partition algorithm to improve efficiency and reduce the database scans, but still wasted scans on infrequent candidate itemset [11].

The primitive association rule mining is mining that describe the association among items of the transaction in the database. A uniform frame work was framed to perform association rule of association rules[12].

In the generalized association patterns, one can add all ancestors for each items from concept hierarchy and then apply the algorithm on the extended transactions [13].

The tones of day to day customer transactions stored in a very large database are processed and discovered multiple level of transactions with relevant attributes. Each of these attribute represented certain concept. Combining these attributes later generates multiple level concepts [5].

For frequent patterns in this new graph model which we call taxonomy superimposed graphs, there may be many patterns that are implied by the generalization and specialization hierarchy of the associated node label taxonomy [14]

Graph databases are able to represent as graphs of any kind of information, where naturally accommodated changes in data can be possible [15, 16].

The disadvantage of these algorithms is

- Number of reads the database transaction n time data scans where n is the size of large nonempty itemset,
- It is an incompetent as it requires wastage memory.
- Huge non interesting rules are discovered.

The proposal method DFIGS is a consequence to overcome the above said drawbacks.

III. THE MODELLING ENVIRONMENT

The general functions of hostel management are to: categorize students, health care, provide facility and allocate resources, but also to control and to manage Hostel activities. Hostels should systematically acquire the information needed to make decisions and to evaluate the effects of these decisions. To fully explore the opportunities for our approach, the propose a

modelling environment based on data warehousing approach allowing to manage the resources, to elaborate planning strategy in food and Hebetate or in maintenance or in disease. So, the role of information is crucial in the methodological framework. Fig. 1 presents the main characteristics of the proposed modelling environment.

This modelling environment takes into account all the processes of extracting, transforming and loading data from disparate databases to the data warehouse until the knowledge discovery in database

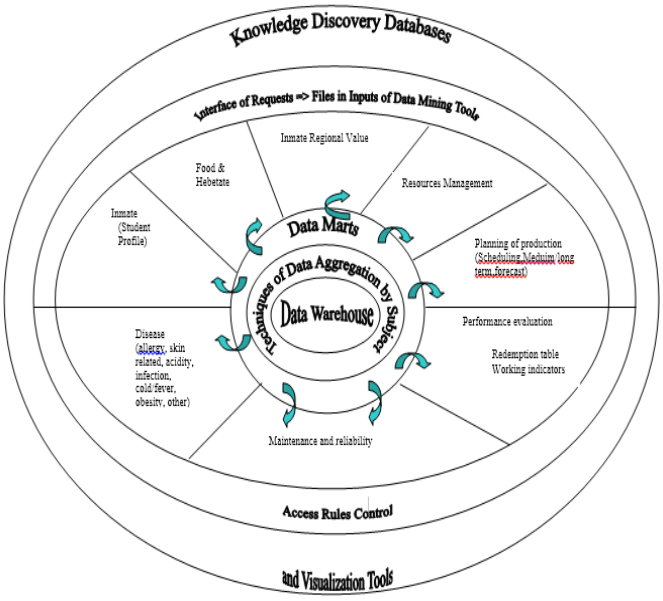

Fig. 1 presents the main characteristics of the proposed modelling environment.

The kernel of the modelling environment depends on the data warehouse architecture. From the data warehouse, we design data marts (small data warehouses) by extraction and aggregation of data according to specific subjects. Each subject is split up on modules concerning the Inmate Profile, the production planning, the Food and Hebetate, the maintenance and reliability and the performance evaluation. Each module concerns one or more activities. All identified modules are in interaction. For example, if they interested by production, they take into account together not only the planning but also the Disease, the resource management, the maintenance, the food and Hebetate , etc. guarantee

The modelling environment will allow a step-by step process to help the hostel administration to:

Evaluate the quality, reliability and controlling the maintenance costs of the accommodation,

Define the inmate of students as well as their profile in order to determine inmate's regional value, Food & Hebetation, etc. andto plan for the medium-term and long term production,

For the Hostel Institution, anticipate the resource needs in terms of planning, and simulate different scenarios to identify the best one.

Study and analyze the inmates' feedback,

Determine the impact of the inmates' behaviours on the food intake,

etc…

The originality of our research is to treat at the same time aspects of design of data warehouse and their exploitation (using knowledge discovery in large databases) within a modelling methodological framework for optimization and performance evaluation. Very few authors look into this problem. Calvanesea et al [10] present a methodology for data warehouse design based on a conceptual representation of the enterprise, which is exploited both in the integration phase of the data warehouse

Information sources and during the knowledge discovery activity on the information stored in the data warehouse. Alshawi et al. (2003) propose a data warehouse providing the student information required to optimize the time that the Institution hostel decides to file the compound of a patent. No modelling environment is proposed. The main works treat only of methodological framework to optimize and to evaluate performances [15].

The kernel of our modelling environment is based on the data warehouse and the performance of the data warehouse depends on the model architecture. In the next section, we present the multidimensional models concerning our data warehouse.

Institution administration aims to study the inmate food intake, acceptable and unacceptable habituate, and regional weather value, are the main reasons for health problem. But the hostel prepares and provides hygienic food and clean environment regularly. Many strategies followed in collecting all these inmate information, finally results to improve residence facility, inmate's likeness, etc.

We have identified four fact tables

- Inmate (this is concerning inmate details like name, address, date of birth, age, religion, city, contact number, etc.)
- Hebetate (the measurements concerning acceptable and unacceptable hebetate)
- Liquid intake(the measurement concerning per day liquid intake)
- Disease (the measurement concerning after solid and liquid intake, acceptable and unacceptable hebetate, low drinking water intake, etc.)
- The other tables concern the dimensional tables (potential analysis axis).
- The hebetate is main problem of inmate diseases, not the hostel environment and hostel food preparation.

1. Inmate details

- Here insert the inmate personal information for further communication, if the incharge want to inform his/her parents.
- Number of time going in and out from hostel(month,week)
- Home town information to know the weather.
- Calculate body weight(compare height and wieght)

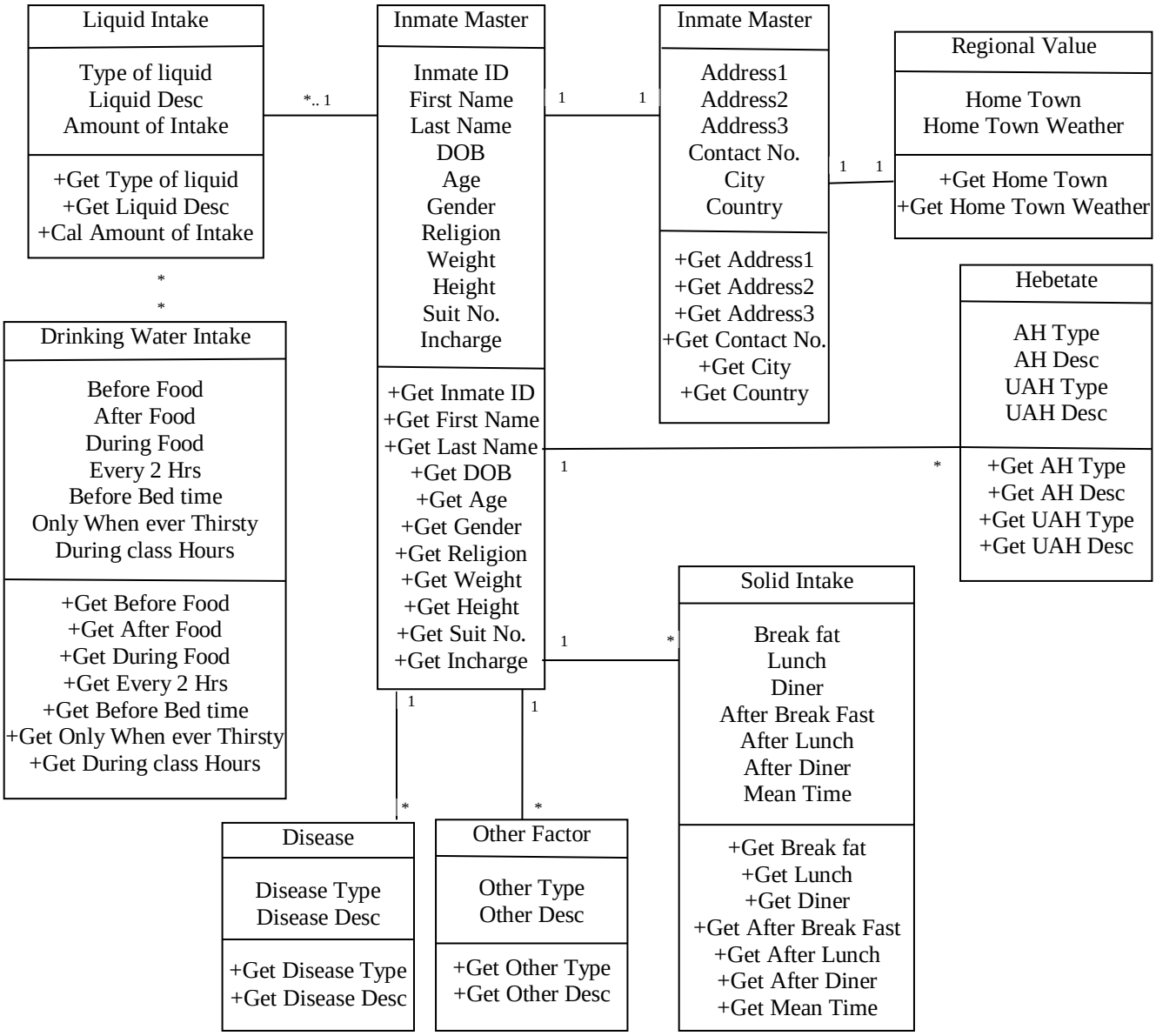

Fig. 2 Food intake and Hebetate Example in UML

2. Hebetate

- This can be divided in to two table that is acceptable and unacceptable hebetate.(inmate no, AH(acceptable hebetate) type,UNH(unacceptable hebetate) type)
- here they point out daily activity (AH type,AH Desc)
- Here they point out unacceptable think like drugs, tobacco, etc.(UNH type,UNH desc).

3.Liquid Intake

- Average intake of drinking water and other liquid food per day. It is very important for human body.
- Some students very eager to take beverages liquid drinks in out side of the campus, it is not good for his/her health.
- The beverages drinks create more side effect to human body like teeth - pain,
- injuring, poor strength, abdomen pain, obesity problem.
- If they take more coffee and tea in his/her carrier affect gallbladder.
- The student take soups and fresh juices, it's very good for health, but most of the student not interest to take like this.

4.Diseases

- This section is mainly focus on students side effect.
- Students focus his/her health problem to leave hostel, but it is not true.

TABLE I. AN EXAMPLE OF TRANSACTION DATABASE

TID	Plain Water	Liqid Intake	Solid Intake	Type of Intake	DISEASE
T1	below 1 litre	soft drinks	tawa fry	vegetarian	Ulcer
T2	1 to 3 litres	hot drinks	boiled	vegetarian	Obesity
T3	3 to 5 litres	beverages	tawa fry	non-vegetarian	Obesity
T4	1 to 3 litres	hot drinks	tawa fry	vegetarian	Acidity
T5	1 to 3 litres	beverages	boiled	vegetarian	Obesity
T6	3 to 5 litres	hot drinks	fast food	non-vegetarian	Obesity
T7	below 1 litre	beverages	tawa fry	non-vegetarian	Ulcer
T8	1 to 3 litres	hot drinks	tawa fry	vegetarian	Acidity
T9	below 1 litre	beverages	fast food	non-vegetarian	Ulcer
T10	1 to 3 litres	hot drinks	fast food	vegetarian	Obesity
T11	3 to 5 litres	beverages	tawa fry	non-vegetarian	Obesity
T12	1 to 3 litres	beverages	boiled	non-vegetarian	Acidity
T13	3 to 5 litres	beverages	tawa fry	vegetarian	Obesity
T14	1 to 3 litres	hot drinks	tawa fry	vegetarian	Acidity
T15	below 1 litre	soft drinks	tawa fry	vegetarian	Ulcer
T16	1 to 3 litres	soft drinks	boiled	non-vegetarian	Obesity
T17	below 1 litre	soft drinks	tawa fry	vegetarian	Ulcer
T18	1 to 3 litres	beverages	boiled	non-vegetarian	Acidity
T19	below 1 litre	hot drinks	boiled	vegetarian	Ulcer
T20	1 to 3 litres	hot drinks	fast food	vegetarian	Obesity

- Most of the health problem arrive from his/her poor hebetate only. Like some students interest chunk food, deep fry items, tawa fry items and late night diner, it is create obesity problem. After food most of them like easy desert and take heave drinking water. some time this type of food hebetate create heart attack disease.
- The other disease main reason is inmate daily activity and friendship, like skin disease, allergy and cold/ fever.
- Unified Modeling Language (UML) [19] is an emerging standard for conceptual data modeling, in object-oriented domain. It is a comprehensive language to model structural schema and dynamic behavior at conceptual level. As far as database design is concerned we are only interested in modeling the static structure of the system. In some ways the UML is quite similar to the Multidimensional model, we now have class diagrams. Figure 2 is the equivalent UML representation of the A part of the logical model

IV. PROPOSED METHOD

After pre-processing the set minimum support threshold is 50%. There are 20 transactions and five different items named Plain Water, Liqid Intake, Solid Intake, Type of Intake and Disease in the database. Most of the rule mining algorithms are

lexicographical order. An example of transaction database is shown in Table I.The items name of the transaction database is used rather than numbers to deal with some worst cases. The requirement of numbering system is to first scan the database to identify the length of each transaction, that means length of the numbers to the items in a transaction, and at the same time assigning the simples(H and L) to the items: Number 1 is assigned as item Plain Water, Number 2 is assigned as item Liqid Intake, Number 3 is assigned as item Solid Intake, Number 4 is assigned as item Type of Intake, Number 5 is assigned as item Disease. This conversion process help us in both constructing the cluster table and building the graph, this process help avoid the need to rescan the transaction database. Next move is the clustering table that can easily reside in the main memory.

In this example, the maximum transaction length is five as shown in table II, there will be at most five clusters, the total number of clusters is five as shown in table III, The presence and absence of an item in a transaction is denoted by H and L in content of the table. After that, the bit vector for each item will be ready and it is an easy process to determine the frequent Transaction Database convert to Symbol (H and L)

TABLE II. TRANSACTION DATABASE CONVERT TO SYMBOL (H AND L)

Item No. / TID	1	2	3	4	5
T1	L	H	L	H	L
T2	H	H	H	H	H
T3	H	H	L	L	L
T4	H	L	L	H	H
T5	H	H	H	H	H
T6	H	L	L	L	H
T7	L	H	L	L	L
T8	H	L	L	H	L
T9	L	H	L	L	L
T10	H	L	L	H	H
T11	H	H	L	L	H
T12	H	H	H	L	L
T13	H	H	L	H	H
T14	H	L	L	H	L
T15	L	H	L	H	L
T16	H	H	H	L	H
T17	L	H	L	H	L
T18	H	H	H	L	L
T19	L	L	H	H	L
T20	H	L	L	H	H

1 itemsets by counting the number of symbol Hs in each transaction, the minimum support threshold is not less than

counting the number of Symbol Hs, but it is considered as a frequent itemset and then building the graph.

TABLE III. THE CLUSTER TABLE FORM THE DATABASE IN TABLE 1

Item No. / TID	1	2	3	4	5
T7	L	H	L	L	L
T9	L	H	L	L	L
T1	L	H	L	H	L
T4	H	L	L	H	L
T6	H	L	L	L	H
T8	H	L	L	H	L
T14	H	L	L	H	L
T15	L	H	L	H	L
T17	L	H	L	H	L
T19	L	L	H	H	L
T3	H	H	L	L	H
T10	H	L	L	H	H
T11	H	H	L	L	H
T12	H	H	H	L	L
T18	H	H	H	L	L
T20	H	L	L	H	H
T2	H	L	H	H	H
T13	H	H	L	H	H
T16	H	H	H	L	H
T5	H	H	H	H	H

The bit vectors for the items are:

BV1=LLLHHHHLLLHHHHHHHHHH
BV2=HHHLLLLHHLHLHHHLLHHH
BV3=LLLLLLLLLHLLLHHLHLHH
BV4=LLHHLHHHHHLHLLLHHHLH
BV5=LLLLHLLLLLHHHLLHHHHH

By counting the number of Symbol Hs in each bit vector is determined the support for each candidate itemset of length 1, as follows: support ({1}) = 70%, support ({2}) = 60%, support ({3}) = 30%, support ({4.}) = 60%, support ({5}) = 45%. Thus the frequent 1 itemsets are :{{1}, {3}, {4}, {5}} as their supports are not less than 45%.

The second step starts by reordering frequent 1 itemsets by providing each one with a sequential number to facilitate the process of constructing the graph, which is making logical( ) and operation between each pair of consecutive frequent 1 itemsets<itemi,itemj>| i<j if the number of symbol Hs in the result is greater than or equal to minimum support threshold, a edge is directed to drawn from itemi to itemj , this process is repeated for all frequent 1 itemsets. The simple directed graph to display frequent k– itemsets, k>=2 is shown in figure 1, and by assigning 35% as a new value to the minimum support threshold, the frequent 2 itemsets will be: {{1,3}, {1,4}, {3,5}}as shown in table III and the graph is constructed by drawing an edge between each pair of frequent items, as shown in figure 2. By counting the number of 1s in each bit vector is determined the support for each candidate itemset of length 2, as follows: support ({1,3}) = 35%, support ({1,4}) = 40%, support ({3,5}) = 45%. Thus the frequent 2 itemsets are :{{1,3}, {1,4}, {3,5}} as their supports are equal and above 35%.

The traverse of graph as if their path is to determine frequent 3 itemsets among three nodes{i,j} and {j,k} then the set {i,j,k} will be frequent 3 itemsets. Here, in this example, {{1,2,5}} is the only frequent 3 itemsets. As there are no extra edges, by assigning 20% as a new value to the minimum support threshold, the frequent 3 itemsets will be: {1,3,5}as shown in table IV and the support for each candidate itemset of length 3, as follows: support ({1,3,5}) = 25%. Thus the frequent 3 itemsets are :{1,3,5} as their supports are equal and above 25%. Finally the algorithm terminates.

TABLE IV. FREQUENT ITEMSETS 2 FROM FREQUENT ITEMSETS 1

Item No. / TID	{1,2}	{1,4}	{2,5}
T7	L	L	L
T9	L	L	L
T1	L	L	L
T15	L	L	L
T17	L	L	L
T19	L	L	L
T4	L	H	L
T6	L	L	H
T8	L	H	L
T14	L	H	L
T12	H	L	L
T18	H	L	L
T3	H	L	H
T10	L	H	H
T11	H	L	H
T20	L	H	H
T2	L	H	H
T16	H	L	H
T13	H	H	H
T5	H	H	H

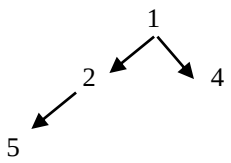

Fig 3 A simple directed graph to display frequent k-itemsets, k>=2

In this regard, as the database contains hundreds and thousands of transactions database and different items, constructing only one graph is not suitable for this work: So construct different graphs for each cluster and find from this graph all frequent itemsets, then combine the subsets of frequent itemsets together to get the whole set of frequent itemsets, and this technique is scalable with all transactions databases of different sizes.

TABLE V. FREQUENT ITEMSETS 3 FROM FREQUENT ITEMSETS 1

Item No. / TID	{1,2,5}
T7	L
T9	L
T1	L
T15	L
T17	L
T19	L
T4	L
T6	L
T8	L
T14	L
T12	L
T18	L
T10	L
T20	L
T2	L
T3	H
T11	H
T16	H
T13	H
T5	H

V. CONCLUSION

This paper proposes the use of the graph database for pre-processing, after the whole transaction database is divided into partitions of variable sizes. Each cluster is considered one at a time by loading the first cluster into memory and calculating large itemsets and the corresponding support counts. Then the second cluster is loaded additionally and cumulative support count is then derived for the second clustered large itemsets. This process is continued for the entire set of clusters and finally the whole large itemsets and the corresponding cumulative support counts. This approach reduces main memory requirement since it considers only a small cluster at a time and hence it is scalable for any large size of the database.

REFERENCES

[1] Willi Klosgen and Jan M. Zytkow, "Hand Book of Data Mining and Knowledge Discovery", Oxford University Press, 2002.

[2] Jiawei Han, Micheline Kamber, "Data Mining Concepts and Techniques", Morgan Kauffman Publishers, San Francisco 2001.

[3] Rakesh agrawal, Tomasz Imielinski and Arun N. Swami, "Data Mining a Performance Perspective", IEEE Transactions on Knowledge and Data Engineering, 1993,Vol.5,pp.914-925.

[4] Rakesh Agrawal and Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", Proceedings of the Twentieth International Conference on Very Large Databases, Santiago, Chile, 1994,pp.487-499.

[5] Jiawei Han, Yongjian Fu, "Mining Multiple Level Association Rules in Large Databases", Proceeding of 1995 Int'l Conf. on Very Large Data Bases (VLDB'95), pp. 420-431, Zurich, Switzerland, 1995.

[6] R. Agrawal, T. Imilienski, A. Swami, "Mining association rules between sets of items in large databases", Proceedings of the ACM SIGMOD Int'l Conf. on Management of Data, Washington, DC, 1993 pp. 207–216.

[7] Han, J., Pei, J., Yin, Y, "Mining frequent Patterns without Candidate Generation" In: ACM-SIGMOD, Dallas , 2000.

[8] F. Berzal, J.C. Cubero, N. Marin, J.M. Serrano, TBAR: An Efficient Method for Association Rule Mining in Relational Databases, Elserier Data and Knowledge, Engineering , pp. 47–64, 2001.

[9] D.W. Cheung, J. Han, V.T. Ng, A.W. Fu, Y. Fu, A fast distributed algorithm for mining association rules, Proceedings of Int'l Conf. on PDIS'96, Miami Beach, Florida, USA, 1996.

[10] R. Agrawal, R. Srikant, "Mining sequential patterns", Proceedings of the 11th Int'l Conf. on Data Engineering (ICDE), 1995.

[11] Ben Franklin, Genealogical Data Mining, 2006.

[12] Show-Jane Yen and Arbee L.P.Chen, "A Graph based Approach for Discovering Various Types of Association Rules", IEEE Transaction on Knowledge and Data Engineering, Vol.13, No.5, pp. 839-845, Sep/Oct. 2001.

[13] R.Srikant and R.Agrawal, "Mining Generalized Association rules", In Proc. Of the 21st Int'l Conf. on Very Large Databases, pp.407-419, Zurich, Switzerland, 1995.

[14] Cakmak. A, Ozsoyoglu. G, "Taxonomy superimposed graph mining", in proceedings of the 11th Int'l Conf. on Extending Database Technology, Advances in Database Technology, ACM Int'l Conf. Proceeding Series, Vol.261, pp.217-228,2008.

[15] Silvescu. A, Caragea D, Anna Atramentov Graph Databases Artificial Intelligence Research Laboratory, Department of Computer Science Iowa State University Ames, Iowa, 2007.

[16] Sadhana Priyadarshini and Debahuti Mishra, "A Hybridized Graph Mining Approach",in proceeding of ICT2010,CCIS 101, pp.356-361,2010.

[17] Haiwei Pan, Xiaolei Tan, Qilong Han, Guisheng Yin, "An Effective Algorithm Based on Association Graph and Matrix for Mining Association Rules", in Proceeding of IEEE2010, 978-1-4244-6977-2/10,2010.

[18] Muthukumar A, Nadarajan R, "Efficient and Scalable Partition based Algorithms for Mining Association Rules", Academic Open Internet Journal ISSN 1311-4360, Vol.19,2006.

AUTHOR PROFILE

Dr.N.Jayaveeran has pursued his Ph.d from Bharathidasan university, Tiruchirappali-24, Tamil Nadu,and he is currently working as Associate Professor and Head, PG and Research Department of Computer Science, Khadir Mohideen College, Adirampattinam, Tamil Nadu. He has published several research articles in international and national journals. His area of research includes Roughset Theory, Data mining, Network Security, Bigdata analysis and cloud computing. He has 29 years of teaching experience and he is currently guiding eight research scholars.

Mr. D. Senthil Kumar pursued Master Degree of Computer Science in 1999 from Bharathidasan University, Tiruchirapalli - 24, Tamil Nadu and Master of Philosophy in 2007 from Bharathidasan University, Tiruchirappali-24, Tamil Nadu. He is currently pursuing PhD in Bharathidasan University and he is currently working as Lecturer in the Department of Computer Science in Bharathidasan University Model College, Aranthangi, Tamil Nadu. He has published several papers in various conference and his research work is mainly focused on Data mining and Cloud computing. He has 14 years of teaching experience.