autonomy in terms of right to change schema or software. A homogeneous DDBMS appears to the user as a single system. The homogeneous system is much easier to design and manage.

A heterogeneous database system is an automated (or semi-automated) system for the integration of heterogeneous, disparate database management systems to present a user with a single, unified query interface. Heterogeneous database systems (HDBs) are computational models and software implementations that provide heterogeneous database integration.

4. Related Work

Association rule mining finds interesting associations and/or correlation relationships among large sets of data items. Association rules show attributes value conditions that occur frequently together in a given dataset. The market basket analysis used association rule mining in distributed environment.

Association rule mining is used to find rules that will predict the occurrence of an item and based on the occurrences of other items in the transaction [10], search patterns gave association rules where the support will be counted as the fraction of transaction that contains an item X and an item Y and confidence can be measured in a transaction the item i appear in transaction that also contains an item X.

The Apriori Algorithm proposed to finds frequent items in a given data set using the ant monotone constraint. Apriori is an influential algorithm in market basket analysis for mining frequent item sets for Boolean association rules. This algorithm in [13] contains a number of passes over the database. During pass k, the algorithm finds the set of frequent itemsets Lk of length k that satisfy the minimum support requirement.

Apriori is designed to operate on databases containing transactions. The purpose of the Apriori Algorithm [11] is to find associations between different sets of data. It is sometimes referred to as "Market Basket Analysis". Each set of data has a number of items and is called a transaction. The output of Apriori is sets of rules that tell us how often items are contained in sets of data. Association rule mining is used to find rules that will predict the occurrence of an item and based on the occurrences of other items in the transaction, search patterns gave association rules where the support will be counted as the fraction of transaction that contains an item X and an item Y and confidence can be measured in a transaction the item i appear in transaction that also contains an item X.

Support (s): - Fraction of transactions that contain both X and Y
Support (X->Y) = P (X∪Y)/T

Confidence(c): - Measure show often items in Y appear in transactions that contain X.
Confidence (X->Y) = Support (X∪Y) / Support (X)

Association rules are created by analyzing data for frequent if/then patterns and using the criteria support and confidence to identify the most important relationships. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true.

Privacy preserving distributed mining of association rule for a horizontally partitioned dataset across multiple sites are computed as follows where I = {i1, i2, .in} be a set of items and T = {T1, T2,…Tn} be a set of transactions where each T I i Í . A transaction Ti contains an item set X Í I only if i X Í T . An association rule implication is of the form X_Y (X ÇY =0) with support S and confidence C if S% of the transactions in T contains X ÈY and C% of transactions that contain X also contain Y. In a horizontally partitioned database, the transactions are distributed among n sites. Support (X $\_Y$) = probe (X ÈY) /Total number of transaction.

The global support count of an item set is the sum of all local support counts.
Support
g(X)=Support1(x)+Support2(x)+……+
Support n(x).

Confidence of rule (X $\_Y$) = Support (X ÈY) / Support(X)
The global confidence of a rule can be expressed in terms of the global support.

Confidence g (X $\_Y$) = Support g (X ÈY) / Support g(X) The basis of this algorithm is the Apriori algorithm that uses K-1 frequent sets.

5. The Fast Distributed Mining Algorithm

The protocols are based on the Fast Distributed Mining (FDM) algorithm like in [2], which is an unsecured distributed version of the Apriori algorithm. Its main idea is that any s-frequent itemset must be also locally s-frequent in at least one of the sites. Hence, in order to find all globally s-frequent itemsets, each player reveals his locally s-frequent itemsets and then the players check each of them to see if they are s-frequent also globally. The stages of the FDM algorithm are as follows:

1) Initialization: It is assumed that the players have already jointly calculated Fk−1 s .The goal is to proceed and calculate $F^k{}_s$.
2) Candidate Sets Generation: Each Pm generates a set of candidate k- itemsets $B^{k,\ m}{}_s$ out of $F^{k-1,m}{}_s \cap F^{k-1}{}_s$ — the (k − 1)-itemsets that are both globally and locally frequent, using the Apriori algorithm.
3) Local Pruning: For each X ∈ $B^{k,\ m}{}_s$. Pm computes supp m(X) and retains only those itemsets that are locally s-frequent. We denote this collection of itemsets by $C^{k,\ m}{}_s$.
4) Unifying the candidate item sets: Each player broadcasts his $C^{k,\ m}{}_s$ and then all players compute $C^k{}_s := S^M{}_{m=1}\ C^{k,m}{}_s$
5) Computing local supports: All players compute the local supports of all itemsets in $C^k{}_s$.

6) Broadcast Mining Results: Each player broadcasts the local supports that he computed. From that, everyone can compute the global support of every itemset in $C^k{}_s$. Finally, $F^k{}_s$ is the subset of $C^k{}_s$ that consists of all globally s-frequent k-itemsets.

With the existence of many large transaction databases, the huge amounts of data, the high scalability of distributed systems, and the easy partition and distribution of a centralized database, it is important to investigate efficient methods for distributed mining of association rules. This study discloses some interesting relationships between locally large and glob-ally large itemsets and proposes an interesting distributed association rule mining algorithm, FDM (Fast Distributed Mining of association rules), which generates a small number of candidate sets and substantially reduces the number of messages to be passed at mining association rules. Our performance study shows that FDM has a superior performance over the direct application of a typical sequential algorithm. Further performance enhancement leads to a few variations of the algorithm.

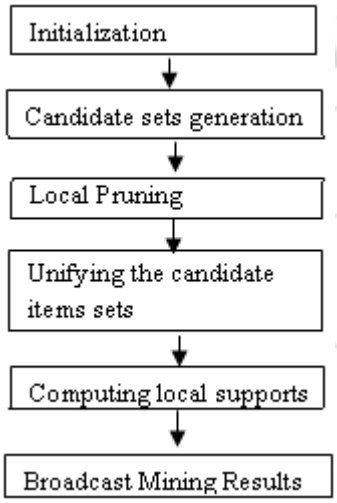

Figure: The Process of FDM Algorithm

6. Conclusion

The problem of computing association rules within a scenario of homogeneous database. Assume that all sites have the same schema, but each site does not have information on different entities. The goal is to produce association rules that hold globally while limiting the information shared about each site. Many protocols have been implemented. In this, focus is based on horizontal partitioned distributed data through a popular association rule mining technique. Protocols exploit the fact that the underlying problem is of interest only when the number of players is greater than two.

References

[1] Tamir Tassa,"Secure mining of association rule in horizontally distributed databases" ,IEEE trans. Konwledege and Data Engg.,Vol. 26, no.2, April 2014J.
[2] M. Kantarcioglu and C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data," IEEE Trans. Knowledge and Data Eng., vol. 16, no. 9, pp. 1026-1037, Sept. 2004.
[3] Krishna Pratap Rao, Adesh chaudhary, Prashant johri "Elliptic Curve Cryptography Based Algorithm for Privacy Preserving in Data Mining", International Journal for research in Applied Science and
[4] P. Jagannadha Varma, Amruthaseshadri,.M. Priyanka, M.Ajay Kumar, B.L.Bharadwaj Varma, " Association Rule Mining with Security Based on Playfair Cipher Technique" (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (1), 201
[5] Prof. Geetika. Narang, Anjum Shaikh, Arti Sonawane, Kanchan Shegar, Madhuri Andhale," Preservation Of Privacy In Mining Using Association Rule Technique", International Journal of Scientific & Technology Research, Volume 2, Issue 3, March 2013
[6] Zhi Liu,Tianhong Sunand Guoming Sang," An Algorithm of Association Rules Mining in Large Databases Based on Sampling ", International Journal of Database Theory and Application Vol.6, No.6 , 2013
[7] Priyanka Asthana, Anju Singh , Diwakar Singh," A Survey on Association Rule Mining Using Apriori Based Algorithm and Hash Based Methods ", International Journal of Advanced Research in Computer Science and Software Engineering. Volume 3, Issue 7, July 2013
[8] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," in The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, July 23-26 2002,
[9] A.C. Yao. Protocols for secure computation. In FOCS, pages 160–164, 1982.
[10] Sotiris Kotsiantis, Dimitris Kanellopoulos Association Rules Mining: A Recent Overview GESTS International Transactions on Computer Science and Engineering, Vol.32 (1), 2006, pp. 71-82
[11] T.Kartikeyan and N.Ravikumar A Survey on Association Rule Mining International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 1, January 2014
[12] Shiny. I.S , S. Gayathri,"Secure Multiparty Computation and Privacy Preserving Data Sharing with Anonymous ID Assignment", International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 International Conference on Humming Bird ,01st March 2014
[13] Chitteni Siva, Selvi Secure Mining of Association Rules in Horizontally Distributed Databases International Journal of Computer Science and Mobile Computing A Monthly Journal of Computer Science and Information Technology IJCSMC, Vol. 3, Issue. 4, April 2014, pg.1079 – 1082