

# Eclat with Large Data base Parallel Algorithm and Improve its Efficiency

Rana Ishita  
Department of Computer  
Science & Engineering  
Parul Institute of  
Engineering & Technology,  
Limda Gujarat, India

Amit Rathod  
Department of  
Information Technology  
Parul Institute of  
Engineering & Technology,  
Limda Gujarat, India

## ABSTRACT

Data mining is the finding the hidden pattern from the huge amount of data. In Data mining, the definition of association rule finds interesting association or correlation relationships among a large set of data items. Association rule mining finding frequent pattern, correlations among the items or object in transactional database or relational database. Basic idea is that the search tree could be divided into sub process of equivalence classes. And since generating item sets in sub process of equivalence classes is independent from each other, we could do frequent item set mining in sub trees of equivalence classes in parallel. So the straightforward approach to parallelize Eclat is to consider each equivalence class as a data. We can distribute data to different nodes and nodes could work on data without any synchronization. Even though the sorting helps to produce different sets in smaller sizes, there is a cost for sorting. Our Research to analysis is that the size of equivalence class is relatively small and this size also reduces quickly as the search goes deeper in the recursion process. Base on time using more than using data we can handle large amount of data so first we develop Eclat algorithm then develop parallel Eclat algorithm then compare with using same data with respect time with the help of support and confidence.

## Keywords

Association rule, Frequent Item, Data Mining, Eclat Algorithm,

Parallel Approach, Parallel Eclat

## 1. INTRODUCTION

Procedure of Data mining is an evaluating data from different perspectives and summarizing into useful information. It has evolved in to a wide area of research because of the theoretical challenges and practical applications were associated with the problem of realizing the interesting and earlier unknown knowledge from real world databases. There are many technique of data mining such as Classification, Clustering, Association rule mining, Regression etc.

### 1.1 Issues in Data Mining

There are many issues from the different Data mining algorithms.

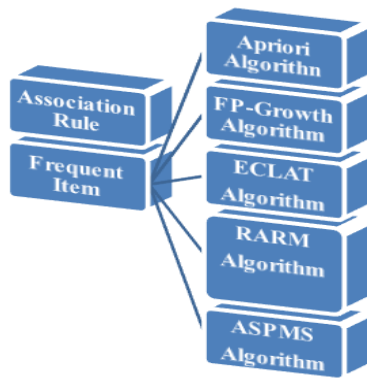
- Noisy data and missing values
- Static data
- Sparse data
- Dynamic data
- Interestingness
- Heterogeneity
- Algorithm efficiency
- Size and complexity of data

### 1.2 Association Rule

Association rule mining itself is right now one of the important topics of research related to Data mining. Still lots of research have been done; and lots of remains to be still explored. Association rule can create analyzing data for frequent pattern using the criteria Support & Confidence to identify the relationship. Support is indicating of how frequently the item appears in the database. Confidence indicates the number of time has been found. In Association rule mining; frequent item set mining include application like market basket analysis, modular fragment mining, web link analysis etc. It aims at finding regularities in the shopping behavior of customers of supermarkets, mail-order companies, on-line shops etc. There are two methods for generating frequent items sets: Sequential methods and Parallel methods. There are various types of algorithm are used in frequent item sets; Apriori, Eclat, FP-Growth etc. There are two main goal of to generate rule; frequent item set generation and rule generation

### 1.3 Frequent Item set Mining

Finding frequent item sets the main problem is; item sets are contained in a user-specified minimum number of transactions and that there are so many possible sets, which renders naive approaches infeasible due to their un-acceptable execution time; frequent pattern mining has been an important subject matter in data mining from many years. A remarkable progress in this field has been made and lots of efficient algorithms have been designed to search frequent patterns in a transactional database. Frequent pattern mining can be used in a variety of real world applications. It can be used in market basket analysis, super markets for selling, product placement on shelves, for promotion rules and in text searching.



**Fig 1: Different types of Algorithm are used in association rule mining<sup>[1]</sup>**

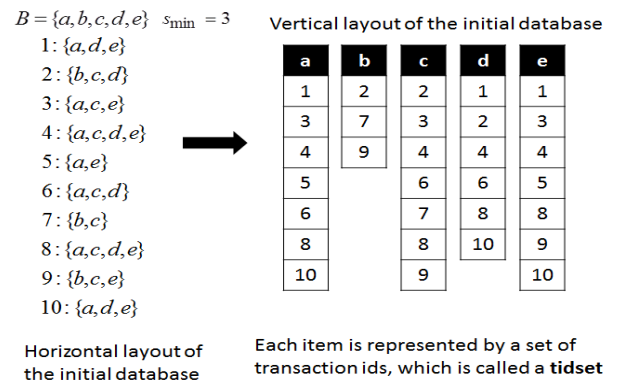
In Association rule to count supports of candidates, we need to go through transactions in the transaction database and check if transactions contain candidates. Since the transaction database is usually very large, it is not always possible to store them into main memory. Furthermore, to check if a transaction containing an item set is also a non-trivial task. So an important consideration in frequent item set mining

## 2. RELATED WORK

**Table 2.1 Related Work**

Sr. No.	Year/Journal	Title	Method	Advantages
1.	2014, Science Direct	Frequent Pattern mining algorithm finding associated frequent patterns for Data Streams: A Survey	-Apriori -FP Growth -RARM -Eclat -ASPMS	-Eclat is better than Apriori. - RARM performs better than Apriori and FP growth algo in terms of execution time. - FP growth requires less database scan as compared to RARM algo.
2.	2014, Science Direct	Study on the Method of Road Transport Management Information Data mining Based on Pruning Eclat Algorithm And Map Reduce	-Apriori Algorithm -Eclat Algorithm	-Improve Efficiency
3.	2013, IEEE	MREclat: an Algorithm for Parallel Mining Frequent Item sets	-Parallel mining algorithm -Eclat algorithm	- Improve the efficiency of mining frequent item set from massive datasets, parallel algo MREclat. -MREclat algo has high scalability and good speed.
4.	2013, Springer	Efficient mining frequent item sets algorithm	-Apriori -CTFI -BCTF	-The candidate item sets generation with help of support count and support count problem so it does not need to generate candidate item sets
5.	2013, IJIRCCE	An Efficient Algorithm for Mining Frequent Item Sets in Data Streams	-Eclat -RARM	-Execution time -No of frequent Items
6.	2010, IEEE	A new Algorithm for Discovering Association Rules	-ECLAT -Apriori -LOGECLAT	-LogEclat algo indicates decrease the time of generating new database for frequent item set and improve the efficiency

algorithms is the representation of the transaction database to facilitate the process of counting support. There are two layouts that algorithms usually employ to represent transaction databases: horizontal and vertical layout.



**Fig 2: Horizontal and Vertical layout**

7.	Member, IEEE	A Transaction Mapping Algorithm for Frequent Item Sets Mining	-Transaction mapping algo -Eclat algo -FP Growth algo	-Frequent Itemset are found in depth-first order along a lexicographic tree as done in Eclat algorithm.  -To generate a transaction tree.  -Minimum support is large, there are more repetitive patterns, and the algorithm runs fast.
----	--------------	---------------------------------------------------------------	-------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

### 3. ECLAT ARCHITECTURE

Eclat means Equivalence class clustering and bottom up lattice traversal. Eclat algorithm is the best known basic algorithm for mining frequent item sets in a set of transaction. Eclat represents the data in vertical data format. Architecture of existing system is shown in figure.

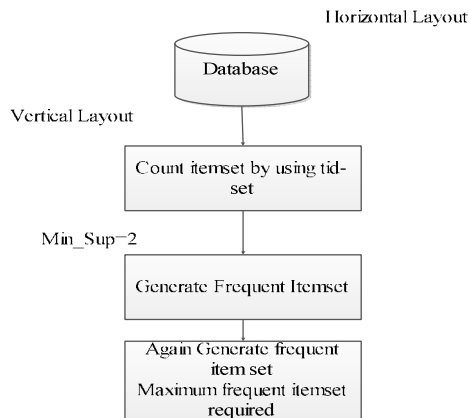


Figure 3.1 Architecture of Eclat Algorithm

#### 3.1.1 Eclat Algorithm

Eclat algorithm is basically a depth-first search algorithm using set intersection. Algorithm Of existing system is shown in figure. Explanation of algorithm step by step is below.

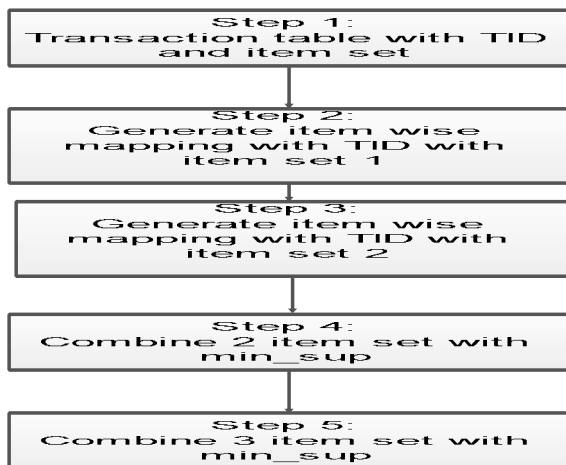


Fig 3.2 Eclat Algorithm

#### Advantages

- Eclat algorithm is faster than Apriori algorithm.

- Databases are no need to scan each item set is created as support count information will be obtained from previous item set.
- Using method depth-first search reduces memory requirements.

#### Disadvantages

- The T-id sets can be long therefore expensive to manipulate.

#### Limitations:

- Eclat algorithm does not take full advantage of Apriori property to reduce the no of candidate item sets explored during frequent item set generation.”
- “In Eclat algorithm, the computation is done by intersection of the T-id sets of the Corresponding k-item sets. “
- “In Eclat algorithm, the classes are solved from bottom to top, according to a reverse lexicographic order.

### 4. PARALLEL ECLAT ARCHITECTURE

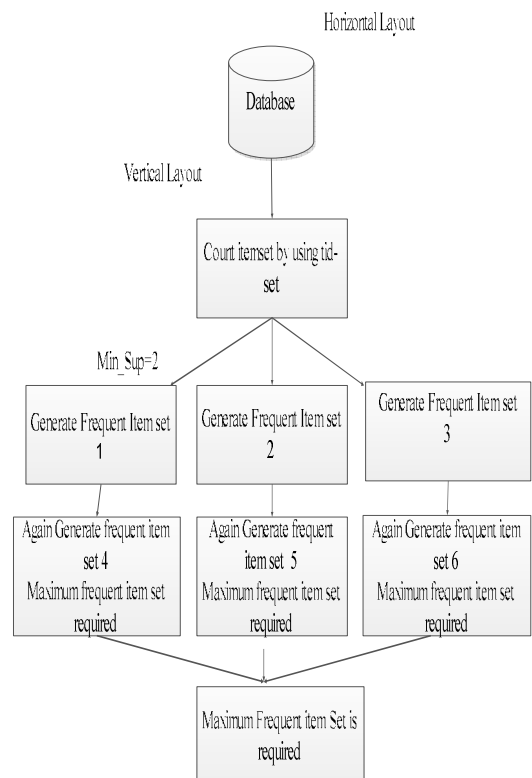


Fig 4.1 Parallel Eclat Architecture

4.1.1 Experimental Parameters

- Record
- Support

5. EXPERIMENTAL RESULT OF ECLAT AND PARALLEL ECLAT ALGORITHM

Implementation of Eclat and Parallel Eclat algorithm would be done using c#.net language with the help of Microsoft Visual Studio 10 and for database SQL Server Management Studio is used. The Data set are used in real data for Agriculture Data.

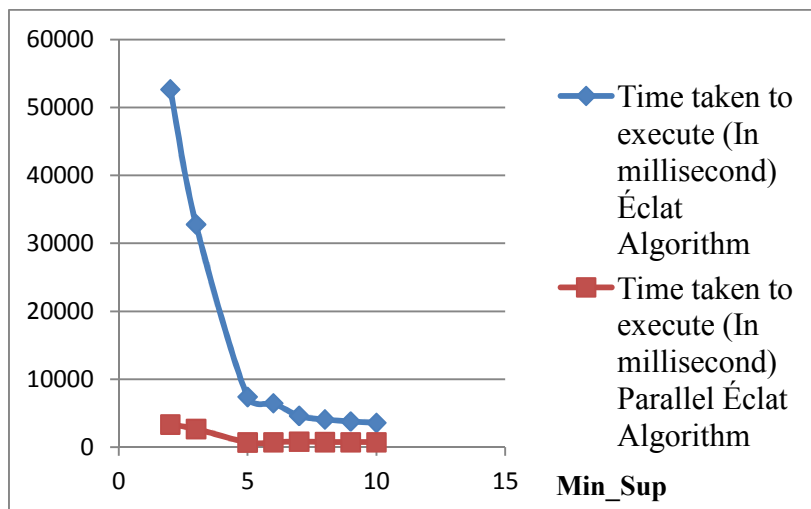
ID	ItemId	ItemName
19	A	Wheat
20	B	Rice
21	C	Rain
22	D	Urea
23	E	Pulse
24	F	Peas
25	G	Carrat
26	H	Phosphate
27	I	Cold
28	J	Vegetables
29	K	Beetroot
30	L	Cucumber
31	M	Water
32	N	Bittergourd

Fig 5.1 Characteristics of experiments Data set

Comparison between Éclat algorithm and Parallel Éclat Algorithm with No of Support count and Its Execution Time

Table 5.1 Comparison Eclat & Parallel Eclat Algorithm with min\_sup count

Number of Support Min_sup	Time taken to execute (In millisecond) Éclat Algorithm	Time taken to execute (In millisecond) Parallel Éclat Algorithm
2	52649	3353
3	32779	2678
5	7406	695
6	6474	729
7	4609	817
8	4082	752
9	3797	734
10	3592	732



5.1.1 Comparison between Eclat and Parallel Eclat Algorithm

Table 5.2 Comparison Eclat and Parallel Eclat Algorithm

Factor	Eclat Algorithm	Parallel Eclat Algorithm
Process	Vertically	Vertically
Database to use	Small Data	Large Data
Execution Time	Maximum time as required.	Take minimum time as Required.
Efficiency	It can be handle data Vertically.	It can be Handle data Parallel.

6. CONCLUSION AND FUTURE ENHANCEMENT

The survey of various frequent item set algorithm is done with each having advantages, disadvantages and limitations over different parameters. The main motivation for frequent item set generation to increase the efficiency and scalability. The proposed work is outline with the basic ideas of Eclat using new parallel approach for vertical data layout. We want to increase efficiency, scalability compare to Eclat algorithm or better. Eclat can be handle small data; but proposed algorithm handle large amount of data. As per our survey the support count was plays an important role in frequent pattern mining algorithm, we proposed better algorithm based on the Parallel Eclat algorithm. Our Experiment result shows the improve efficiency of the execution time and its perform better result than the Existing algorithm. The future work will be focus on different Mining techniques will be very important in order to

conduct advanced analysis and such as determining trends to finding interesting patterns, on streaming data. Data streams data are from various sources, and it has much confidential information also so we can protect these confidential data by applying a privacy techniques.

## 7. REFERENCES

- [1] Shamila Nasreen, Muhammad Awais Azamb, Khurram Shehzad, Usman Naeem, Mustansar Ali Ghazanfar “Frequent Pattern mining algorithm finding associated frequent patterns for Data Streams: A Survey” 2014, Science Direct
- [2] Xiaofeng Zheng <sup>a</sup>, Shu Wang <sup>a\*</sup> “Study on the Method of Road Transport Management Information Data mining Based on Pruning Eclat Algorithm and Map Reduce” 2014, Science Direct
- [3] Zhigang Zhang, Genlin Ji\*, Mengmeng Tang “MREclat: an Algorithm for Parallel Mining Frequent Item sets” 2013, IEEE
- [4] Marghny H. Mohamed • Mohammed M. Darwieesh “Efficient mining frequent item sets algorithm” 2013, Springer
- [5] Dr. S.Vijayarani, Ms. P. Sathya “An Efficient Algorithm for Mining Frequent Item Sets in Data Streams” 2013, International Journal of Innovative Research in Computer and Communication Engineering
- [6] Kan Jin “A new Algorithm for Discovering Association Rules” 2010, IEEE
- [7] Mingjun Song, and Sanguthevar Rajasekaran “A Transaction Mapping Algorithm for Frequent Item Sets Mining” Member, IEEE
- [8] R. Agrawal, T. Imielinski, and A.N. Swami, "Mining association rules between sets of items in large databases," in ACM SIGMOD International Conference on Management of Data, Washington, 1993.
- [9] R. Agrawal, and R. Srikant, "Fast algorithms for mining association rules," in 20th International Conference on Very Large Data Bases, Washington, 1994.
- [10] J. Han, J. Pei, and Y. Yin, "Mining frequent patterns without candidate generation," in ACM SIGMOD International Conference on Management of Data, Texas, 2000.
- [11] M.J. Zaki, S. Parthasarathy, M. Ogihara, and W. Li, "New algorithms for fast discovery of association rules," in Third International Conference on Knowledge Discovery and Data Mining, 1997.
- [12] Avani M. Sakhapara, Bharathi H. N. “Comparative Study of Apriori Algorithms for Parallel Mining of Frequent Itemsets,” International Journal of Computer Applications, 2014.
- [13] Tipawan Silwattananusarn1 and Assoc.Prof. Dr. KulthidaTuamsuk2 “Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012”, International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012.
- [14] Han, J., Kamber, M.: “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, 2006.
- [15] Wan Aezwani Bt Wan Abu Bakar, Zailani B Abdullah, Md. Yazid B. Md Saman, Masita@Masila Bt abd Jalil, Mustafa B. Man “Vertical Association Rule Mining :Case Study Impkementation with Relational DBMS. “ ISTMET 2015.
- [16] <http://www.google.com>
- [17] <http://www.wikipedia.com>