



SDI Review Form 1.6

Journal Name:	Asian Journal of Research in Computer Science
Manuscript Number:	Ms_AJRCOS_41045
Title of the Manuscript:	A PARALLEL APRIORI -TRANSACTION REDUCTION ALGORITHM USING HADOOP -MAPREDUCE IN CLOUD
Type of the Article	Original Research Article

General guideline for Peer Review process:

This journal's peer review policy states that **NO** manuscript should be rejected only on the basis of '**lack of Novelty**', provided the manuscript is scientifically robust and technically sound. To know the complete guideline for Peer Review process, reviewers are requested to visit this link:

(<http://www.sciencedomain.org/page.php?id=sdi-general-editorial-policy#Peer-Review-Guideline>)



SDI Review Form 1.6

PART 1: Review Comments

	Reviewer's comment	Author's comment (if agreed with reviewer, correct the manuscript and highlight that part in the manuscript. It is mandatory that authors should write his/her feedback here)
Compulsory REVISION comments	<p>Abstract</p> <p>Apriori algorithm is a classical algorithm of association rule mining and widely used for generating frequent item sets. However, the original Apriori algorithm has many drawbacks such as it needs to scan the dataset many times to discover all frequent itemset and generate huge number of candidate itemset. To overcome these limitations, researchers have made a lot of improvements to the Apriori such as candidate generation, without candidate generation, transaction reduction, partitioning, and sampling. When it comes to mine massive data, these algorithms failed to prove efficiency because limitation of the processing capacity, storage capacity, and main memory constraints.</p> <p>The aim of this study is to propose a parallel Transaction Reduction MapReduce Apriori algorithm (TRMR-Apriori) which is reduce unnecessary transaction values and transactions from the dataset in parallel manner to overcome above problems. The experiments show that TRMR-Apriori is able to achieve better execution time to discover frequent itemset those of previous sequential ARM algorithms such as Apriori, AprioriTid, Eclat, and FP-Growth and the previous parallel algorithms such as PApriori, MRApriori, and Modified Apriori with different condition on homogeneous computing environment using Hadoop-MapReduce platform in cloud. Overall, the TRMR-Apriori shows the strength to extract the frequent itemset from massive dataset in cloud.</p> <p>The abstract shows that the researcher wants to use Apriori in a new way by making improvements. The aim of the study shows to use in a parallel environment.</p> <p>One suggestion is that Apriori is not proposed for the parallel environment. So I suggest that not to use the word drawback especially in abstract section.</p> <p>Introduction</p> <p>Data mining and knowledge discovery has emerged to extract useful, interesting, and unknown patterns and knowledge from huge amount of database. Association Rule Mining (ARM) is one of the most important and popular technique of data mining which find interesting correlation or association between set of items or attributes and also frequent patterns in large database [1,2].</p> <p>Parallel and distributed computing is a better solution to overcome the above problems. Many researches have been carried out for parallelizing the Apriori algorithm. Apriori parallel algorithms handles gigantic dataset on various platforms with different configurations. There are many major challenges such as to achieve efficacious load balancing, utilize total memory system, intended new algorithm for memory utilize, consider different data layout, produce effective parallel algorithm, reduce communication cost among processors, system failure, data recovery, simplify parallel programming issue, manage scalability, and high availability.</p> <p>Cloud systems which can be effectively employed to handle parallel mining since they provide scalable storage and processing services, as well as software platforms for developing and running data analysis environments. We exploit Cloud computing platforms for running big data mining processes designed as a combination of several data analysis steps to be run in parallel on Cloud computing elements.</p> <p>From literature study it is clear that parallel implementation of Apriori algorithm is more efficient and has good performance compare to the sequential Apriori algorithm. The use of Hadoop-MapReduce framework provides further advantages such as the Apriori algorithm requires frequent transaction set scanning; Hadoop provides HDFS storage with good parallel reading-writing characteristics, thus the time repeatedly read transaction set is greatly shortened compared with the traditional storage system, which makes it possible to mine large amount of data on Hadoop with Apriori algorithm. The calculation with Apriori algorithm can be regarded as process of counting, and this process is suitable for MapReduce model, so Apriori algorithm has the natural characteristics of MapReduce [22].</p> <p>Introduction is given in detail about the previous research and about the challenges of the field.</p> <p>Problem statement</p> <p>Researchers have been proposed new algorithms based on the Apriori algorithm and try to improve the efficiency by making some modifications, such as reducing the number of passes over the database; reducing the size of the database to be scanned in every pass; pruning the candidates by different techniques such as direct hash pruning, transaction reduction, partitioning, sampling, dynamic itemset counting, vertical dataset layout, and frequent pattern tree [8,17,21,23].</p>	



	<p>The researcher discuss that new algorithms were proposed based on Apriori and tell that they are going to do some thing new on the basis of the previous research.</p> <p>The objective of this study are</p> <p>To investigate the performance of different sequential ARM algorithms in cloud environment and parallel Apriori algorithm using Hadoop-MapReduce in cloud.</p> <p>To propose a transaction reduction algorithm for parallel Apriori based on HadoopMapReduce.</p> <p>The outcomes of this study are to set-up and configure the cloud environment to test big data analysis and improving parallel Apriori algorithm using transactional reduction method. Cloud Data Mining offers tremendous potential for analyzing and extracting the useful information in various fields of human activities: finance, banking, medicine, genetics, biology, pharmacy, and marketing. Cloud provides technology that can handle huge amounts of data, which cannot be processed efficiently and at reasonable cost using standard data mining technologies and techniques. Hadoop-MapReduce is the current trend in the field of huge data processing.</p> <p>Based on the algorithms and technologies developed by large Internet companies, there is a quite widespread ecosystem of solutions for processing and analysis of huge amounts of data. Proposed transaction reduction algorithm may increase the performance compared to earlier well developed versions of sequential and parallel versions Apriori algorithms. In this section the researcher tell that a new way is proposed in parallel environment.</p> <p>Methodology</p> <p>This study consists 6 phases to accomplish this research: problem identification, massive dataset collection (benchmark data), set-up and configure cloud environment, implement sequential algorithms in cloud, Test the proposed transaction reduction Apriori algorithm as well as existing some parallel Apriori in cloud environment., evaluation. In this part the researcher discuss about the different steps or phases of the research.</p> <p>PROPOSED TRANSACTION REDUCTION MAPREDUCE APRIORI ALGORITHM (TRMR-APRIORI)</p> <p>RESULTS AND DISCUSSION</p> <p>Experiments that have been carried out to verify the performance of the proposed parallel transaction reduction Apriori algorithm on Hadoop-MapReduce. The results of the experiments are presented in tables as well as appropriate graphs. The description and analysis about the performance improvement in the proposed method are compared against the performance of existing well known sequential algorithm as well as parallel Apriori on Hadoop-MapReduce problem researches. The results are divided into two categories, one comparison of the proposed method against sequential algorithms in cloud environment. The second category is the comparison of the proposed method against some parallel Apriori algorithm which were implemented on Hadoop-MapReduce in unique environment.</p> <p>The research discuss about the experiments and that the research is also compared to previous known algorithms.</p> <p>CONCLUSION This study attempted to examine the performance of extracting frequent itemset from massive dataset. The main aim of this study was discover whole frequent itemset from huge dataset on cloud environment using Hadoop-MapReduce parallel techniques. An algorithm proposed to extract entire frequent itemset using transaction reduction technique that is reduce unessential transaction values and transactions from the dataset in parallel manner. The transaction reduction is one of the best way to find the frequent itemset from massive dataset with efficiently. Our experimental results also exhibit the performance of our proposed parallel transaction reduction algorithm compared with existing sequential as well as parallel algorithms on cloud environment. The proposed Transaction Reduction MapReduce Apriori (TRMR-Apriori) algorithm was tested on Amazon Web Service (AWS) cloud environment. UNDER PEER REVIEW Moreover, the experiments results show that the proposed TRMR-Apriori algorithm was implemented to extract entire frequent itemsets from big dataset with effectively and efficiently by comparing existing sequential Apriori algorithms as well as parallel HadoopMapReduce Apriori algorithms on unique environment. This study mainly proposed an Apriori parallel algorithm using transaction reduction method. The proposed algorithm was experimented using Hadoop-MapReduce techniques on cloud environment. The implications of finding in the context of the research objectives addressed by this study are discussed in the following. To achieve the first objective, we tested selected existing well established sequential frequent itemset mining algorithms in association rule mining such as Apriori, AprioriTid, Eclat, and FP-Growth on cloud environment. In this test we analyzed the performance varying the minimum support (min_sup) value using selected all benchmark dataset. We was analyzed with existing Apriori parallel algorithms which based on the HadoopMapReduce platform with different conditions such as varying min_sup value, increase number of nodes, and replicated the transaction size to evaluate the existing algorithms performance. To achieve the second objective, the proposed TRMR-Apriori algorithm was</p>	
--	---	--



SDI Review Form 1.6

	<p>compared their performance with experimental results which were achieved by objective one such as existing well established sequential frequent itemset mining algorithms in association rule mining (Apriori, AprioriTid, Eclat, and FP-Growth). In this test we analyzed and compared the performance varying the minimum support (min_sup) value using selected all benchmark dataset. The proposed TRMR-Apriori algorithm was tested in different conditions such as varying min_sup value, increase number of nodes, and replicated the transaction size to evaluate the proposed algorithm performance. The proposed TRMR-Apriori algorithm was analyzed with existing Apriori parallel algorithms which based on the Hadoop-MapReduce platform. We implemented our proposed transaction reduction MapReduce algorithm (TRMR-Apriori) on Amazon web service (AWS) to evaluate the performance. We used EC2, S3, EMR services from AWS and uploaded selected and pre-processed bench mark dataset to EC2 and S3. We also used SPMF open source data mining library for the experimental results of existing versions of sequential Apriori algorithm such as Apriori, AprioriTid, FP-Growth, and Eclat on cloud environment.</p> <p>In this section aims and the obtained results were discussed. They discuss that how the results were generated and how the propose algorithm is efficient.</p> <p>REFERENCES 1. Fayyad, U., Piatetsky-Shapiro, G. & Smyth, P. 1996. The Kdd Process for Extracting Useful Knowledge from Volumes of Data. Communications of the ACM 39(11): 27-34. 2. Agrawal, R. & Srikant, R. 1994. Fast Algorithms for Mining Association Rules. Proceeding 20th international conference on very large data bases, VLDB, 487-499. 3. Han, J. & Kamber, M. 2006. Data Mining, Southeast Asia Edition: Concepts and Techniques. Morgan kaufmann. 4. Patel, A. B., Birla, M. & Nair, U. 2012. Addressing Big Data Problem Using Hadoop and Map Reduce. Engineering (NUICONE), 2012 Nirma University International Conference on, 1-5. 5. Park, J. S., Chen, M.-S. & Yu, P. S. 1995. Efficient Parallel Data Mining for Association Rules. Proceedings of the fourth international conference on Information and knowledge management, 31-36</p> <p>In last the references were given.</p> <p>Overall the study is a good contribution to the field of research.</p> <p>The word Case study is used inappropriately. May be the researcher is unaware about the concept of case study. It will be better to use other appropriate word rather then case study, if not possible then convert whole article as a case study after getting the real sense of case study concept</p> <p>The word drawback is not a suitable word in the abstract section for Apriori. As the algorithm was not proposed(Apriori) for the parallel environment.</p>	
Minor REVISION comments		
Optional/General comments		

Reviewer Details:

Name:	<i>Irshad Ullah</i>
Department, University & Country	<i>Pakistan</i>