



SDI Review Form 1.6

Journal Name:	Journal of Advances in Mathematics and Computer Science
Manuscript Number:	Ms_JAMCS_39635
Title of the Manuscript:	Improved FTWeightedHashT Apriori Algorithm on Big Data using Hadoop-MapReduce Model
Type of the 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	1. Abstract is too large. It should be precise. No need to write place, particular dataset name, and value of execution time, etc. in abstract. 2. After line 37, introduce the need of algorithms for big data. 3. Line 43, mention the study [6] is on map reduce. 4. There is not sufficient literature review. Manuscript is incomplete in many directions. Following section must be in such type of manuscripts: Introduction, Literature Review/Related Work, Proposed Work/Algorithms, Methodology (if any), Experimental Details/Setup, Results and Discussion, Conclusion. 5. Add more recent and relevant papers in related works. For examples, S. Singh, R. Garg, P. K. Mishra, Observations on factors affecting performance of MapReduce based Apriori on Hadoop cluster, IEEE International Conference on Computing, Communication and Automation, 2016, pp. 87-94. Y. Xun, J. Zhang, X. Qin, FiDooP: Parallel Mining of Frequent Itemsets Using MapReduce, in IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2016, vol. 46, no. 3, pp. 313-325. D. Apiletti, E. Baralis, T. Cerquitelli, P. Garza, F. Pulvirenti, Luca Venturini, Frequent Itemsets Mining for Big Data: A Comparative Analysis, Big Data Research 6. Section 2 is not with appropriate title. What is meant be experimental details here? 7. Instead of using only index number of papers, use author name too. For example, in line 68, it should be, Singh et. Al. [5] has 8. Line 73, experiments should be sequential environment. Line 74, suitable experimental setup should be Hadoop cluster. 9. Line 76, authors mentioned the particular value of time, but time varies on different type of cluster setup. 10. Line 87, correct the sentence. Also here, authors mentioned about weighted hash tree apriori and hash table trie from paper [5-6], these terms must be explained here or in introduction section. 11. Line 94, correct the sentence, dataset multi scans should be multi scans of dataset. 12. Line 96, FTWeightedHashT is proposed by authors, explain the naming convention, e.g. FT stands for? 13. Line 97, follows: a should be follows. A 14. Line 98, there should not be part A 15. Line 99, a Hadoop should be HDFS 16. Line 101, ith should be th (th as superscript) 17. Proposed Algorithms requires more explanations. 18. Line 104, explain the calculation of WPi. What is chapter 2 here? 19. Line 105, how using transaction filtering is sufficient, also correct the sentence. Line 110, Transaction filtering is proposed by authors? How? It is already in sequential implementation of apriori and map reduce implementation of apriori in following paper (Give citation to this paper). S. Singh, R. Garg, P. K. Mishra, Observations on factors affecting performance of MapReduce based Apriori on Hadoop cluster, IEEE International Conference on Computing, Communication and Automation, 2016, pp. 87-94. 20. Line 127, what is weight of itemsets? What is weighted support? How it is different form support of itemsets? 21. Lines 131 to 136, are not clear, they need more explanations. 22. In figure 1, suffixes of F and W are not visible. 23. Is only insertion of filter transaction in existing weightedHashT algorithm, a contribution of authors?	1. Abstract is re-written as required. 2. Done. 3. Done. 4. Manuscript Following sections had been edited as required. 5. Done. More recent and relevant papers in related works are added. 6. Done. The title is edited into "Proposed Work". 7. Done. Author name is added. 8. Done. The sentence is Clearfield. 9. There was no need to mention the results of their proposed work because the mentioned reference explains everything. 10. The sentence is corrected and the terms are explained in related works section. 11. The sentence is corrected. 12. Done. 13. Done. 14. "Part A" is removed. 15. Done. 16. The Equation property is disabled in the journal template. However, we fix it 17. Explanations of the proposed algorithm are supported. 18. There was no need to mention the calculation of Wpi here but explained in detail in chapter two of the thesis. Sorry. "as explained in chapter 2" is removed. 19. It explained and the sentence is corrected. We applied not proposed transition filtering technique so the word "proposed" is removed. 20. The term of weight is explained. 21. Noted in the text. 22. That's because the template is converted the equations to images. Suffixes are now almost visible. 23. No. our contribution is explained in the fourth step of the proposed work. The other contribution is using only one VM instead of 2 VMs. Look Proff. To be familiar with big data technologies like Hadoop is a contribution for me. We are in Yemen, War, no electricity, no good internet signal, no food, no safety...etcetera. Please do a search about yemen I have spent two years to mold this work, GOD KNOWS. You should encourage me that this work is born in Yemen 24. Done. 25. Simple explanation of the main areas and functions of Pseudo code are added 26. Done. 27. Done. 28. No. For that, the table is removed. And a good explanation about the results obtained are added. 29. two scans of our proposed algorithm is one for frequency of itemsets and the other scan is for generating the frequent itemset tree. three scans of WeightedHashT is one scan for frequency and one scan for



SDI Review Form 1.6

	24. Line 230, correct the sentence. 25. In pseudo code presented in figure 4, explain the used notations. Also the used functions like Cal_Weight(), generateTree() etc. 26. Section 3.3, need more explanation for the results obtained. 27. Line 301, 2170 ms time is obtained at what support/weighted support, on which dataset? On which dataset the results in table 1 obtained? 28. In table 1, three algorithm's execution times are compared. Have all three algorithms implemented by authors? 29. Explain the number of scans in each algorithm, why it is 2 or 3 or unknow? 30. The obtained time is in ms, use some larger dataset or at lower support so that time can be obtained in seconds and then compare.	generating the tree and one scan for obtain the frequent itemsets. Unknown scans of HashTableTrie because the paper did not mentioned the no. of scans. However, the table is already removed. 30. Done. The difference was clear.
Minor REVISION comments	1. Correct the sentence about big data, in line 22-23. 2. In line 34, ARM should be in parenthesis. Regression should be regression. 3. After line 37, give a brief description of ARM problem. 4. In line 42, no space will be in bracket. 5. Correct the sentence in line 56 6. Section 2.1, title should be: Design of Proposed FTWeithtedHashT Apriori Algorithm or Designing Proposed FTWeithtedHashT Apriori Algorithm 7. In section 3.1, give the references of datasets. In abstract, authors mentioned three datasets, but here only one dataset is used for evaluation? 8. In section 3.2, Line 290, In which particular paper 2 VMs have used? What is meant by one VM for Hadoop and one for eclipse?	1. Done 2. Done. 3. Done. A brief description is added. 4. Done. 5. Done. 6. Done. 7. Done. The origins of datasets are added. 8. for example, [6] used two VMs.
Optional/General comments	1. Manuscript need proof readings, a number of syntax and semantic errors must be corrected. 2. In keywords list, add the keyword- frequent itemset mining 3. Reference style is not uniform in references section. If journal specify format then follow that. 4. In reference 5, Journal name should be International Journal of Computer Applications. 5. References are not in order of use in text.	1. Manuscript is almost reviewed. 2. Done. The keyword is added. 3. Thank you for the advice. 4. Done. 5. Am sorry for that. *****

As per the guideline of editorial office we have followed VANCOUVER reference style for our paper.

Kindly see the following link:

<http://sciencedomain.org/archives/20>