FIUT Approach for Frequent Pattern Complex Query Management

V. S. Suresh Kumar

Assistant Professor, Department of Computer Science and Engineering, Nandha College of Technology,
Erode- 638052, Tamilnadu, India, Email:sureshkumar@nandhaetch.org

Abstract - To solve the scalability and load balancing challenges in the parallel mining algorithms for frequent item sets, overcoming by applying FiDoop. FiDoop incorporates the frequent items metric tree or FIU-tree rather than conventional FP trees, thus attaining compressed storage and evade the obligation to build conditional pattern base. Process improves the performance of FiDoop by equalization I/O load across information nodes of a cluster. Decomposition of M-Set occurs till M-2 terms. This process is repeatedly expend till the complete decomposition method is accomplished. The results reveal that for large low-dimensional datasets, FiDoop performs better than FiDoop-HD, where FiDoop-HD outperforms FiDoop for high-dimensional data processing. FiDoop-HD improves the performance of FiDoop when long item sets are decomposed. Experiments show our proposed solution is efficient and scalable.

Keywords - Big Data, Map-Reduce Framework, Parallel Mining.

I. INTRODUCTION

In the 21st century, it is increasingly inseparable from the network, people visit dozens or even hundreds of pages, or upload photos or speech every day, which makes the data on the network into a geometrical growth, and the traditional technical architecture has become increasingly unable to meet the current needs of the vast amounts of data. Therefore, researching large processing and storage become a lot of and a lot of well-liked these days

II. PRE-PROCESSING

As the size of the increasing data day by day, the frequent item set techniques are out of date. As a remedy to this issue FIUT algorithm is implemented on Map-Reduce framework. Here we using FIUT algorithm to avoiding building conditional patterns and to achieve compressed storage using the Hadoop framework. They work under the three job levels. Synthetic datasets are used for the Experiments.

III. FREQUENT ONE ITEM SETS GENERATION

The first level of Map-Reduce job is mining frequent item sets. A transaction database is partitioned into multiple input split files stored by the HDFS across multiple data nodes of a Hadoop cluster. A Number of the clerks is dead supported variety of input split. Each mapper look at the input split, all transaction stored in a pair of key by the record reader. Then, mapper computes the frequencies of items and generates local one-item sets. Next, these one-item sets with the same key emitted by different mapper are sorted and merged in a specific reducer, which further produces global one item sets. Finally, infrequent items are pruned by applying the minimum support and consequently, frequent item set are generated and stored in the form of pair as the output from the first Map-Reduce job. Thus its lead as input to the second Map-Reduce job in Fi-Doop.

IV. ALL K ITEMS GENERATION

The main aim of the second Map-Reduce job to enhance the second round of scanning on the database to cut back irregular items from each record. The second job marks Associate in nursing thing set as a k-item set if it contains k frequent items. All plotter of the second job takes transactions as input. Then, the mapper emits a set of pair, in which item sets are composed of the number of the items produced by pruning and the set of items. These pairs obtained by the second Map-Reduce job's mapper are combined and shuffled for the second job's reducers. After performing arts the mix operation, each reducer emits key/value pairs, where the key is the number of each item set and the value is each item set and its count. It is necessary to make sure that frequent things in every group action ought to retain their lexicographic order so as to facilitate successive part.

V. WORK FLOW OF FIUT BASED MAPREDUCE

Figure 1 shows the workflow of FIUT based Map reduce.

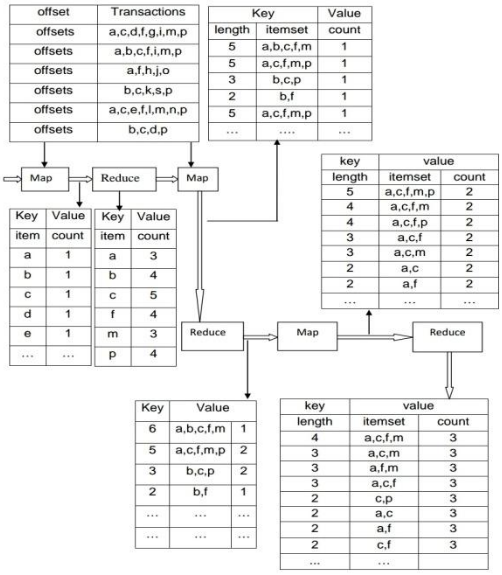

Fig. 1 Workflow of FIUT based Map reduce

VI. Frequent K Item Sets Generation

The current phase emphasis the: A) item-set decomposing B) FIU tree construction and C) mining Mapper's main goal is twofold: 1) small sized sets are obtained by decomposing the K-item sets by doing Map-reduce, the set lies in the range of 2 to k − 1 and 2) FIU tree have to be constructed with result of the same length. The last map reducer to highly scalable work. the Mappers process can be processed in parallel. Lead to improves data storage efficiency and I/O performance. The Map function of the job generates a set of key/value pairs, in which the key is the number of items in an item set and the value is an FIU-tree that is comprised of non leaf and leaf nodes. Non leaf nodes include item-name and node-link leaf nodes include item-name and its support. In doing so, item sets with the same number of items are delivered to a single reducer.

VII. Parameters for Evaluation

The performance for proposed methods can be evaluated by using the following parameters. Parameters which are considered for evaluating the experiments are:

1. Minimum support
2. Scalability

VIII. Minimum Support Count

Minimum support count plays the important role in mining frequent item sets. When we increase the minimum support threshold the running time of the proposed algorithm reduces. A small minimum support slows down the performance of the evaluated algorithms. This is because an increasing number of items satisfy the small minimum support when the minimum support is decreased; it takes an increased amount of time to process the large number of items. Figure 2 shows the execution time of four different minimum support counts

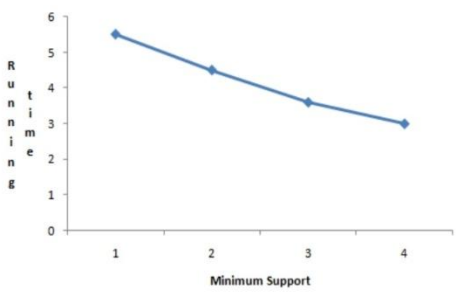

Fig. 2 Depiction of Minimum counts

IX. SCALABILITY

In this experiment, by evaluating the scalability proposed algorithm clearly shows the size of the data increases. The parallel mining process is decreased by a large data that has to be scanned twice. Thus its lead to large scanning time.

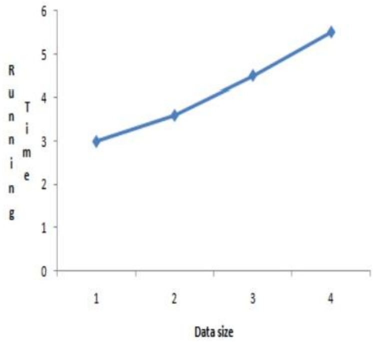

Fig. 3 Depiction of different data set

X. CONCLUSION AND FUTURE WORK

As application of the parallel mining of frequent itemsets using FIU Tree on Map-Reduce framework. We incorporate the Frequent Itemset metric Tree rather than conventional FP trees, thus attaining compressed storage and evade the obligation to build conditional pattern base. At the end of the third Map-Reduce job all frequent K itemsets are generated. To evaluate the performance of the proposed FIUT algorithm on Map-Reduce framework we use synthetic datasets in our experiments. As a future research direction the distributed cache technique is used to store the intermediate result of each MapReduce job which will significantly improves performance of parallel mining of frequent itemsets using FIUT on MapReduce framework.

REFERENCES

[1] Chang E.Y., Li H., Wang Y., Zhang D. and Zhang M. (2008), '*PFP: Parallel FP-growth for query recommendation*', in Proc. ACM Conf. Recommend.Syst., Lausanne, Switzerland, pp. 107–114.

[2] Chang W.L., Chen P.L. and Lin K.W. (2011), '*A novel frequent pattern mining algorithm for very large databases in cloud computing environments*', in Proc. IEEE Int. Conf. Granular Comput. (GrC), Kaohsiung, Taiwan, pp. 399–403.

[3] Chunyan H., Hong S., Huaxuan Z. and Shiping s. (2013), '*The study of improved FP-growth algorithm in MapReduce*' in Proc. 1st Int. Workshop Cloud Comput. Inf. Security, Shanghai, China, 2013, pp. 250–253S.

[4] Cong S, Han J., Hoeflinger J. and Padua D. (2005) '*A sampling-based framework for parallel data mining'*, in Proc. 10th ACM SIGPLAN Symp. Prin. Pract. Parallel Program, Chicago, IL, USA, pp. 255–265.

[5] Dean J. and Ghemawat S. (2008), '*MapReduce: Simplified data processing on large clusters'*, Commun. ACM, vol. 51, no. 1, pp. 107–113

[6] Nandagopal S, Karthik S, and Arunachalam VP., "*Mining of meteorological data using modified apriori algorithm*", European Journal of Scientific Research, Vol.47 No.2, pp.295-308, 2010

[7] Saveetha P and Arumugam S, "*Study on Improvement in RSA Algorithm and its Implementation*", International Journal of Computer & Communication Technology, Vol.3 No.6,pp.78, 2012.

[8] Nithya K, KalaivaaniPCD and ThangarajanR, "*An enhanced data mining model for text classification*", International Conference on Computing, Communication and Applications,pp.1-4,2012.

[9] DhivyaaCR, NithyaK and SaranyaM, "*Automatic detection of diabetic retinopathy from color fundus retinal images*", International Journal on Recent and Innovation Trends in Computing and communication,Vol.2 ,Issue 3,ISSN:2321-8169,2012.

[10] Vijayakumar, M., Prakash, S. and Parvathi, R.M.S. "*Inter Cluster Distance Management Model with Optimal Centroid Estimation for K-Means Clustering Algorithm*," WSEAS Transactions on Communications, Issue 6, Vol. 10, pp. 182-191, June 2011.

[11] Vijayakumar M and Parvathi RMS, "*Concept mining of high volume data streams in network traffic using hierarchical clustering*", European Journal of Scientific Research, Vol.39,No.2, pp:234-242,January 2010.

[12] Prakash S, Vijayakumar M and Parvathi RMS, "*A novel method of mining association rule with multilevel concept hierarchy*",International Journal Computer Application(IJCA),pp:26-29,2011.

[13] Saranya M and Nithya K, "*Campus Navigation and Identifying Current Location through Android Device to Guide Blind People*", International Research Journal of Engineering and Technology (IRJET), Vol.02,Issue : 08,Nov 2015.

[14] Nandagopal S, Arunachalam VP and Karthik"*A Novel Approach for Mining Inter-Transaction Item sets*", European Scientific Journal,Vol.8,No:14,pp:92-108,2012.