ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

A Hybrid Algorithm Combining Weighted and Hash T Apriori Algorithms in Map Reduce Model Using Data Analysis Cloud Platform

A.Arul Mary

Assistant Professor, Department of Computer Science, Anni Vailankanni Arts and Science College, Bharathidhasan University, Thanjavur, India

ABSTRACT: Data mining performs a primary role in knowledge analysis that blessings improvement of enterprise. In modern generation, distributed data mining turns into a first-rate studies space owing to distributed computing. digital distribution and distributed records mining receives its most vital impact on massive facts analysis with the assistance of cloud and grid computing. This studies paintings plays statistics mining in an exceedingly cloud surroundings and Hadoop platform. affiliation rule set of rules, Apriori, to guide Map Reduce parallel computing platform. we tend to enforce and assess the planned set of rules on Amazon EC2 Map Reduce platform. This work mines distributed statistics in an exceedingly cloud surroundings the utilization of hybrid apriori association rule algorithmic rule which mixes the benefits of each hash-t and weighted apriori algorithms. to explore "huge information" on clouds, it's miles terribly essential to analysis data processing methods based mostly completely on cloud computing paradigm from each theoretical and sensible views finally the work compares the general performance with the present implementation of weighted and hash-t algorithms. The performance of our approach is manifested with the help of the initial experimental results documented during this paper.

KEYWORDS*:* Weighted apriori; Association rule mining; Mapreduce; Hadoop; Cloud; Data mining;

I. INTRODUCTION

The increasing ability to come up with large quantities information|of information|of information} brings potentials to get and utilize valuable knowledge from data. data processing has been an efficient tool to research information from totally different angles and obtaining helpful info from information. It may also facilitate in predicting trends or values, classification of knowledge, categorization of knowledge, and to search out correlations, patterns from the dataset. On the opposite hand, utilizing the large quantity information|of knowledge|of information} presents technical challenges as data storage and transfer approaches has to take care of prohibitory amounts of knowledge. The management of knowledge resources and dataflow between the storage and reckon resources is turning into the most bottleneck. Analyzing, visualizing, and spreading these massive information sets has become a significant challenge and information intensive computing is currently thought of because the "fourth paradigm" in scientific discovery when theoretical, experimental, and machine science [15]. To shift to the current fourth paradigm, researchers ought to set up strategically to perform information analysis. Cloud computing, a business model containing a pool of resources, provides an efficient paradigm for this purpose. during this context, cloud computing may be a distributed computing paradigm that permits massive datasets to be sliced and allotted to obtainable pc nodes wherever the information is processed domestically, avoiding network-transfer delays. This makes it attainable for folks to know and utilize the trillions of rows {of information|of information|of knowledge} in an exceedingly data center. we tend to powerfully believe that cloud computing is an efficient platform for data processing. to realize a lot of expertise in cloud-assisted data processing, during this paper, we tend to use association rule primarily based algorithmic program, Association rule mining aims to extract attention-grabbing correlations, patterns, associations among sets of things within the dealing info or different information repositories. The Apriori algorithmic program [5] is that the most generally used

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

algorithmic program for association rule mining. The input file size of Apriori is sometimes quite massive and distributed in nature. Therefore, a cloud may be a perfect platform for this algorithmic program. However, the classical Apriori algorithmic program wasn't designed to be performed within the parallel setting of cloud as a result of the unvarying approach to induce the frequent sets causes continual scans of the disk. This high I/O overhead makes running the algorithmic program in clouds impractical. though there are some paralleled association rule algorithms [16, 17, 18, 19], implementing these algorithms is troublesome as a result of programmers ought to take care of challenges of method communication and synchronization [20]. it's particularly troublesome to take care of failures that occur often in information intensive computing.

The MapReduce model has conjointly been employed in machine learning applications [21]. A key good thing about MapReduce is that it mechanically handles failures, concealment the complexness of fault-tolerance from the software engineer [11]. we decide to use Hadoop because the implementation, because it provides cheap and reliable storage, and tools for analyzing structured and unstructured information. However, changing a sequent algorithmic program to the parallel MapReduce format as needed by Hadoop desires a considerabl effort. There might arise things that the algorithms might not be effectively enforced within the mapreduce format, or implementing might need larger overheads and not catch up on the benefits Hadoop provides. In our work, we've got not solely revised Apriori to the MapReduce format, however conjointly improved its parallel performance in the slightest degree stages

II. LITERATURE SURVEY

This work is related to two main areas of datamining:
Association rule mining and Apriori, Mining association rule.

Association rule mining and Apriori:

Association rules are widely used in various areas, such as telecommunication networks, marketing and risk management, and inventory control. Many companies keep large quantities of their day to day transactions, which can be analyzed to learn the purchasing trend or pattern of customers. Such valuable insight can be used to support a variety of business-related applications, such as marketing, promotion of products, and inventory management. Besides market-based data analysis, association rules are also mined in the field of bioinformatics, medical diagnosis, web mining and scientific data analysis. The concept of strong rules was used to find association rules in items sold in supermarkets. An association rule defines a relation between two set of items. For example, *{A, B}=>{C}* in a purchase relation, would indicate if a person buys *A* and *B* together, he/she is more likely to also buy *C*.

Mining association rule:

Mining association rule consists of the following two steps: (1) finding the item sets that are frequent in the data set: The frequent item sets are set of those items whose support(*sup(item)*)in the data set is greater than the minimum required support(*min_sup*). Considering the above example all of the three items *A, B* and *C* belongs to the frequent item set and *sup(A, B)* and *sup(C)* would be greater than *min_sup*. The support of an item set is defined as proportion of transactions that contains the item set. (2) generating interesting rules from the frequent item sets on the basis of confidence (*conf*). The confidence of the above rule will be *sup(A, B)* divided by *sup(C)*. If the confidence of the rule is greater than the required confidence, then the rule can be considered as an interesting one. The performance of theassociation rule mining mostly depends on the first step i.e. generation of the frequent item set. As is evident, the algorithm does not require the details to be specified, such as like the number of dimensions for the tables or the number of categories for each dimension, as each item of transaction is considered. Therefore, this technique is particularly wellsuited for data and text mining of huge databases.

Many times, an association rule mining algorithm generates extremely large number of association rules. In most cases, it is impossible for users to comprehend and validate a large number of complex association rules. Therefore, it is important to generate only “interesting” rules, “non-redundant” rules, or rules satisfying certain criteria, such as coverage, leverage, lift, or strength. The Apriori algorithm is one of the most wildly used algorithms to generate association rules [2]. Apriori algorithm was designed to run on databases containing transactions. It is a ‘bottom up’ approach as candidate items are first generated and then the database is scanned to count the support for candidate items exceeding minimum support.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

The number of items in candidate subsets is increased one at a time with iteration. These candidate sets are converted to frequent subsets once their support count is matched with the required minimum support. The iteration would stop when no frequent subset can be generated. In the candidate generation phase, an additional pruning step is used, to check that all the subsets of the candidate are frequent. This helps in reducing the size of candidate set before scanning the data base. In this method, the number of times the input file will be read will depend on the number of iteration is required to get the maximum items in the frequent itemset. For example if the maximum number of items in the frequent itemset is 15, then the whole input file will be read a minimum of 15 times.
The Apriori Algorithm is based on the Apriori property:
 If an item set *I* does not satisfy the minimum support threshold (i.e., *P (I)* < *min_sup*), then *I* is not frequent.
 If an item *A* is added to the item set *I*, then the resulting item set (i.e., $I \vee A$) cannot occur more frequently than *I*, i.e., *P* ($I \vee A$) < *min_sup*.
The detailed Apriori Algorithm can be found in [2].

III. BACKGROUND

1) Association rule mining and Apriori: Association rules square measure wide utilized in varied areas, like telecommunication networks, selling and risk management, and internal control. several corporations keep massive quantities of their day to day transactions, which might be analyzed to find out the buying trend or pattern of shoppers. Such valuable insight will be accustomed support a spread of business-related applications, like selling, promotion of product, and inventory management. Besides market-based knowledge analysis, association rules also are strip-mined within the field of bioinformatics, diagnosis, net mining and scientific knowledge analysis. The thought of sturdy rules was employed by Agarwal et al [1] to seek out association rules in things sold-out in supermarkets. Associate in Nursing association rule defines a relation between 2 set of things. as an example, => in a very purchase relation, would indicate if an individual buys A and B along, he/she is additional doubtless to additionally purchase C. Mining association rule consists of the subsequent 2 steps: (1) finding the item sets that square measure frequent within the knowledge set: The frequent item sets square measure set of these things whose support(sup(item))in the info set is bigger than the minimum needed support(min_sup). Considering the higher than example all of the 3 things A, B and C belongs to the frequent item set and sup(A, B) and sup(C) would be larger than min_sup. The support of Associate in Nursing item set is outlined as proportion of transactions that contains the item set. (2) generating fascinating rules from the frequent item sets on the idea of confidence (conf). the arrogance of the higher than rule are going to be sup(A, B) divided by sup(C). If the arrogance of the rule is bigger than the specified confidence, then the rule will be thought of as a stimulating one. The performance of the association rule mining largely depends on the primary step i.e. generation of the frequent item set. As is obvious, the algorithmic rule doesn't need the small print to be specific, like just like the range of dimensions for the tables or the amount of classes for every dimension, as every item of dealings is taken into account. Therefore, this method is especially wellsuited for knowledge and text mining of giant databases. Many times, Associate in Nursing association rule mining algorithmic rule generates extraordinarily sizable amount of association rules. In most cases, it's not possible for users to grasp and validate an oversized range of advanced association rules. Therefore, it's vital to come up with solely “interesting” rules, “non-redundant” rules, or rules satisfying sure criteria, like coverage, leverage, lift, or strength. The Apriori algorithmic rule is one among the foremost wildly used algorithms to come up with association rules [2]. Apriori algorithmic rule was designed to run on databases containing transactions. it's a ‘bottom up’ approach as candidate things square measure initial generated and so the information is scanned to count the support for candidate things extraordinary minimum support. the amount of things in candidate subsets is redoubled one at a time with iteration. These candidate sets square measure born-again to frequent subsets once their support count is matched with the specified minimum support. The iteration would stop once no frequent set will be generated. within the candidate generation section, an extra pruning step is employed, to see that every one the subsets of the candidate square measure frequent. This helps in reducing the scale of candidate set before scanning the info base. during this methodology, the amount of times the input data are going to be browse can depend upon the amount of iteration is needed toget the utmost things within the frequent itemset. as an example if the utmost range of things within the frequent itemset is fifteen, then the total input data are going to be browse a minimum of fifteen times. The Apriori algorithmic rule relies on the Apriori property: • If Associate in Nursing item set I doesn't satisfy the minimum support threshold (i.e., P (I) < min_sup), then I isn't frequent. • If Associate in Nursing

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

item A is value-added to the item set I, then the ensuing item set (i.e., I ∨ A) cannot occur additional often than I, i.e., P (I ∨A) < min_sup.

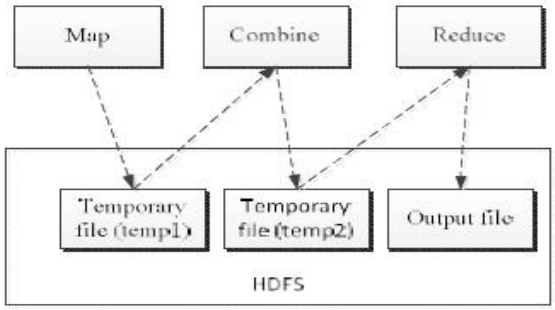

Fig.1 Flow diagram1

IV. BACKGROUND AND RELATED WORK

MapReduce associated Hadoop: MapReduce may be a programming paradigm and an associated parallel and distributed implementation for developing and capital punishment parallel algorithms to method huge datasets on clusters of goods machines [11]. Algorithms area unit per MapReduce victimization 2 functions: map and cut back. A MapReduce job typically splits the input data-set into freelance chunks that area unit processed by the map task in a very fully parallel manner. The map task maps the task into key and worth. The framework types the outputs of the map tasks, that area unit then inputted to the cut back tasks. The input of cut back and output of map should have same information sort. Typically each the input and also the output of employment area unit hold on in a very file-system. The MapReduce framework consists of one master Job hunter and one slave Task hunter per cluster-node. The master is chargeable for planning a job's element tasks on the slaves, watching them and re-executing the unsuccessful tasks. the quantity of times to re-execute a unsuccessful job is configurable. The slaves execute the tasks as directed by the master. The MapReduce framework operates completely on <key, value> pairs. that's to mention that the framework views the input of employment as a collection of <key, value> pairs and produces a collection of <key, value> pairs because the output of the task. The output of map and also the output of cut back don't got to be of a similar sort. The output of cut back may be used as input for one more (different) map perform to develop MapReduce algorithms that complete in multiple rounds. Hadoop may be a free, Java-based programming framework that supports the process of huge information sets in a very distributed computing atmosphere [22]. Hadoop is intended to run on an outsized variety of machines that donnt share any memory or disks. It creates clusters of machines and coordinates works. Hadoop consists of 2 key services: reliable information storage victimization the Hadoop Distributed classification system (HDFS) and highperformance parallel processing victimization Map/Reduce. HDFS splits user information across servers in a very cluster. It uses replication to avoid information loss even once multiple nodes fail. It conjointly keeps track of locations of replicated files. Moreover, Hadoop implements MapReduce, the aforesaid distributed multiprocessing system. Hadoop was designed for clusters of goods, shared-nothing hardware. No special programming techniques area unit needed to run analyses in parallel victimization Hadoop MapReduce as most existing algorithms will work while not several changes. MapReduce takes advantage of the distribution and replication of knowledge in HDFS to unfold execution of any job across several nodes in a very cluster. albeit some machines fail, Hadoop continues to work the cluster by shifting work to the remaining machines. It mechanically creates an extra copy of the info from one in every of the replicas it manages. As a result, clusters area unit selfhealing for each storage and computation while not requiring intervention by systems directors

V . METHODOLOGY

Overview

In this section, we tend to gift a way to design the Apriori algorithmic program in order that it are often effectively dead on MapReduce. we tend to additionally illustrate a way to implement the revised Apriori on Hadoop. First, the information|input file|computer file} set is saved on the HadoopDistributed classification system (HDFS) as a result of

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

HDFS will hold large chunks of information and HDFS will facilitate data localization for MapReduce task. The computer file is split into elements and assigned to the clerk. The output from the clerk is things (keys) and every item is count as worth. Thereafter, the output is taken in by the combiner. The combiner combines all the count values associated with a selected item (key). The result from the combiner is then taken in by the reducer for any combining associated rundown of the values equivalent to an item (key). once obtaining the total of all of the values of associate item, the reducer checks whether or not the total worth exceeds a given threshold value; if therefore, the total worth in conjunction with the item square measure written to the output. The total worth of associate item is discarded if it's but the minimum support threshold worth min_sup. This method would generate frequent-1 item set. Frequent-1 item set consists of set of 1 item pattern. The candidate-2 item set is generated exploitation the frequent - one item set within the clerk. The count of every pattern inside the candidate -2 item set is checked exploitation the computer file appointed to the clerk. Again, the output is given to the combiner, that accumulates and sums the of candidate item set with the corresponding worth. The "Reducer" any reduces and combines the info. within the finish solely those candidate sets whose worth exceeds the worth of minimum threshold square measure written to the computer file. an equivalent method is perennial. The candidate set is generated exploitation previous iteration's frequent item set. The iteration stops once no a lot of candidate set are often generated for associate iteration, or the frequent item set of the previous iteration can not be found. Once the frequent-n item set is generated and saved as output, the association rules square measure generated. The association rule confidence calculation needs the support count of item set that forms the rule. The item set in conjunction with its support count is hold on within the computer file, every line with a special frequent item set tab-separated with the support count. The frequent-n item set resides within the output folder's 'n' named sub-folder. Once confidence is found to be 100% the rule is generated.

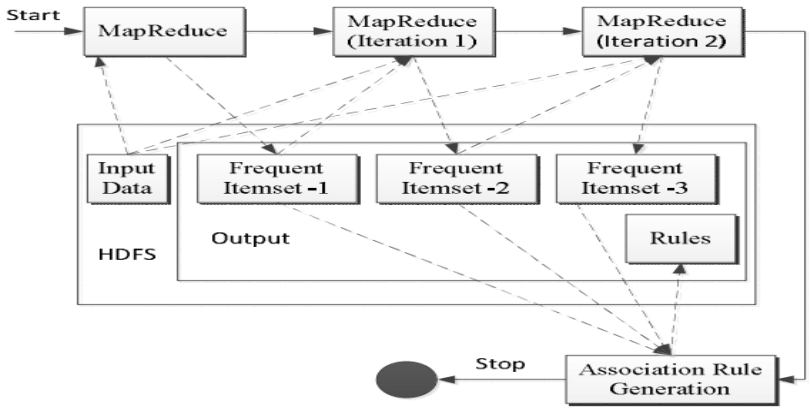

Figure.2. Data flow diagram 2 showing two iteration

VI. PROPOSED WORK

Big data applied to datasets whose size is beyond the ability of commonly used software tools to capture, manage and process the data within a tolerable elapsed time. Based on the previous work, it has been proposed to overcome the drawbacks of existing algorithms.

1. Weighted Apriori:

1. Itemset combination will be generated frequently. it will increase the candidate itemsets.
2. The weight computation for the each transaction will take more time to execute.
3. Less Accuracy
4. Not depent on data deviation

2. Hash Tree Apriori

1. High computational requirement

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

2. High memory utilization
3. Node processing requires high time to compute. The benefits of both the algorithms could also not be bypassed, so that a new algorithm is invented with the positives of both weighted apriori and hash t apriori algorithms. To overcome the negatives and to combine the positive approaches of both the algorithms of previous work [14] a new tree-Based hybrid approach is being introduced in the proposed work.

3. Benefits of the hybrid approach

1. It will reduce the computation time as well as the accuracy of the frequent item prediction.
2. It will avoid the candidate set generation.
3. It will generate the tree structure and analyze
   their height, weight and reach ability for each node. [Reach ability refers to reach the end point of the transaction of each item in the tree]
4. The work definitely gives best outcome for the mining of frequent item sets with the state-of-art technologies like hadoop hdfs, vmware and eucalyptus platforms.
5. Such a combination of modified algorithm in hadoop VMWare environment doesn't exist earlier with respect to the literature survey.

Before discussing the new algorithm few points ae to be explained. Frequent Item mining is the process of predicting the frequent patterns in the dataset. It will be applicable to real-time data such as Gene/Protein Sequence, Public usage, etc. Here prediction of combination and relationship between items is the most important to predict the association between the items. In the proposed system the selected datasets (Mushroom & Retail) contain numerous numbers of transactions. From these transactions first it is needed to predict the items in the transactional database. For predicting the frequent items and association rules in the transactional database it generally consumes lot of memory storage and time. To overcome the problem the new proposal computes the frequent pattern without the generation of candidate item sets. Tree based mining approach is used to reduce the consumption of the memory usage in the association rule prediction during reducing phase. It uses correlation based approaches to optimize the process.

4. Hybrid Apriori

The hybrid algorithm mines the transaction database which is considered to be the data for distribution. It finds the frequent item sets in the first iteration and also finds the relationship matrix. For each transaction it scans the database for counts and obtains the subset of transaction that is the candidate. For each candidate, count is increased. Then using the relationship matrix frequent tree is constructed. If the tree is having a single transaction path then for each combination (which is there in combination tree) of the nodes in the transaction path it generates pattern and that pattern with support count more than the minimum support count of nodes already in the combination tree, will be stored as a result tree. The same manner is followed throughout the construction of combination tree and result tree and the process is iterated.

The flow diagram of the proposed work is given in Fig.3. The HWFS (H stands for Height, W stands for Weight, F stands for Frequency and S stands for support) Apriori in the figure.1. refers to the concept of finding Height, Weight, Frequency & Support in same apriori tree. The flow diagram given above clearly depicts the map and reduce functions are working in a distributed environment with tree based mining approach to reduce the consumption of the memory usage in the association rule prediction.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

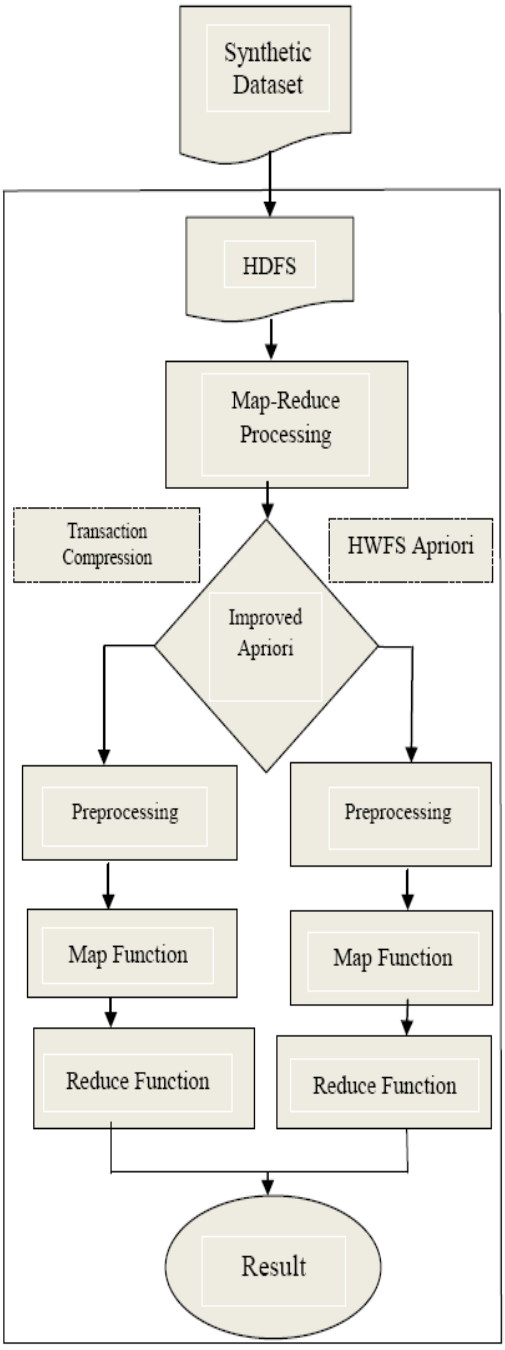

Fig.3 Flow diagram3

The diagram in fig.4. depicts the way of representing the logic of the tree formation. First the process starts with item prediction which finds the frequency of items, secondly relation prediction is done and a relationship matrix is formed. Then reachability prediction is being done to find the direct and indirect relationship using the matrix.

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

Support/confidence are computed which is the weighted support and confidence and can be used in association evaluation to find the association rules.

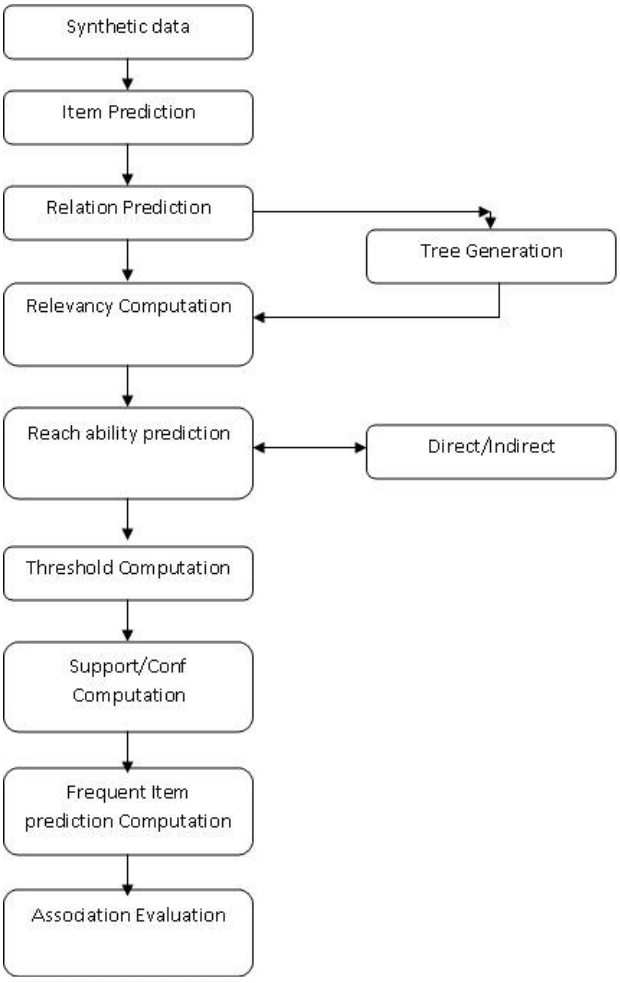

Fig.4 Flow diagram4

VII. OPTIMIZATION

The ways in which for rising the rule is assumed over. Any rule can provide higher performance if mathematical model is employed during a appropriate manner. during this projected rule it's identified that performance is improved by employing a correlation model within the tree construction section. data processing during a distributed atmosphere for certain should prove the information integrity before and when the distribution, whether or not any integrity and security violations are done to the information. during this rule this method has been done exploitation Hash Message Authentication Code [9] and is verified to be correct.

VIII. COMPUTING ENVIRONMENT

The new rule is enforced in java exploitation eclipse platform. Eucalyptus is getting used to form the cloud platform and eucalyptus nodes are created in VMWare machines. it'll facilitate to search out the results of the frequent item mining within the real cloud atmosphere. This improves the performance of the new mplementation. Another VMWare machine has been got wind of for hadoop distributed filing system (HDFS). Here hadoop are used because the

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 3, March 2016

execution node within the eucalyptus atmosphere to handle the mining method. 2 informationsets Retailitem and Mushroom are taken from UCI data repository. The hadoop atmosphere here has four nodes that area unit handling the datasets exploitation HDFS. secret writing has been dead and result's obtained.

IX. CONCLUSION

Cloud computing provides efficient solutions of storing and analyzing mass knowledge. during this paper, we've studied parallel association rule mining strategy within the cloud computing atmosphere. above all, we have a tendency to enforced the improved Apriori formula on MapReduce programming model on the Hadoop platform. The improved formula will proportion to giant knowledge set with relatively less price. Moreover, this distributed formula may cater to the distributed nature of the input file. what is more, details of the management of the distributed systems, like knowledge transferring among nodes and node failures area unit taken care by Hadoop, that adds an excellent deal of hardiness and quantifiability to the system. This work on implementing a replacement improved weighted hash T apriori formula, that is being termed as a hybrid formula, during a hadoop - map cut back atmosphere and obtains results employing a retail knowledge set and mushroom dataset. optimisation and knowledge integrity {are also|also area unit|are} taken care that are all necessary within the success of any formula particularly distributed cloud data processing. the most effective formulas in association rule mining has been chosen which supplies higher insight within the work however in future the work is increased by having completely different type of algorithm, with optimum tree pruning ways which is able to eliminate the not often used things during a higher approach. The analysis work could any proceed to focus on the importance for still higher tree construction. a search work couldn't be confined to fastened range of distributed nodes. that the same work is tested with eight and sixteen range of nodes. clearly the writing can work and provides higher results. Next section of modification within the formula is worn out the tree construction step to provide a additional optimized tree.

REFERENCES

[1] Agrawal, R. & Srikant. R. Fast algorithms for mining association rules. *Proceedings of the 20th VLDB Conference Santiago, Chile (VLDB '94)*, 1994,pp: 487-499.
[2] Aouad, L.M., et al., . Distributed frequent itemsets mining in heterogeneous platforms. *J. Eng. Comput. Archtecture*: 1(2), 2007.
[3] Bhagyashree Ambulkar. & Vaishali Borkar.,. 'Data Mining in Cloud Computing', *IJCA*, 2012
[4] Domenico Talia., 'Grid-based Distributed Data Mining Systems, Algorithms and Services', *SIAM Data Mining Conference*, 2006.
[5] Grudzinski, P. & Wojciechowski. M. , Integration of candidate hash trees in concurrent processing of frequent itemset queries using apriori. *Control Cybernetics*.:38(1), 2009.
[6] Kambatla, K., et al., . Towards optimizing hadoop provisioning in the cloud. Proceedings of the 2009 Conference on Hot Topics in Cloud Computing *(HotCloud'09)*: 2009,Article No. 22.
[7] Juan Li , Pallavi Roy , Samee U.et.al.. Data mining using clouds: An experimental implementation of apriori over mapreduce. *Proceeding of the 12th IEEE International Conference on Salable Computing and Communication (ScalCom 2102)*, 2012,Changzhou, China.
[8] Robert Grossman. & Yunhong Gu.,. 'Data Mining Using High Performance Data Clouds: Experimental Studies Using Sector and Sphere', *NSF*, 2004.
[9] Federal Information Processing Standards Publication, The Keyed-Hash Message Authentication Code (HMAC), *Information Technology Laboratory, National Institute of*
Standards and Technology, Gaithersburg, MD, 2001.
[10] Sun, K. & Bai.F. . Mining weighted association rules without preassigned weights. *IEEE T. Knowl.Data En., 20(4):* 2008, 489-495.
[11] Wang, K. & Thomas Su.M.Y.. Item selection by hub-authority profit ranking. *In SIGKDD, 2002*, pp: 652-657.
[12] Wang, W., Jeong yang & Philip.S., Efficient mining of weighted association rules. *KDD 2000, Boston, MA USA*, 2000, pp: 270-274.
[13] Zhou zhao, Da Yan & Ng, W., Mining probabilistically frequent sequential patterns in large uncertain databases., *IEEE transactions on knowledge and data engineering*, 2014, vol. 26, no. 5.
[14] R.Sumithra, Sujni Paul and D. Ponmary Pushpa Latha, Apriori Association Rule Algorithms using VMware Environment.
[15] T. Hey, S. Tansley, and K. Tolle. The Fourth Paradigm: Data-Intensive Scientic Discovery. Microsoft Research, 2009.
[16] G. Buehrer, S. Parthasarathy, S. Tatikonda, T. Kurc, and J. Saltz. Toward terabyte pattern mining: an architecture-conscious solution. In Proceedings of the 12th ACM SIGPLAN symposium on Principles and practice of parallel programming, PPoPP '07, pages 2–12, New York, NY, USA, 2007. ACM.
[17] M. El-Hajj and O. Zaiane. Parallel leap: large-scale maximal pattern mining in a distributed environment. In Parallel and Distributed Systems, 2006. ICPADS 2006. 12th International Conference on, volume 1, page 8 pp., 0-0 2006.
[18] W. Fang, K. K. Lau, M. Lu, X. Xiao, C. K. Lam, Y. Yang, B. He, Q. Luo, P. V. Sander, and K. Yang. Parallel data mining on graphics processors. Technical Report 07, The Hong Kong University of Science & Technology, 2008.