International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2015): 78.96 | Impact Factor (2015): 6.391

A Review on Scale up the Performance of Collaborative Filtering Algorithm

Barkha A. Wadhvani[1], Sameer A. Chauhan[2]

[1]Department of Computer Science, Government Engineering College, Modasa, India

[2]Department of Information Technology, Government Engineering College, Modasa, India

Abstract: ***Now a days everything goes digital from retail stores to Government offices. Also Digital India scheme is declared for digital empowerment where every information is digitally available. In this era of Information or Internet explosion data and choices are increases tremendously. More and more choices demands good recommendations so recommendation system came into the picture. It is very useful substitute for user to discover items or information that they might not have found by themselves easily. Collaborative filtering is the process of filtering for the information or for the patterns using techniques involving collaboration among viewpoints, multiple users or data sources etc. In Literature it is proven the most efficient technique to provide recommendations. This paper describes the techniques to scale up the performance of collaborative filtering algorithm to provide recommendation.***

Keywords: Recommendation system, Scale up, Collaborative filtering algorithm, Distributed computation, Compute intensive tasks, TOPSIS, Hadoop, Cloud computing, CUDA, BigData

1. Introduction

In this digital world, everything goes digital whether it is information, banking or shopping. People prefers to be the part of on-line trending and mostly they prefers to buy goods on-line so that they can compare the features and price of the goods and can do smart shopping to save their time and money. But day by day with the growth of digitization, choices are increased exponentially. As more and more choices confuses the users and waste their time for searching the items they need. So instead of comparing or searching items recommendation system will help users to find the items they might need or like according to users past behaviors, feedback or purchases. It is information filtering system and it filters from the millions of items and predicts the items that are best suitable, required and most likely purchased by the users.

Now a days people requires suggestions, recommendation or reviews from other users by spoken words, general reviews, news reports and so forth before considering or purchasing something. So instead of referring all this things recommendation system saves users' time by providing recommendations easily and efficiently and moreover according to user's choices. Recommendation system helps people to find through available articles, jokes, grocery, restaurants, hotels, movies, music, books etc.

Recommendation system is of three types - Content based Filtering, Collaborative filtering and association rule based. In content based filtering it provides recommendation through contents of users profile like feedback, behavior history and user's item profile. In collaborative filtering it recommends the items by implicit or explicit ratings. In association rule based, it recommend the items to users which is well associated (in any manner or more combine purchased e.g. Bread with Butter) with the items they purchased or interested in.

The rest of the paper is structured as follows. In Section II we introduced the Collaborative Filtering Process, its Types and challenges. In Section III we present the analysis and related work. Finally, Section IV concludes this paper.

2. Collaborative Filtering

Collaborative Filtering Algorithm is the most widely used and effective proven method for providing recommendations. It recommends items from the users past ratings collectively. It is the process of filtering for the information or for the patterns using techniques involving collaboration among viewpoints, multiple users or data sources etc. Collaborative Filtering Algorithm is a type of compute intensive application. It has more time complexity e.g. (xy) where 'x' is number of users and 'y' is number of items. As suppose if the similarity between users and items is to be predicted, and if items and users are increases accordingly then complexity also increases. There are millions of items and users are participated in the system so computation time is also increases. It might not provide the recommendations in tolerable time. So Recommendation System is the wonderful example of Big Data computations.

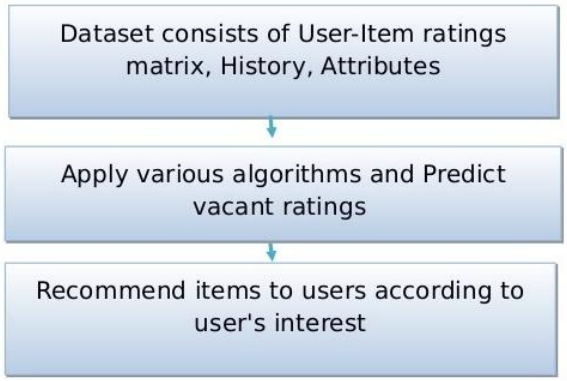

Figure 1: Collaborative Filtering Process

A. Major Challenges

- Sparsity: Sparsity is the lack of information or ratings. This problem occurs when users rated very few items from the millions of items. So at that time it will be very difficult to provide recommendations.
- Scalability: When numbers of existing users and items grow tremendously and exponentially, Recommendation system will suffer serious scalability problems, with computational resources going beyond acceptable levels in terms of time.
- Shilling Attacks: It happens when people give very positive ratings on their own products or items and give negative ratings on other opposition items. Now ultimately recommendations in collaborative filtering is depend on ratings so the item that have more positive ratings is recommended the most and which has less ratings is less recommended. So recommendations can be biased.
- Gray Sheep: It refers to the users whose opinions or rating does not match or very less match with other users so it will be very difficult to recommend them.
- Cold Start: Cold start problem is the new item or user problem. When new item or user included to the dataset then it has no rating at all or no sufficient rating, so collaborative filtering fails to provide recommendation.

B. Collaborative Filtering Approaches

Collaborative filtering algorithm has two phase approaches - Model based and Memory based. In Model based approach, it provides recommendations by developing the model based on the users ratings. It extracts some information from the dataset, and uses that as a "model" to make recommendations without having to use the complete dataset every time. It uses cluster based collaborative filtering and Bayesian based methods to make models. Cluster can be made based on users characteristics, Features, Locality, interests and many more. As items are divided into few groups or clusters according to some characteristics or features, it can solved scalability and sparsity problem [1]. To solve the sparsity problem mostly singular value decomposition (SVD) is used [2]. It is the most favorable technique for the matrix factorization. New approaches for SVD are proposed to effectively reduct the dimensions and it outperforms the traditional SVD [3]. The problem with the model based approach is loss of some important information sometime as it reduct the dimensions in creating model and it considers only few characteristics sometimes to solved the sparsity and scalability challenge.

Memory based approach has two main techniques - User based and Item based. It working on items-users matrix with explicit ratings given by users. In user based approach, similarity between the users are calculated based on their ratings and from that most similar users item preferences or ratings, the unrated items are predicted for the particular user and recommend the top-K items to that user. In Item based, similarities between the items are calculated and predict the unknown ratings for the users based on the similarity of the items. To compute the similarities between the user-user or item-item, many methods are used like Cousine correlation coefficient, Pearson Correlation coefficient, Jaccard coefficient, Euclidean, Ratio based and so forth. Among them Pearson correlation coefficient is the most efficient and benchmark to find the correlations [4]. To select user based or item based approach is totally depends on the recommendations the system wants to provide. Item based approach is most efficient approach than user based as user's behavior or choices may change on time basis but the items do not change frequently. So item-item correlations can be found offline also. And also users in size are mostly more than the item count or item size. Moreover good recommendations should be scalable over large dataset and generate recommendations in some sub seconds or as quick as it can. Item-item correlation can overcome scalability and sparsity challenge [5]. In [6] authors have introduced Ratio based approach. In ratio based collaborative filtering approach, it calculates the ratio between the ratings of items which are rated commonly by users. Ratio based is compared with the traditional item based approach with Pearson correlation coefficient and conclude that Ratio based approach is less expensive than other memory based algorithms and also more efficient and scalable for item based collaborative filtering algorithm. If recommendation system wants to provide recommendations according to some similarities between users or user's characteristics, behavior or features then user based approach is used. In [7, 8], authors proposed a novel approach for providing quality recommendations by considering user location. According to [9, 10] if social trust taken into the consideration by taking only users which are socially connected in some way then it increase the prediction accuracy of recommendations.

3. Related Work

Recently data are increases right away and it is beyond the capacity of storage and processing so called Big Data [11]. It is very difficult to process that data using traditional software's or databases. [39, 40] discussed the big data issues and technologies that is necessary to be considered to handle big data. If we are talking about ecommerce or any others online business, users and items both are just increases momentarily so we can consider Recommendation system as a notable example of Big data. Now as the users and items are increases it will become burdensome to provide recommendations in bearable time. As in collaborative filtering algorithm, it needs to calculate the correlation between the user-user(u) or item-item(i). Now if there are millions of users or items then it requires approximately (u*u) or (i*i) computations only to calculate correlation and after that other computations are required to predict the unknown ratings and to provide recommendations. So it becomes the compute intensive task and it go through scalability problem in terms of time, storage and computation. It is expected to scale up the performance of collaborative filtering algorithm for recommendation system to achieve most desirable performance.

In [12], compared various methods including user based, item based - Pearson Correlation Coefficient (PCC) and Log Likelihood Ratio Similarity (LLR). They concluded that the increased in the contents can provide more accurate results. In [13], authors build user based generic restaurant recommendations using apache mahout platform and it uses the some threshold to generate to scale up the performance in terms of computations, time and quality.

Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) is a Multi-Criteria Decision Analysis Method by considering alternatives that have shortest geometric distance from the positive ideal solution and the longest geometric distance from the negative ideal solution [41]. Authors [42] introduced TOPSIS to find out a set of initial items and then recommend items users might like in the set of initial items according to similarity. It considering the multiple implicit attributes to provide more accurate and reasonable recommendations. Authors [43-45] combined various approaches with TOPSIS to provide ideal recommendations according to multiple attributes. TOPSIS has less dependency unlike other recommendation algorithms so it is more scalable. TOPSIS can also solves the cold start user and cold start item problem, Sparsity, Grey sheep and Shilling Attack as it can work well on multi-attribute and each attribute have different weights.

In order to scale up the performance of collaborative filtering algorithm, distribute the data to other systems or nodes or we can say parallelly execute that data on multiple nodes can provide result in time efficient manner. Authors [14] implemented item based approach by distributing the expensive computations on different systems using map reduce phases. It uses the cloud computing platform and evaluated the results by taking speedup metric. Authors concluded that parallel execution can excessively increases the scalability. In [15], authors implemented user based collaborative filtering approach on cloud computing platform by map reduce processing model to overcome the scalability challenge. They divide the calculation by user Id but the problem in Map Reduce processing framework is that it needs to initialize the mapper so for the number of users it required that number of initialization so it becomes resource and time consuming so it needed the process which takes the sizable computations but the less input data that required the frequent mapping initialization. Mostly all used the same dataset that is Movielense which is the touchstone dataset to provide recommendations and for evaluation purpose. A new method Threshold based Similarity Transitivity (TST) is proposed by [16]. TST is the threshold based and it primary filters out the fallacious similarity by setting a threshold. It implemented and designed to be scalable with Map Reduce framework and on the cloud computing platform. A neoteric approach to provide medicine recommendation which recommend the medicine according to symptoms [17]. It uses model based approach by creating clusters according to functional description of drugs and based on user collaborative filtering it provides recommendations. For that authors have used cluster based algorithm on cloud computing platform to overthrow the shortcomings such as expensive computations, cold start and sparsity.

Authors [18] proposed ClubCF approach based cloud computing for big data applications. It implemented to provide the service recommendations. Primarily services are merged into the clusters to make it scalable as the cluster has the much lesser services than the whole system. It costs less online computation time and provide more accurate results as in each cluster services are more related with each other. They concluded that ClubCF increases the coverage and solved the sparsity challenge in a certain degree.

A Distributed Stochastic Gradient Descent (DSGD) algorithm matrix factorization which is provides the scalability and works good on sparse data. DSGD implemented in distributed environment Hadoop with the slight modification to cluster setup and parallel paradigm and evaluated the good performance in terms of scalability and accuracy [19]. Recommendation system is supremely computational so it is optimal for Hadoop. A book recommendation engine is conferred by [20]. It uses the apache mahout with map reduce and Hadoop to overcome the data run-over. Hybrid approach that is the combination of content based and collaborative based is proposed in [21]. In this paper authors determined the self-organizing mapping (SOM) to optimize the improved k-means (IK) clustering in collaborative filtering. It is worked to recommend TV programs. Finally it concluded that this approach reduced the Mean Square Error(MAE) and overcome scalability by using clustering and distributed platform. [22] paper has also adopted the Map Reduce Hadoop platform for scaling up the performance of very large production level system by creating clusters.

Item based approach has better scalability than the user based approach [23]. In this authors implemented and compared both the approach and result showed that the execution time improves by about 30% with every supplement of a node into the Hadoop cluster. Micro-video Recommendation system proposed in [24]. To meet the data scale and performance of computing and for other aspects authors used slope one algorithm on Hadoop platform. To provide more accurate, scalable and trustworthy recommendations authors of [25] presented parallel approach that is based on social relationship. They finalized that their approach improves the scalability, coverage and cold start challenges. To unite the ability of some common PC to process the large scale data in short time pipeline approach with item based algorithm can be used by creating Hadoop cluster[26]. Pipeline approach utilizes all the computing resources and make all of them busy. Web service recommendation system is presented in [27] which is implemented on Hadoop platform to improve scalability in big data environment. It considered the positive and negative preferences of users to achieve more efficiency and accurate recommendations.

There are other parallel platforms also which can be used to scale up the performance of collaborative filtering algorithm. One of the parallel computing platform is CUDA (Compute Unified Device Architecture) provides a straightforward means of describing inherently parallel computations [28]. CUDA enables the drastic increases in computing performance by harnessing the power of GPU. This paper presents the CUDA parallel platform in detail. [29, 30] both have implemented recommendation system using CUDA parallel platform to scale up the performance. As the memory available on the GPU card is small authors used compression techniques to reduce the data volume so that the less number of iterations and computations required for similarity calculations. They concluded that parallel GPU implementations outperforms and also scale well with the increase with users or items and provides result in timely manner.

To scale up the performance not only the parallel or distributed platform required but the processing techniques, scheduling methods in cluster, scaling methods, replication, locality of data, database etc should also taken into the consideration. In [31] authors investigated the scalability capacities over storing, managing and processing and performance of different data management approaches. They identified Postgres-XL, parallel RDBMS, Hadoop, MapReduce, Spark, a next generation Map Reduce with in-memory processing framework, Cassandra, a peer-to-peer NoSQL data store and its details so any of them can be selected according to processing, data and type of system. They concluded that we should focus on minimize the data transfer as it can introduce some kind of delay in processing. MATRIX task execution framework is overwhelming the scaling limitations of Hadoop towards extreme scales [32]. Authors compared YARN and MATRIX by taking different examples and workloads for all scales. There are two types of scaling approach we can adopt in cluster- scale-up and scale-out. Scale-up approach is working well with the small or medium sized data (KB or MB) and Scale-out approach is perform better for large sized (GB, TB) data [33]. Scale-up is adding more capacity, power or resources to existing system itself and Scale-out is adding more machines with the existing machines to upgrade the performance. Scaling approach should be selected according to type of workload. For example scale-up approach is more efficient for write intensive workloads such as scan and sort. Scale-out is good for iterative algorithms as Join, K-mean, Page rank etc because they can effectively used CPU and memory by caching intermediate files [33]. As nowadays solely using either scale-up or scale-out cluster to run a workload cannot achieve high performance so authors [34] explored building the hybrid scale-up/out Hadoop architecture that is outperforms the traditional Hadoop architecture. To scale up the performance data locality also taken into the consideration. [35] proposed a new join algorithm CHMJ which is distributes the data according to hash values of join property which increases the data locality and availability. Authors concluded that CHMJ is five times efficient than Hive(Data Warehouse). Data locality is very important for scaling up the performance. Consider a scenario of application where users or items size is small or medium and in that case if data going to be distributed in cluster then it also requires the large amount of intermediate results over the network, which is inefficient so as the data locality of jobs is increases performance is also increases. To provide personalized recommendations it may require joining multiple tables so at that time proper Join algorithm should be selected and also considering data locality factor to scale the performance up. In [36], authors presented a parallel replica placement in HDFS to improve throughput. As stated by [37], as the load increases with the number of users, HBase-HDFS performs better than MySQL-HDFS. HDFS is write once and read multiple times storage system. HBase provides the random reads and random writes from HDFS storage. So according to application database model should be selected. Performance of the application can also be improved by considering efficient task scheduling algorithm for MapReduce process. [38] proposed the novel scheduler dynamic priority multi-queue scheduler (DPMQS) for MapReduce to increase the data locality and priority of jobs that are near to completing their Map phase so that it can bridge the time gap between the start of reduce jobs and execution of reduce tasks. MapReduce processing is also considered to be important to improve the performance in terms of time.

4. Conclusion

With the increase in digitization, data are growing extravagantly. As the users and items are also increases exponentially in any system, it will be very onerous to provide recommendations to the users in timely manner. Collaborative filtering is most extensively used technique for providing recommendation. There are various methods and algorithms are available to resolve the challenges of Collaborative filtering approach and to scale up its performance.

Approach like TOPSIS can be combined with the collaborative filtering algorithm to provide more accurate results by considering multiple attributes. Including Social network connections to collaborative filtering, more trustworthy recommendations and coverage can be achieved. And many challenges like Gray sheep, Cold-start; Shilling attacks etc can be solved. Moreover other environments like hadoop, cuda, cloud etc can be used to scale up the performance of collaboration algorithm.

References

[1] Xingyuan Li, "Collaborative filtering recommendation algorithm based on cluster, " Proceedings of 2011 International Conference on Computer Science and Network Technology, Harbin, 2011, pp. 2682-2685
[2] M. Sunitha Reddy and T. Adilakshmi, "Music recommendation system based on matrix factorization technique -SVD," 2014 International Conference on Computer Communication and Informatics, Coimbatore, 2014, pp. 1-6.
[3] Z. Sharifi, M. Rezghi and M. Nasiri, "New algorithm for recommender systems based on singular value decomposition method," ICCKE 2013, Mashhad, 2013, pp. 86-91.
[4] B. M. Sarwar, G. Karypis, J. A. Konstan, and J. Riedl. Item-based collaborative filtering recommender algorithms. In WWW10, 2001.
[5] G. Linden, B. Smith, and J. York, "Amazon.com recommendations: item- to-item collaborative filtering, " IEEE Internet Computing, vol. 7, no. 1, pp. 76–80, 2003
[6] Yaqiu Liu, Zhendi Wang and Man Li, "Ratio-based collaborative filtering algorithms, " 2008 2nd International Symposium on Systems and Control in Aerospace and Astronautics, Shenzhen, 2008, pp. 1-5.
[7] Z. Gao, Z. Lu, N. Deng and K. Niu, "A novel collaborative filtering recommendation algorithm based on user location, " 2016 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Nantou, 2016, pp. 1-2.
[8] J. Bu, X. Shen, B. Xu, C. Chen, X. He and D. Cai, "Improving Collaborative Recommendation via User-Item Subgroups, " in IEEE Transactions on Knowledge and Data Engineering, vol. 28, no. 9, pp. 2363-2375, Sept. 1 2016.

[9] W. Wenjuan and L. Zhubing, "A personalized recommendation strategy based on trusted social community, " 2015 10th International Conference on Computer Science & Education (ICCSE), Cambridge, 2015, pp. 496-499.
[10] P. Yu, "Recommendation method for mobile network based on user characteristics and user trust relationship, " 2016 IEEE International Conference on Big Data Analysis (ICBDA), Hangzhou, 2016, pp. 1-6.
[11] S. Kaisler, F. Armour, J. A. Espinosa and W. Money, "Big Data: Issues and Challenges Moving Forward, " 2013 46th Hawaii International Conference on System Sciences, Wailea, Maui, HI, 2013, pp. 995-1004.
[12] G. R. Bamnote and S. S. Agrawal, "Evaluating and Implementing Collaborative Filtering Systems Using Apache Mahout, " 2015 International Conference on Computing Communication Control and Automation, Pune, 2015, pp. 858-862.
[13] U. Farooque, "Implementing user based collaborative filtering to build a generic product recommender using Apache mahout, " 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom), New Delhi, 2016, pp. 984-987.
[14] J. Jiang, J. Lu, G. Zhang and G. Long, "Scaling-Up Item-Based Collaborative Filtering Recommendation Algorithm Based on Hadoop, " 2011 IEEE World Congress on Services, Washington, DC, 2011, pp. 490-497.
[15] Z. D. Zhao and M. s. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," 2010 Third International Conference on Knowledge Discovery and Data Mining, Phuket, 2010, pp. 478-481.
[16] F. Xie, Z. Chen, H. Xu, X. Feng and Q. Hou, "TST: Threshold based similarity transitivity method in collaborative filtering with cloud computing, " in Tsinghua Science and Technology, vol. 18, no. 3, pp. 318-327, June 2013.
[17] Y. Zhang, L. Wang, L. Hu, X. Wang and M. Chen, "COMER: Cloud-based medicine recommendation, " 10th International Conference on Heterogeneous Networking for Quality, Reliability, Security and Robustness, Rhodes, 2014, pp. 24-30.
[18] R. Hu, W. Dou and J. Liu, "ClubCF: A Clustering-Based Collaborative Filtering Approach for Big Data Application, " in IEEE Transactions on Emerging Topics in Computing, vol. 2, no. 3, pp. 302-313, Sept. 2014.
[19] M. Pozo and R. Chiky, "An implementation of a Distributed Stochastic Gradient Descent for Recommender Systems based on Map-Reduce, " 2015 International Workshop on Computational Intelligence for Multimedia Understanding (IWCIM), Prague, 2015, pp. 1-5.
[20] U. Sahu, A. K. Tripathy, A. Chitnis, K. A. Corda and S. Rodrigues, "Personalized recommendation engine using HADOOP, " 2015 International Conference on Technologies for Sustainable Development (ICTSD), Mumbai, 2015, pp. 1-6.
[21] Z. Ma, Y. Yang, F. Wang, C. Li and L. Li, "The 21 Based Improved K-Means Clustering Collaborative Filtering Algorithm in TV Recommendation System, " 2014 Second International Conference on Advanced Cloud and Big Data, Huangshan, 2014, pp. 288-295.
[22] J. Katukuri, T. Könik, R. Mukherjee and S. Kolay, "Recommending similar items in large-scale online marketplaces, " 2014 IEEE International Conference on Big Data (Big Data), Washington, DC, 2014, pp. 868-876.
[23] P. Ghuli, A. Ghosh and R. Shettar, "A collaborative filtering recommendation engine in a distributed environment," 2014 International Conference on Contemporary Computing and Informatics (IC3I), Mysore, 2014, pp. 568-574.
[24] S. Shang, M. Shi, W. Shang and Z. Hong, "A micro-video recommendation system based on big data, " 2016 IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS), Okayama, 2016, pp. 1-5.
[25] D. Lalwani, D. V. L. N. Somayajulu and P. R. Krishna, "A community driven social recommendation system, " 2015 IEEE International Conference on Big Data (Big Data), Santa Clara, CA, 2015, pp. 821-826.
[26] Z. L. Zhao, C. D. Wang, Y. Y. Wan, Z. W. Huang and J. H. Lai, "Pipeline Item-Based Collaborative Filtering Based on MapReduce, " 2015 IEEE Fifth International Conference on Big Data and Cloud Computing, Dalian, 2015, pp. 9-14.
[27] R. Subhashini and G. Akila, "Valence arousal similarity based recommendation services," 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, 2015, pp. 1-4.
[28] M. Garland et al., "Parallel Computing Experiences with CUDA, " in IEEE Micro, vol. 28, no. 4, pp. 13-27, July-Aug. 2008.
[29] C. H. Nadungodage, Y. Xia, J. J. Lee, M. Lee and C. S. Park, "GPU accelerated item-based collaborative filtering for big-data applications, " 2013 IEEE International Conference on Big Data, Silicon Valley, CA, 2013, pp. 175-180.
[30] Richa and P. Bedi, "Parallel context aware recommender system using GPU and JCuda," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, 2016, pp. 1383-1389.
[31] H. Chihoub and C. Collet, "A Scalability Comparison Study of Data Management Approaches for Smart Metering Systems," 2016 45th International Conference on Parallel Processing (ICPP), Philadelphia, PA, 2016, pp. 474-483.
[32] K. Wang et al., "Overcoming Hadoop Scaling Limitations through Distributed Task Execution," 2015 IEEE International Conference on Cluster Computing, Chicago, IL, 2015, pp. 236-245.
[33] Taekyung Yoo, Minsub Yim, Ilgyun Jeong, Yunsu Lee and Seung-Tae Chun, "Performance evaluation of in-memory computing on scale-up and scale-out cluster, " 2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN), Vienna, 2016, pp. 456-461.
[34] Z. Li and H. Shen, "Designing a Hybrid Scale-Up/Out Hadoop Architecture Based on Performance Measurements for High Application Performance," 2015 44th International Conference on Parallel Processing, Beijing, 2015, pp. 21-30.

[35] Y. Zhao et al., "A data locality optimization algorithm for large-scale data processing in Hadoop," 2012 IEEE Symposium on Computers and Communications (ISCC), Cappadocia, 2012, pp. 000655-000661.
[36] Patel Neha M., Patel Narendra M, M. I. Hasan, Shah Parth D and Patel Mayur M, "Improving HDFS write performance using efficient replica placement, " 2014 5th International Conference - Confluence The Next Generation Information Technology Summit (Confluence), Noida, 2014, pp. 36-39.
[37] Mehul Nalin Vora, "Hadoop-HBase for large-scale data, " Proceedings of 2011 International Conference on Computer Science and Network Technology, Harbin, 2011, pp. 601-605.
[38] X. Dai and B. Bensaou, "Scheduling for response time in Hadoop MapReduce," 2016 IEEE International Conference on Communications (ICC), Kuala Lumpur, 2016, pp. 1-6.
[39] H. Hu, Y. Wen, T. S. Chua and X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial, " in IEEE Access, vol. 2, no., pp. 652-687, 2014.
[40] M. C. Okur and M. Büyükkeçeci, "Big data challenges in information engineering curriculum, " 2014 25th EAEEIE Annual Conference (EAEEIE), Cesme, 2014, pp. 1-4.
[41] Hwang, C.L.; Lai, Y.J.; Liu, T.Y. (1993). "A new approach for multiple objective decision making". Computers and Operational Research. 20: 889–899.
[42] X. Wang and C. Zhou, "A collaborative filtering recommendation algorithm using user implicit demographic information, " 2012 7th International Conference on Computer Science & Education (ICCSE), Melbourne, VIC, 2012, pp. 935-939.
[43] Larasati, A. H. Setyaningrum and L. K. Wardhani, "Development Decision Support System of Choosing Medicine Using TOPSIS Method (Case Study: RSIA Tiara), " 2016 6th International Conference on Information and Communication Technology for The Muslim World (ICT4M), Jakarta, Indonesia, 2016, pp. 160-165.
[44] Hanane, O. Brahim and F. Bouchra, "CCSD and TOPSIS methodology for selecting supplier in a paper company, " 2016 4th IEEE International Colloquium on Information Science and Technology (CiSt), Tangier, 2016, pp. 275-280.
[45] X. Lu, W. Chang and S. Zhou, "Quality evaluation of diesel engine using an improved TOPSIS method based on K-means algorithm," 2016 35th Chinese Control Conference (CCC), Chengdu, 2016, pp. 9628-9633.