A Survey on Data Placement and Workload Scheduling Algorithms in Heterogeneous Network for Hadoop

Ruchi Mittal
Department of CSE
Rayat Bahra Institute of Engg. & Tech.
Patiala, Punjab

Harpreet Kaur
Department of CSE
Rayat Bahra Institute of Engg. & Tech.
Patiala, Punjab

ABSTRACT
The elastic scalability and fault tolerance of the cloud computing has led to a wide range of real world applications. However, processing requirements of Big Data in these applications pose a humongous challenge for achieving desired performance levels. MapReduce is an effective parallel distributed programming model for handling large unstructured datasets in cloud applications. Hadoop, an open source implementation of the MapReduce model, is currently being employed for high performance processing of Big Data. The current Hadoop implementation considers the nodes of a cluster in a homogeneous environment where each node has the same computing capacity and workload. But in real world applications the nodes may have different computing capacities and workloads resulting in a heterogeneous environment. In such heterogeneous environment the default Hadoop implementation does not yield the expected performance. This paper includes a survey on the algorithms proposed by different authors on (a) data placement strategies and (b) workload scheduling for Hadoop in heterogeneous network.

Keywords
Cloud Computing, Big Data, MapReduce, Hadoop, Heterogeneous Network.

1. INTRODUCTION
Nowadays there has been an exponential growth in the data produced every minute. To store, process and analyze such large volume of data, called Big Data, is proven to be a big challenge these days. Parallel computing is a method adopted for the purpose. MapReduce, a programming model is proven to be an efficient parallel data processing paradigm for large scale clusters. Hadoop, an open source implementation of the MapReduce model, can be used developed to process thousands of megabytes of the data on the Linux platform. Hadoop is also adopted by Amazon and Facebook for the processing and analysis of large volumes of data.

The MapReduce model partitions a program into multiple smaller tasks which are executed individually and in parallel. The results of them are combined by the model to give a single output. As compared to earlier parallel programming models, MapReduce model provides several benefits. First, in this model, new nodes can be added easily in the cluster without much modifications and the system works the same. Second, the system doesn't get much affected on the failure of a single node as the task is automatically allocated to some other idle node in the model.

Hadoop is an open source implementation of the MapReduce model for parallel processing of the large datasets. Heterogeneity and Data locality are the main factors affecting the performance of Hadoop system in the Hadoop architecture. In the classic homogeneous Hadoop system, all the nodes have the same processing ability and hard disk capacity. All the nodes are assigned same workloads and data required for their processing is often local requiring lesser data movement between the nodes. The data needed to be written in HDFS are partitioned in to several smaller data blocks of similar size and are assigned to the nodes equally. In such homogeneous environment load balancing can be maintained easily.

However in real world applications the nodes may be of different processing ability and hard disk capacity. With the default Hadoop strategy, the faster nodes may finish their tasks with their local data at a great speed. After finishing the task with local data faster nodes can work with nonlocal data which may be present on slower nodes which requires more data movement between the nodes, thus affecting the Hadoop performance.

The rest of the paper is organized as follows. Section II presents an overview on the basic terms of the MapReduce programming model, Hadoop and Hadoop Distributed File System (HDFS). Section III comprises the work presenting and describing the different algorithms on the workload scheduling and data placement strategies. Section IV and Section V explains the conclusion drawn from the survey and opportunities for future research work.

2. BACKGROUND
2.1 MapReduce Model
MapReduce is a parallel programming model used for large sets of data processing in a distributed environment with a large number of computing nodes. This model is proposed by Google in 2004. An application which is to be executed in MapReduce model is called a job and is divided into 'map tasks' and 'reduce tasks'. This model works on the strategy of 'divide and conquer'. The large set of data which is stored for processing is divided into the blocks of same size. These blocks are then allocated to the nodes which process them in parallel. The result of the blocks with same map function is composed of <key, value> pairs. The reduce nodes combine these intermediate outputs and generate the final output data.

2.2 Hadoop
Hadoop is an open source implementation of the MapReduce model supported by Apache Software Foundation. Hadoop MapReduce and Hadoop Distributed File System (HDFS) are two main parts of the Hadoop System. MapReduce handles the tasks of parallel computing and HDFS manages the data management. The jobs are classified into tasks and are executed in parallel by the MapReduce. The data are classified into blocks and are handled by the HDFS. These tasks and data blocks are assigned to the nodes of the cluster. Hadoop adopts the master/slave architecture in which the master is the JobTracker and slave is the TaskTracker.

JobTracker is responsible for the job scheduling and task distribution and TaskTracker is responsible for performing the tasks and to return the result to the JobTracker. These use heartbeat messages for communication. High-end PC's are not necessary for high performance computing while working with Hadoop. Several general PC's can be used for the purpose with Hadoop with which high performance platform can be build saving large amounts of money.

2.3 Hadoop Distributed File System (HDFS)

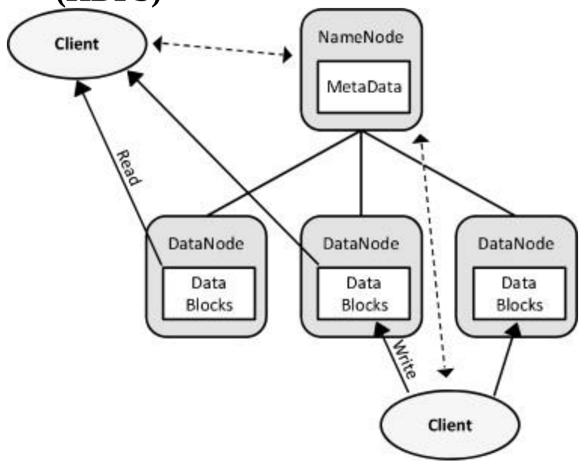

Fig 1: An overview of HDFS Read and Write [5]

Based on the Google File System, the HDFS is implemented by Yahoo! which is used with the MapReduce model. In HDFS (figure 1), the master is the NameNode and slave is the DataNode. The complete file system and file information is stored and managed by the NameNode. The files written in the HDFS are also partitioned into same sized blocks and assigned to the DataNode by the NameNode. The data blocks are stored by the DataNode.

3. RELATED WORK

The structured survey starts from the definition of the terms-workload scheduling and data placement strategies for Hadoop which will guide the work description further.

3.1 Workload Scheduling

There are three main schedulers which come along with Hadoop. These are FIFO, Fair Scheduling and Capacity Scheduling. In FIFO, all the jobs are loaded into a queue and are scheduled to execute accordingly as these are loaded in the queue. In Fair Scheduling, a minimum number of map and reduce slots are allocated to the jobs, i.e. each job receives a fair share of cluster's resources. In Capacity Scheduling, each queue shares the cluster's computational resources according to its priority. These schedulers were designed to work in the homogeneous network. But in real world applications, clusters are of different computing capabilities and storage capacity. So the expected output cannot be obtained from these in heterogeneous networks. In this, we present the new scheduling algorithms proposed by different authors to work in heterogeneous networks to enhance the Hadoop's performance.

3.2 Data Placement

The data placement strategy of the MapReduce for data placement consists of two functions: Map and Reduce. Jobs are divided into map and reduce tasks to be executed by Mapper and Reducer. First, the input is loaded into the HDFS and data are partitioned into equal sized data blocks and each block is handled by the mapper for data processing. As an intermediate output, a <key, value> pair is generated which is given to the reducer which merge them generating a single output (figure 2).

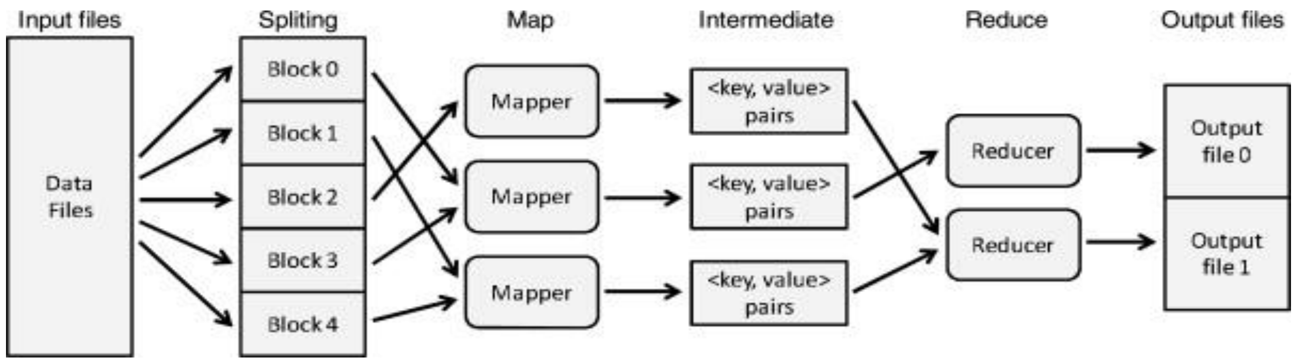

Fig 2: An overview of MapReduce Model [5]

Table 1. Proposed Workload Scheduling Algorithms along with their performance.

Author's Name	Publication	Title	Proposed Work	Results
Julio C.S. Anjos, Ivan Carrera, Wagner Kolberg, Andre Luis Tibols, Luciana B. Arantes, Claudio R. Geyer	Future Generation Computer Systems, Jan. 2015.	MRA++: Scheduling and data placement on MapReduce for heterogeneous environments	The proposed MRA++ design considers the heterogeneity of the nodes during the distribution of data, scheduling of tasks and in job control. A training task is set for information gathering before the distribution of data.	The new algorithm attains 70% of more performance gain in 10 Mbps networks by nullifying the introduced delay in setup phase.
Zhou Tang, Min Liu, Almoalmi Ammar, Kenli Li, Keqin Li	The Journal of Supercomputing, Nov. 2014.	An Optimized MapReduce Workload Scheduling Algorithm for Heterogeneous Computing	This paper considers the large data processing workflow as DAG which consists of MapReduce jobs. The proposed algorithm calculates the priorities of the jobs by categorizing them into I/O intensive and computing intensive and the slots are allocated accordingly. Then the workflow is scheduled for the tasks according to the data locality.	The experimental results show that the schedule length and the parallel speedup for the workflow task can be improved with the proposed algorithm.
Feng Yan, Ludmila Cherkasova, Zhuoyao Zhang, Evgenia Smirni	7th IEEE International Conference on Cloud Computing, July 2014.	Optimizing Power and Performance Trade-offs of MapReduce Job Processing with Heterogeneous Multi-Core Processors	Based on different job types, as large jobs requiring faster output or small interactive jobs requiring faster response time, this scheduler called DyScale, is designed to allocate jobs accordingly to fast or slow cores in a heterogeneous cluster by creating virtual resource pools by using priority scheduling.	With this scheduler, smaller jobs can be executed up to 40% faster and the output of large jobs can be achieved 40% higher in the simulation study.
Jessica Hartog, Renan DelValle, Madhusudhan Govindaraju, Maichael J. Lewis	IEEE International Congress on Big Data, July 2014	Configuring A MapReduce Framework For Performance Heterogeneous Clusters	In this, the proposed MapReduce framework called MARLA, divides a task into subtasks and delays the binding of data to the subtask's process.	The proposed work can improve the performance for few upgraded nodes but didn't affect it equally.
Xiaolong Xu, Lingling Cao, Xinheng Wang	IEEE Systems Journal, 2014.	Adaptive Task Scheduling Strategy based on Dynamic Workload Adjustment (ATSDWA) for Heterogeneous Hadoop Clusters.	Using ATSDWA, the TaskTrackers can adjust themselves according to the load change at runtime and according to their computing capacity can obtain the tasks while realizing the self-regulation.	The proposed algorithm can be beneficial for both the JobTracker by avoiding its overloading and for the TaskTracker by reducing the task execution time giving more stability to the heterogeneous Hadoop cluster. It can be applied to both heterogeneous and homogeneous environments.

Table 1 (Continued)

Author's Name	Publication	Title	Proposed Work	Results
Aysan Rasooli, Douglas G. Down	Future Generation Computer Systems, 2014.	COSHH: A Classification And Optimization Based Scheduler for Heterogeneous Hadoop Systems	A new Hadoop scheduling algorithm is designed and implemented considering the heterogeneity both at the application and cluster level. The main goal of the proposed algorithm is to improve the average completion time of jobs.	As compared to the well known scheduling algorithms-FIFO and Fair Scheduling, this algorithm yields moderate output under minimum share satisfaction, fairness and locality metrics.
Bin Ye, Xiaoshe Dong, Pengfei Zheng, Zengdong Zhu, Qiang Liu, Zhe Wang	8th ChinaGrid Annual Conference, IEEE 2013.	A Delay Scheduling Algorithm based on History Time in Heterogeneous Environments	In this, by considering the history time of the completed tasks and Delay Scheduler's strategy, a new algorithm is proposed for multi-user Hadoop Cluster.	With the proposed scheduler, more jobs can be assigned to the suitable slots for better performance of the system.
Quan Chen, Minyi Guo, Qianni Deng, Long Zheng, Song Guo, Yao Shen	The Journal of Supercomputing, June 2013.	HAT: History based Auto-Tuning MapReduce in Heterogeneous Environments	The proposed History based Auto-tuning MapReduce scheduler tunes the weight of each map and reduce tasks by their value in history tasks and uses them to calculate the progress of current tasks. It then adapts automatically to the continuous changing environment by regularly monitoring the progress of tasks.	The performance of MapReduce applications can be increased by 37% with Hadoop and by 16% with LATE scheduler by applying the proposed HAT scheduler to them.
Matei Zaharia, Andy Konwinski, Anthony D. Joseph, Randy Kartz, Ion Stocia	8th USENIX Symposium on Operating Systems Design and Implementation, ACM, 2008.	Improving MapReduce Performance in Heterogeneous Environments	In this, a robust scheduling algorithm, Longest Approximate Time to End (LATE), is proposed which uses the estimated completion times of jobs which are expected to hurt the response time the most.	On Amazon's Elastic Compute Cloud, the proposed LATE algorithm performs much better than the Hadoop's default speculative scheduling algorithm.
Quan Chen , Daqiang Zhang, Minyi Guo, Qianni Deng, Song Guo	10th IEEE International Conference on CIT, 2010	SAMR: A Self-adaptive MapReduce Scheduling Algorithm In Heterogeneous Environment	The proposed algorithm classifies the nodes into slow nodes using the historical information and further classifies them into slow map nodes and slow reduces nodes. It then launches the backup tasks in the meanwhile.	In heterogeneous environments, the proposed algorithm reduces the execution time by 24% for Sort applications by 17% for Wordcount applications as compared to default Hadoop scheduling mechanism.
Visalakshi P and Karthik TU,	IJCSNS, April 2011	MapReduce Scheduler Using Classifiers for Heterogeneous Workloads	The proposed scheduler executes at the JobTracker. At the JobTracker when a message is received by it from the TaskTracker, the scheduler, from the list of the MapReduce jobs, selects a task that is expected to yield maximum throughout.	This maintains a balance between the CPU bound and IO bound jobs by effectively classifying them and preventing the re-launching of the jobs.

Table 2. Proposed Algorithm on Data Placement along with their performance

Author's Name	Publication	Title	Proposed Work	Results
Chia-Wei Lee, Kuang-Yu Hsieh Sun-Yuan Hsieh , Hung-Chang Hsiao	Elsevier-Big Data Research, July 2014.	A Dynamic Data Placement Strategy for Hadoop in Heterogeneous Environments	The proposed Dynamic Data Placement (DDP) algorithm works in two phases: in First phase, in the Hadoop distributed file system the input is written and data are allocated to the nodes and in second phase, the capacity of the nodes is calculated and reallocation of data takes place accordingly.	The performance of the algorithm is evaluated on two functions- Wordcount and Grep mainly in heterogeneous environment for Hadoop .The algorithm with average improvement of 14.5% performs 24.7% better in case of Wordcount application . And for Grep application it performs 32.1% better with average improvement of 23.5%.
Xiaofei Hou, Ashwin Kumar T K, Johnson P Thomas, Vijay Vardharajan	4th International Conference on Big Data and Cloud Computing, IEEE 2014.	Dynamic Workload Balancing for Hadoop MapReduce	By analyzing the information obtained from the log files of Hadoop, this proposed dynamic algorithm balances the workload between the different busiest racks on the Hadoop cluster by shifting the tasks between them and idle racks.	The simulation results show that the remaining time of the tasks, which belong to the busiest racks in Hadoop Cluster, can be decreased by more than 50%.
Krish K.R., Ali Anwar, Ali R. Butt	22nd International Symposium on Modeling, Analysis & Simulation of Computer and Telecommunication Systems, IEEE 2014.	ΦSched: A Heterogeneity-Aware Hadoop Workflow Scheduler	In this, the information regarding the behavior of various leading Hadoop applications in the heterogeneous Hadoop cluster is merged into the hardware-aware ΦSched scheduler to improve the resource-application match. An instance of Hadoop Distributed File System (HDFS) is also configured on all the participating clusters to ensure the data locality.	As compared to the hardware oblivious scheduling, using the proposed method performance improvement of 18.7% can be achieved by managing four different clusters. The I/O throughput and the average I/O rate can be enhanced by 23% and 26% respectively with HDFS improvement.
Ashwin Kumar T K, Jongyeop Kim, K M George, Nohpill Park	ICACCI, IEEE 2014.	Dynamic Data Rebalancing in Hadoop	In this paper, depending on the number of incoming parallel MapReduce jobs, the proposed algorithm balances the data by dynamically replicating it with minimum cost of data movement.	For Hadoop with the proposed algorithm, MapReduce job's service time can be decreased by 30% and resource utilization can be improved up to 50%.
Zhao Li, Yao Shen, Bin Yao, Minyi Guo	International Journal of Parallel Programming, 2013.	OFScheduler: A Dynamic Network Optimizer for MapReduce in Heterogeneous Cluster	In this OFScheduler, a dynamic network optimizer is proposed which relieves the network traffic during MapReduce job execution by reducing the bandwidth competition.	For multipath heterogeneous cluster, the proposed scheduler's simulation results show 24 ~ 63% better performances of MapReduce jobs and bandwidth utilization.

Table 2. (Continued)

Author's Name	Publication	Title	Proposed Work	Results
Yuanquan Fan, Weiguo Wu, Haijun Cao, Huo Zhu, Xu Zhao, Wei Wei	7th ChinaGrid Annual Conference, IEEE 2012.	A Heterogeneity Aware Data Distribution and Rebalance Method in Hadoop Cluster	The proposed method first computes the nodes computing capability based on the log information about the history tasks. Then data is divided into different sized blocks according to the nodes computing capacity. Further the dynamic data migration policy aims at the transfer of data from slow DataNode to headmost DataNode during execution time.	The experimental results show that the execution time is reduced by 5% in Wordcount benchmark and by 9.6% in Sort Benchmark. The data locality can also be increased by 18.8% with Wordcount and by 8.3% with Sort benchmark approximately.
Jiong Xie, Shu Yin, Xiaojun Ruan, Zhiyang Ding, Yun Tian, James Majors, Adam Manzanares, Xiao Qin	International Symposium on Parallel and Distributed Processing, Workshops and PhD Forum, 2010.	Improving MapReduce Performance through Data Placement in Heterogeneous Hadoop Clusters	With the proposed method, data across the nodes can be balanced adaptively to improve the performance of data intensive application running on Hadoop cluster.	Using the proposed data placement scheme, output of Wordcount and Grep can be increased up to 33.1% and 10.2% with an average of 17.3% and 7.1%.

4. CONCLUSION

A lot of research work is being carried out on Hadoop these days especially for heterogeneous networks. The work includes the data locality, resource allocation and scheduling issues of the Hadoop. Various performance issues are being considered in these papers. Hadoop is greatly accepted by many large scale IT companies due to its parallel and distributed programming paradigm-MapReduce.

5. FUTURE SCOPE

According to my survey, there is lot of scope for research work in the heterogeneous environments. A lot of work has been proposed for the homogeneous environments. But in real world applications, homogeneous environments are impractical. As in real world, nodes are of heterogeneous nature as they have different computing capacity and different storage capacity. So, in future, algorithms can be proposed for heterogeneous environment for increasing the Hadoop's performance. The proposed algorithm can be based on any of the issues like scheduling, data locality, resource allocation to the nodes etc.

6. ACKNOWLEDGEMENTS

I would like to thank Er. Harpreet Kaur, AP & HOD of CSE Department, Rayat Bahra Group of Institues, Patiala for having faith in me and guiding me time to time in the completion of this survey paper.

7. REFERENCES

[1] Julio C.S. Anjos, Ivan Carrera, Wagner Kolberg, Andre Luis Tibols, Luciana B. Arantes, Claudio R. Geyer, "MRA++: Scheduling and data placement on MapReduce for heterogeneous environments", in Future Generation Computer Systems, vol. 42, pp. 22-35, January 2015.Xiaofei Hou, Ashwin Kumar T K, Johnson P Thomas, Vijay Vardharajan, "Dynamic Workload Balancing for Hadoop MapReduce", in proceedings of 4th International Conference on Big Data and Cloud Computing, IEEE, Dec. 2014.

[2] Zhou Tang, Min Liu, Almoalmi Ammar, Kenli Li, Keqin Li, "An Optimized MapReduce Workload Scheduling Algorithm for Heterogeneous Computing", in The Journal of Supercomputing, Nov. 2014.

[3] Krish K.R., Ali Anwar, Ali R. Butt , "ΦSched: A Heterogeneity-Aware Hadoop Workflow Scheduler", in proceedings of 22nd International Symposium on Modelling, Analysis & Simulation of Computer and Telecommunication Systems, IEEE, pp. 255-264, Sep. 2014.

[4] Chia-Wei Lee, Kuang-Yu Hsieh Sun-Yuan Hsieh , Hung-Chang Hsiao, "A Dynamic Data Placement Strategy for Hadoop in Heterogeneous Environments ", in Big Data Research, vol. 1, pp. 14-22, July 2014.

[5] Feng Yan, Ludmila Cherkasova, Zhuoyao Zhang, Evgenia Smirni, "Optimizing Power and Performance Trade-offs of MapReduce Job Processing with Heterogeneous Multi-Core Processors", in proceedings of 7th IEEE International Conference on Cloud Computing,pp. 240-247, July 2014.

[7] Jessica Hartog, Renan DelValle, Madhusudhan Govindaraju, Maichael J. Lewis, "Configuring A MapReduce Framework For Performance Heterogeneous Clusters", in proceedings of IEEE International Congress on Big Data, pp. 120-127, July 2014.

[8] Aysan Rasooli, Douglas G. Down, "COSHH: A Classification and Optimization Based Scheduler for Heterogeneous Hadoop Systems", Future Generation Computer Systems, vol. 36, pp. 1-15, July 2014.

[9] Xiaolong Xu, Lingling Cao, Xinheng Wang, "Adaptive Task Scheduling Strategy based on Dynamic Workload Adjustment (ATSDWA) for Heterogeneous Hadoop

Clusters", in IEEE Systems Journal, issue 99, pp. 1-12, June 2014.

[10] Ashwin Kumar T K, Jongyeop Kim, K M George, Nohpill Park, "Dynamic Data Rebalancing in Hadoop", in proceedings of IEEE/ACIS 13th International Conference on Computer and Information Science, pp. 315- 320, June 2014.

[11] Zhao Li, Yao Shen, Bin Yao, Minyi Guo, "OFScheduler: A Dynamic Network Optimizer for MapReduce in Heterogeneous Cluster", in International Journal of Parallel Programming, Oct. 2013.

[12] Bin Ye, Xiaoshe Dong, Pengfei Zheng, Zengdong Zhu, Qiang Liu, Zhe Wang, "A Delay Scheduling Algorithm based on History Time in Heterogeneous Environments", in proceedings of 8th ChinaGrid Annual Conference, IEEE, pp. 86-91, Aug. 2013.

[13] Sutariya Kapil B., Sowmya Kamath S., "Resource Aware Scheduling in Hadoop for Heterogeneous Workloads based on Load Estimation", in proceedings of 4th International Conference on Computing, Communications and Networking Technologies, pp. 1-5, July 2013.

[14] Quan Chen, Minyi Guo, Qianni Deng, Long Zheng, Song Guo, Yao Shen, "HAT: History based Auto-Tuning MapReduce in Heterogeneous Environments", in The Journal of Supercomputing, vol. 64, pp. 1038-1054, June 2013.

[15] Yuanquan Fan, Weiguo Wu, Haijun Cao, Huo Zhu, Xu Zhao, Wei Wei, "A Heterogeneity Aware Data Distribution and Rebalance Method in Hadoop Cluster", in proceedings of 7th ChinaGrid Annual Conference, IEEE, pp. 255-264, Sep. 2012.

[16] Visalakshi P and Karthik TU, "MapReduce Scheduler Using Classifiers for Heterogeneous Workloads", in IJCSNS, vol. 11 no. 4, April 2011.

[17] Jiong Xie, Shu Yin, Xiaojun Ruan, Zhiyang Ding, Yun Tian, James Majors, Adam Manzanares, Xiao Qin, "Improving MapReduce Performance through Data Placement in Heterogeneous Hadoop Clusters", in proceedings of International Symposium on Parallel and Distributed Processing, Workshops and PhD Forum, pp. 1-9 Apr. 2010.

[18] Quan Chen ,Daqiang Zhang, Minyi Guo, Qianni Deng,Song Guo, "SAMR: A Self-adaptive MapReduce Scheduling Algorithm In Heterogeneous Environment", in proceedings of 10th IEEE International Conference on CIT, pp. 2736-2743, 2010.

[19] Matei Zaharia, Andy Konwinski, Anthony D. Joseph, Randy Kartz, Ion Stocia, "Improving MapReduce Performance in Heterogeneous Environments", in proceedings of 8th USENIX Symposium on Operating Systems Design and Implementation, pp. 29-42, ACM Press, 2008.

[20] Ivanilton Polato, Reginaldo Re, Alfredo Goldman, Fabio Kon, "A comprehensive view of Hadoop research", in Journal of Network and Computer Applications, vol. 46, pp. 1-25, Nov. 2014.

[21] B G. Babu, Shabeera T P, Madhu Kumar S D, "Dynamic Colocation Algorithm for Hadoop", in proceedings of IEEE International Conference on Advances in Computing, Communications and Informatics, pp. 2643- 2647, Sep. 2014.

[22] S. Sujitha, Suresh Jaganathan, "Aggrandizing Hadoop in terms of Node Heterogeneity & Data Locality", in proceedings of IEEE International Conference on Smart Structures & Systems, pp. 145-151, Mar. 2013.