

# Sorting Technique- An Improvement towards HDD Utilization in HADOOP

Jyoti Karlupia  
Computer Science &  
Engineering Golden College  
of Engineering and  
Technology  
Gurdaspur, India

Rohit Mahajan  
Computer Science &  
Engineering Golden College  
of Engineering and  
Technology  
Gurdaspur, India

Mohit Angurala  
Asst. Prof.  
GNDU Regional Campus,  
Gurdaspur,  
Punjab, India

Kewal Krishan  
Department of Computer  
Science, R K Arya College,  
Nawanshshr, Punjab  
kkbanth11@gmail.com

**Abstract**— A data is tremendously increasing everyday due to updates that are updated after a particular time interval. Therefore it is very hard to handle such a large data that is updated on common basis. Ample number of techniques was proposed in order to enhance the performance and efficiency. In this paper, a technique called as “sorting technique” is implemented so as to increase the mean value. Load of hadoop cluster will decrease in steeply manner when this technique is implemented. The results clearly defines the efficiency of the technique which is used and hence it also proves to be efficient than the traditional techniques.

**Keywords**—ios, TMV etc

\*\*\*\*\*

## I. INTRODUCTION

It is very difficult to maintain the data which is quite frequently updated on regular basis. In order to deal with such huge unstructured type of data the technology of hadoop is introduced in the market. Therefore, this technique is heavily used in the industries where the database is too large to handle. There are many big advantages or characteristics of Hadoop such as scalability as one of the advantage of Hadoop. Another key benefit is that it is cost-effective. Below figure illustrates the advantages of Hadoop:

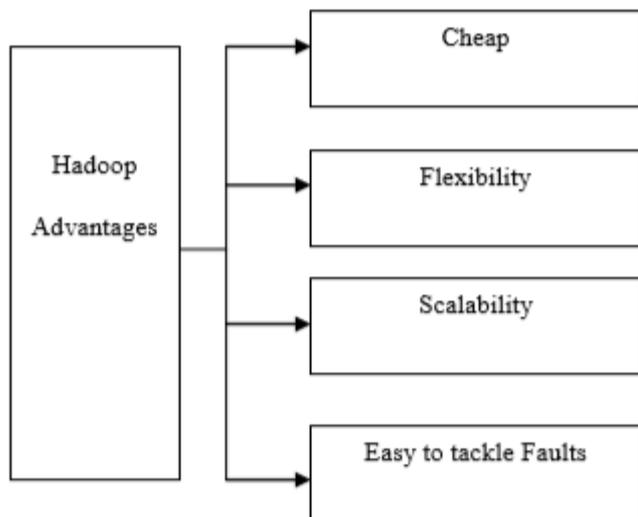


Figure 1: Advantages of Hadoop

## II. PROBLEM FORMULATION

Updates are getting increased every day, therefore to enhance performance as well as efficiency; use of data mining application is used. In the previous work they have proposed i2mapreduce that is incremental processing

extension to Map Reduce & widely used framework for mining big data. Moreover traditional work has also implemented fine grain incremental processing. The traditional work has used four algorithms and i2map reduce technique to improve performance of hadoop and calculated the mean value of Hadoop. This means value will actually show or prove performance of hadoop. Actually, the improvement is necessary because they did not sorted jobs to slaves and mean value was not good. We will improve mean value and sort the jobs to slaves and mean value will be better than the previous work. We will implement a new technique that is sorting technique to further enhance the performance and the Mean value. Firstly we will create hadoop cluster and machines. Then we will connect with master slaves and then we will calculate mean value using machines.

## III. IMPLEMENTATION WORK & RESULTS

Our implementation work shows total three experimentations which includes different input output sorting factor (ios factor). Different outputs are generated for each case when ios factor is varied for different machines.

In Our work we showed three experimentations which includes different input output sorting factor (ios factor). We generated different outputs for each case and ios factor varies for different machines.

**Experiment 1:** when ios factor is 10

hdd_utilization of machines	Total Mean Value
hdd_m1 ios is 10	1.8911
hdd_m2 ios is 10	6.3129
hdd_m3 ios is 10	6.3129

Figure 2: Readings for hdd utilization of machines when ios factor is 10

Experiment 3: when ios factor is 5  
 When we set the ios factor is 5 mean value is calculated. Result is shown in the following figures:

hdd_utilization of machines	Total Mean Value
hdd_m1 ios is 5	3.3109
hdd_m2 ios is 5	9.0516
hdd_m3 ios is 5	9.0516

Figure 6: Readings for hdd utilization of machines when ios factor is 5

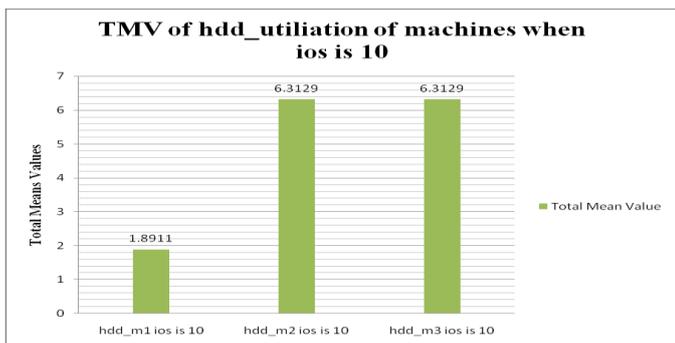


Figure 3: Total mean Value (TMV) for HDD utilization of machines when ios is 10

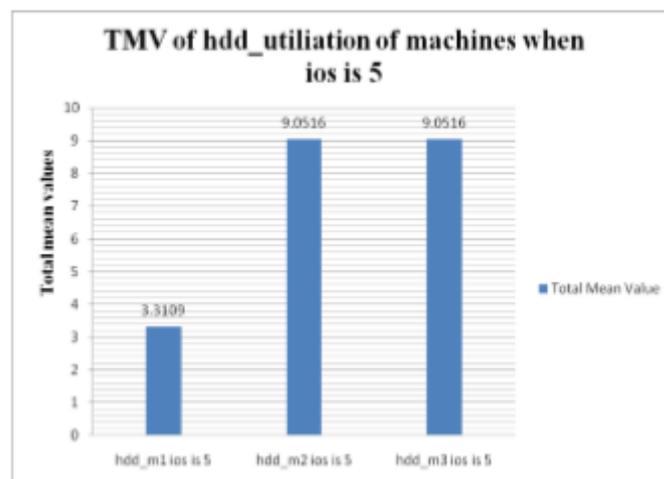


Figure 7: Total Mean Value (TMV) for HDD utilization of machines when ios factor is 5

Experiment 2: when ios factor is 8

In experiment 2 we took ios factor equal to 8. And we calculated total mean value of machine for HDD utilization of machine and the result is gathered for machine. Mean value is calculated when ios factor is 8.

hdd_utilization of machines	Total Mean Value
hdd_m1 ios is 8	2.1276
hdd_m2 ios is 8	6.9934
hdd_m3 ios is 8	6.9934

Figure 4: Readings for hdd utilization of machines when ios factor is 8

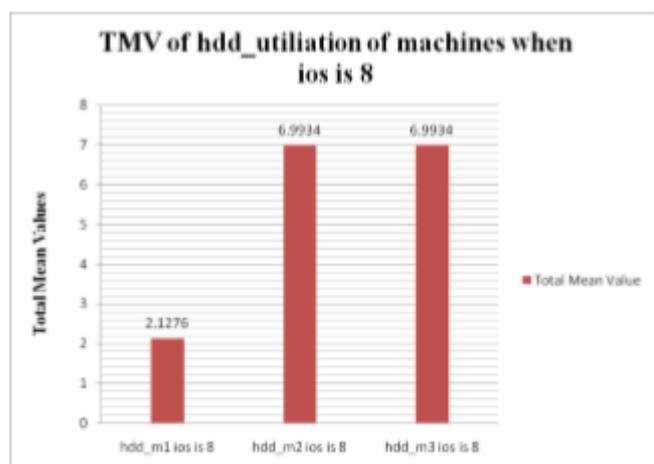


Figure 5: Total mean Value (TMV) for HDD utilization of machines when ios factor is 8

#### IV. CONCLUSION

i2mapreduce & fine grain incremental processings were used in the previous research papers. But still the performance or output was not upto the mark. In this research work total three experiments are carried out in which simulations are carried out according to ios factor which indicates performance comparisons. 3 ios factors are 5, 8 & 10. Out of these 3 ios factors when ios factor is set as 5 then the performance outcome is better than when ios is set as 8 or 10. Hence experimental results clearly indicate that ideal value of 5 is considered to be better than other two.

#### REFERENCES

- [1]. Arnab Nandi, Cong Yu, Philip Bohannon, Raghu Ramakrishna(2012),” Data CubeMaterialization and Mining over MapReduce”. IEEE Trans. Knowl. Data Eng. 24(10): 1747-1759
- [2]. Borthakur, D., Gray, J., Sarma, J. S., Muthukkaruppan, K., Spiegelberg, N., Kuang, H., & Aiyer, A. (2011, June). Apache Hadoop goes real time at Facebook. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (pp. 1071-1080).ACM.

- 
- [3]. Eltabakh, M. Y., Tian, Y., Özcan, F., Gemulla, R., Krettek, A., & McPherson, J. (2011). CoHadoop: flexible data placement and its exploitation in Hadoop. *Proceedings of the VLDB Endowment*, 4(9), 575-585.
- [4]. Leverich, J., & Kozyrakis, C. (2010). On the energy (in) efficiency of hadoop clusters. *ACM SIGOPS Operating Systems Review*, 44(1), 61-65. Ling Hu, Wei-Shinn Ku, Spiridon Bakiras, Cyrus Shahabi (2013) "Spatial Query Integrity with Voronoi Neighbors," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 4, pp. 863-876.
- [5]. Bhandarkar, M. (2010, April). MapReduce programming with apache Hadoop. In *Parallel & Distributed Processing (IPDPS), 2010 IEEE International Symposium on* (pp. 1-1).
- [6]. Liu, X., Han, J., Zhong, Y., Han, C., & He, X. (2009, August). Implementing WebGIS on Hadoop: A case study of improving small file I/O performance on HDFS. In *Cluster Computing and Workshops, 2009. CLUSTER'09. IEEE International Conference on* (pp. 1-8). IEEE. *Location-Based Spatial Queries in Mobile Environments; Engineering*, vol. 25, no. 2, pp. 260-273.