1. **Research Title -**

**Research study – An OpenSimplex approach: Performance optimization techniques for Hadoop MapReduce system by optimizing Hadoop system configuration parameters.**

1. ***Abstract* –**

Hadoop MapReduce is a popular framework for distributed storage and processing of large datasets and is used for big data analytics. It has various configuration parameters which play a key role in deciding the performance i.e., the execution time of a given big data processing job. Default values of these parameters do not result in satisfactory performance and therefore it is important to tune them. However, there is inherent difficulty in tuning the parameters due to two important reasons - first, the parameter search space is large and second, there are cross-parameter interactions. In this paper, we are conducting research on Hadoop MapReduce system’s configuration parameters, to develop new performance optimization techniques for Hadoop MapReduce system. As achieving optimal results from a Hadoop implementation begins with highly efficient system configurations. The research work mainly focuses on deep study of 190 configuration parameters of Hadoop system which includes study of parameters of Linux file system, parameters of Hadoop, Map/Reduce-Specific Configurations and Hadoop job parameter. Also, the goal of research study is to develop a dimensionality-free method which can automatically tune the configuration parameters by considering the cross-parameter dependencies. The new ‘OpenSimplex approach’ an optimization technique which will analyze the Hadoop system cross parameter configurations based on application setup and tune them to achieve desired system performance.

1. **References-**
2. **White Paper IBM- Developer Works:**  Analyze and optimize cloud cluster Performance. Use configurable parameters to monitor and tune the performance of a cloud Hadoop cluster
3. **White paper – Intel:** Hadoop\* Deployments Designing the solution stack to maximize productivity while limiting energy consumption and total cost of ownership
4. H. Lu, C. Hai-Shan and H. Ting-Ting, "Research on Hadoop Cloud Computing Model and its Applications," 2012 Third International Conference on Networking and Distributed Computing, Hangzhou, China, 2012, pp. 59-63.
5. Xin Daxin, Liu Fei. Research on optimization techniques for Hadoop cluster performance [J]. Computer Knowledge and Technology, 2011,8(7):5484~5486.

1. D. Wu and A. Gokhale, "A self-tuning system based on application Profiling and Performance Analysis for optimizing Hadoop MapReduce cluster Configuration, 20th Annual International Conference on High Performance Computing, Bangalore, 2013, pp. 89-98.
2. Z. Bei *et al*., "RFHOC: A Random-Forest Approach to Auto-Tuning Hadoop's Configuration," in *IEEE Transactions on Parallel and Distributed Systems*, vol. 27, no. 5, pp. 1470-1483, May 1 2016.
3. C. Li *et al*., "An Adaptive Auto-configuration Tool for Hadoop," *2014 19th International Conference on Engineering of Complex Computer Systems*, Tianjin, 2014, pp. 69-72.
4. C. O. Chen, Y. Q. Zhuo, C. C. Yeh, C. M. Lin and S. W. Liao, "Machine Learning-Based Configuration Parameter Tuning on Hadoop System," *2015 IEEE International Congress on Big Data*, New York, NY, 2015, pp. 386-392.
5. J. Kim and N. Park, "Identification of the Optimal Hadoop Configuration Parameters Set for Mapreduce Computing," *2015 3rd International Conference on Applied Computing and Information Technology/2nd International Conference on Computational Science and Intelligence*, Okayama, 2015, pp. 108-112.
6. K. Wang, X. Lin and W. Tang, "Predator — An experience guided configuration optimizer for Hadoop MapReduce," *4th IEEE International Conference on Cloud Computing Technology and Science Proceedings*, Taipei, 2012, pp. 419-426.