both containernet and real cluster to run various tasks, we will be able to ensure that our algorithm can increase the performance of Hadoop system, especially under heterogeneous setup.

REFERENCES

[1] D. Reinsel, J. Gantz, and J. Rydning, "The digitisation of the world: from edge to core," *IDC White Paper*, 2018.

[2] Y. Sun, Y. Shi, and Z. Zhang, "Finance Big Data: Management, Analysis, and Applications," *Int. J. Electron. Commer.*, vol. 23, pp. 9–11, 2019.

[3] M. Nakagami, J. A. B. Fortes, and S. Yamaguchi, "Job-Aware Optimization of File Placement in Hadoop," *2019 IEEE 43rd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2, pp. 664–669, 2019.

[4] X. Luo and X. Fu, "Configuration optimisation method of Hadoop system performance based on genetic simulated annealing algorithm," *Cluster Computing*, pp. 1–9, 2018.

[5] P. Garraghan, X. Ouyang, R. Yang, D. McKee, and J. Xu, "Straggler Root-Cause and Impact Analysis for Massive-scale Virtualized Cloud Datacenters," *IEEE Transactions on Services Computing*, vol. 12, pp. 91–104, 2019.

[6] H.-G. Kim, "Effects of Design Factors of HDFS on I/O Performance," *J. Comput. Sci.*, vol. 14, pp. 304–309, 2018.

[7] D. Choi, M. Jeon, N. Kim, and B.-D. Lee, "An Enhanced Data-Locality-Aware Task Scheduling Algorithm for Hadoop Applications," *IEEE Systems Journal*, vol. 12, pp. 3346–3357, 2018.

[8] X. Du, Y. Liu, and C. Zhao, "A Hadoop Yarn Scheduling Based on Node Computing Capability and Data Locality in Heterogeneous Environments," 2018.

[9] K. Midoun, W.-K. Hidouci, M. Loudini, and D. Belayadi, "RTSBL: Reduce Task Scheduling Based on the Load Balancing and the Data Locality in Hadoop," 2018.

[10] P. Zhang, C. Li, and Y. Zhao, "An Improved Task Scheduling Algorithm Based on Cache Locality and Data Locality in Hadoop," *2016 17th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, pp. 244–249, 2016.

[11] K. Kalia and N. Gupta, "A Review on Job Scheduling for Hadoop Mapreduce," *2017 International Conference on Next Generation Computing and Information Systems (ICNGCIS)*, pp. 75–79, 2017.

[12] A. Sharma and G. Singh, "A Review of Scheduling Algorithms in Hadoop," 2020.

[13] A. M. S. Lakshmi, N. S. Chandra, and M. BalRaju, "Optimised Capacity Scheduler for MapReduce Applications in Cloud Environments," 2019.

[14] H. Chen and D. Cui, "SLA-based Hadoop Capacity Scheduler Algorithm," 2015.

[15] J. A. Murali and T. Brindha, "Analysis of Scheduling Algorithms in Hadoop," 2018.

[16] J. V. Gautam, H. B. Prajapati, V. K. Dabhi, and S. Chaudhary, "Empirical Study of Job Scheduling Algorithms in Hadoop MapReduce," *Cybernetics and Information Technologies*, vol. 17, pp. 146–163, 2017.

[17] Y. Xu *et al.*, "RAPID: Avoiding TCP Incast Throughput Collapse in Public Clouds With Intelligent Packet Discarding," *IEEE Journal on Selected Areas in Communications*, vol. 37, pp. 1911–1923, 2019.

[18] P. Pandey, S. Singh, and S. Singh, "Cloud computing," in *ICWET*, 2010.

[19] B. T. Rao, N. V. Sridevi, V. K. Reddy, and L. S. S. Reddy, "Performance Issues of Heterogeneous Hadoop Clusters in Cloud Computing," *ArXiv*, vol. abs/1207.0894, 2012.

[20] S. Shankland, "Google spotlights data center inner workings," *CNET*. https://www.cnet.com/news/google-spotlights-data-center-inner-workings/ (accessed Jun. 07, 2020).

[21] M. Liroz-Gistau, R. Akbarinia, D. Agrawal, and P. Valduriez, "FP-Hadoop: Efficient processing of skewed MapReduce jobs," *Information Systems*, vol. 60, pp. 69–84, 2016.

[22] R. Patgiri and R. Das, "rTuner: A Performance Enhancement of MapReduce Job," in *ICCMS 2018*, 2018.

[23] S. Ghemawat *et al.*, "Performance Tuning and Scheduling of Large Data Set Analysis in Map Reduce Paradigm by Optimal Configuration using Hadoop," 2019.

[24] X. Hua, M. C. Huang, and P. Liu, "Hadoop Configuration Tuning with Ensemble Modeling and Metaheuristic Optimization," *IEEE Access*, vol. 6, pp. 44161–44174, 2018.

[25] M. A. Rahman, A. Hossen, J. Hossen, C. Venkataseshaiah, T. Bhuvaneswari, and A. Sultana, "Towards machine learning-based self-tuning of Hadoop-Spark system," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 15, p. 1076, 2019.

[26] W. Wang, Y. Shi, X. Liu, Y. Feng, and N. Tao, "Hadoop Performance Tuning based on Parameter Optimization," 2018.

[27] Y. Guo, J. Rao, C. Jiang, and X. Zhou, "Moving Hadoop into the cloud with flexible slot management and speculative execution," *IEEE Transactions on Parallel and Distributed systems*, vol. 28, no. 3, pp. 798–812, 2016.

[28] Y. Guo, J. Rao, C. Jiang, and X. Zhou, "Moving Hadoop into the Cloud with Flexible Slot Management and Speculative Execution," *IEEE Transactions on Parallel and Distributed Systems*, vol. 28, pp. 798–812, 2017.

[29] X. Huang, L. Zhang, R. Li, L. Wan, and K. Li, "Novel heuristic speculative execution strategies in heterogeneous distributed environments," *Computers & Electrical Engineering*, vol. 50, pp. 166–179, 2016.

[30] D. C. Vinutha and G. T. Raju, "Evolutionary Approach based Scheduler for Speculative Task Execution," *2019 1st International Conference on Advances in Information Technology (ICAIT)*, pp. 485–490, 2019.

[31] Q. Liu, W. Cai, J. Shen, Z. Fu, X. Liu, and N. Linge, "A speculative execution strategy based on node classification and hierarchy index mechanism for heterogeneous Hadoop systems," *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pp. 889–894, 2017.

[32] M. Zaharia, A. Konwinski, A. D. Joseph, R. H. Katz, and I. Stoica, "Improving MapReduce Performance in Heterogeneous Environments," in *OSDI*, 2008.

[33] S. R. Pakize, "A Comprehensive View of Hadoop MapReduce Scheduling Algorithms," 2014.

[34] M. Beckert and R. Ernst, "Response time analysis for sporadic server-based budget scheduling in real time virtualisation environments," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 16, no. 5s, pp. 1–19, 2017.

[35] F. Kaltenberger, C. Roux, M. Buczkowski, and M. Wewior, "The OpenAirInterface application programming interface for schedulers using Carrier Aggregation," in *2016 International Symposium on Wireless Communication Systems (ISWCS)*, 2016, pp. 497–500.

[36] M. Peuster, J. Kampmeyer, and H. Karl, "Containernet 2.0: A Rapid Prototyping Platform for Hybrid Service Function Chains," *2018 4th IEEE Conference on Network Softwarization and Workshops (NetSoft)*, pp. 335–337, 2018.