REFERENCES

[1] "Hadoop - Big Data Overview - Tutorialspoint", Tutorialspoint.com, 2020. [Online]. Available: https://www.tutorialspoint.com/hadoop/hadoop_big_data_overview.htm. [Accessed: 06- Dec- 2020].

[2] "Hadoop YARN Architecture - GeeksforGeeks", GeeksforGeeks. [Online]. Available: https://www.geeksforgeeks.org/hadoop-yarn-architecture/. [Accessed: 09- May- 2018].

[3] "HDFS Tutorial - A Complete Hadoop HDFS Overview - DataFlair", DataFlair. [Online]. Available: https://data-flair.training/blogs/hadoop-hdfs-tutorial/. [Accessed: 18- Apr- 2018].

[4] B. Mishra, "Big Data Analysis Using Hadoop Map Reduce," Int. Res. J. Comput. Sci., vol. 07, no. 05, pp. 114–122, 2020, doi: 10.26562/irjcs.2020.v0705.005.

[5] R. K. Chawda and G. Thakur, "Big data and advanced analytics tools," in 2016 Symposium on Colossal Data Analysis and Networking (CDAN), 2016, pp. 1–8, doi: 10.1109/CDAN.2016.7570890.

[6] "Why only one of the 5 Vs of big data really matters", IBM Big Data & Analytics Hub. [Online]. Available: https://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters. [Accessed: 06- Sep- 2018].

[7] P. Chandarana and M. Vijayalakshmi, "Big Data analytics frameworks," in 2014 International Conference on Circuits, Systems, Communication and Information Technology Applications (CSCITA), 2014, pp. 430–434, doi: 10.1109/CSCITA.2014.6839299.

[8] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," IEEE Netw., vol. 28, no. 4, pp. 46–50, 2014, doi: 10.1109/MNET.2014.6863131.

[9] Y. Perwej, B. Kerim, M. S. Adrees, and O. E. Sheta, "An Empirical Exploration of the Yarn in Big Data," Int. J. Appl. Inf. Syst., vol. 12, no. 9, pp. 19–29, 2017, doi: 10.5120/ijais2017451730.

[10] Y. Zhai, Y. Ong, and I. W. Tsang, "The Emerging 'Big Dimensionality,'" IEEE Comput. Intell. Mag., vol. 9, no. 3, pp. 14–26, 2014, doi: 10.1109/MCI.2014.2326099.

[11] B. Mandal, S. Sethi, and R. K. Sahoo, "Architecture of efficient word processing using Hadoop MapReduce for big data applications," in 2015

International Conference on Man and Machine Interfacing (MAMI), 2015, pp. 1–6, doi: 10.1109/MAMI.2015.7456612.

[12] “Hadoop - Introduction,” Tutorialspoint. [Online]. Available: https://www.tutorialspoint.com/hadoop/hadoop introduction.htm. [Accessed: 18-Dec-2018].

[13] K. Siddardha and C. Suresh, “Big Data Analytics : Challenges , Tools and Limitations,” vol. 0869, no. 3, pp. 40–43, 2016.

[14] H. M. Admin, “Top 10 Open Source Big Data Tools,” Digital Trans- formation Blogs Bigdata IoT M2M Mobility Cloud, 31-May-2016. [Online]. Available: http://www.happiestminds.com/blogs/top-10-open- source-big-data-tools/. [Accessed: 05-Sep-2020].

[15] "Big Data Analytics | Qubole", Qubole. [Online]. Available: https://www.qubole.com/big-data-analytics/. [Accessed: 06- Dec- 2018].

[16] N. Do and T. Srivastava, "Nobody Tells You - 5 things Big Data 'CAN' and 'Cannot' Do", Analytics Vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2015/11/5-big-data-can-cannot/. [Accessed: 06- Dec- 2018].

[17] “Natural Language Toolkit¶,” Natural Language Toolkit - NLTK 3.5 documentation. [Online]. Available: http://nltk.org/. [Accessed: 25-Apr- 2020].

[18] Tweepy. [Online]. Available: http://tweepy.github.com/. [Accessed: 25- May- 2020].

[19] S. Shengtao, W. Aizhi, and L. Xiaoyang, “The study of a hierarchical hadoop architecture in multiple data centers environment,” Open Cybern. Syst. J., vol. 9, no. 1, pp. 131–137, 2015, doi: 10.2174/1874110X01509010131.

[20] C. Y. Wang, T. L. Tai, S. Jui-Shing, C. Jyh-Biau, and S. Ce-Kuen, “Federated MapReduce to transparently run applications on multicluster environment,” Proc. - 2014 IEEE Int. Congr. Big Data, BigData Congr. 2014, pp. 296–303, 2014, doi: 10.1109/BigData.Congress.2014.50.

[21] Y. Luo and B. Plale, “Hierarchical MapReduce programming model and scheduling algorithms,” Proc. - 12th IEEE/ACM Int. Symp. Clust. Cloud Grid Comput. CCGrid 2012, pp. 769–774, 2012, doi: 10.1109/CC- Grid.2012.132.

[22] Y. Luo, Z. Guo, Y. Sun, B. Plale, J. Qiu, and W. W. Li, “A hierarchical framework for cross-domain MapReduce execution,” ECMLS’11 - Proc. 2nd Int. Work. Emerg. Comput. Methods Life Sci., pp. 15–22, 2011, doi: 10.1145/1996023.1996026.

[23] L. Wang, J. Tao, Y. Ma, S. U. Khan, J. Kołodziej, and D. Chen, “Software design and implementation for mapreduce across distributed data centers,” Appl. Math. Inf. Sci., vol. 7, no. 1 L, pp. 85–90, 2013, doi: 10.12785/amis/071L13.

[24] H. C. Yang, A. Dasdan, R. L. Hsiao, and D. S. Parker, “Map-reduce- merge:

Simplified relational data processing on large clusters," Proc. ACM SIGMOD Int. Conf. Manag. Data, pp. 1029–1040, 2007, doi: 10.1145/1247480.1247602.

[25] I. Tomasic, A. Rashkovska, and M. Depolli, "Using Hadoop MapReduce in a Multicluster Environment," Proc. 36th Int. Conv. MIPRO, pp. 369–374, 2013.

[26] DeZyre. [Online]. Available: https://www.dezyre.com/article/hadoop- cluster-overview-what-it-is- and-how-to-setup-one/356. [Accessed: 25- Mar-2020].

[27] Joe, C. Vijesh, Jennifer S. Raj, and S. Smys. "Big Data Analytics: Tools, Challenges, and Scope in Data-Driven Computing", International Conference on Mobile Computing and Sustainable Informatics, pp. 709-719. Springer, Cham, 2020

[28] A. Sungheetha and R. Sharma, "TransCapsule Model for Sentiment Classification", Journal of Artificial Intelligence, vol. 2, no. 3, pp. 163-169, 2020. Available: 10.36548/jaicn.2020.3.003.

[29] S. Voruganti, "Map Reduce a Programming Model for Cloud Computing Based On Hadoop Ecosystem," Int. J. Comput. Sci. Inf. Technol., vol. 5, no. 3, pp. 3794–3799, 2014.

[30] Spark.apache.org. n.d. Apache Spark™ - Unified Analytics Engine For Big Data. [online] Available at: <https://spark.apache.org/> [Accessed 7 August 2020]..

[31] Storm.apache.org. n.d. Apache Storm. [online] Available at: <https://storm.apache.org/>.

[32] Chen He ; Weitzel, D. ; Swanson, D. ; Ying Lu, "HOG: Distributed Hadoop MapReduce on the Grid". High Performance Computing, Networking, Storage and Analysis (SCC),IEEE 2012.

[33] H.Lin, X.Ma, J. Archuleta, W. Feng, M. Gardner, and Z. Zhang, "MoonL MapReduce on opportunistic environments," in Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing. ACM, 2010, pp. 95-106.

[34] A. Iordache, C. Morin, N. Parlavantzas, E. Feller, et al. "Resilin: Elastic MapReduce over Multiple Clouds". 13th IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing. IEEE/ACM 2013.

[35] "Amazon EC2", Amazon Web Services, Inc.. [Online]. Available: http://aws.amazon.com/ec2/.

[36] B. Sharma, T. Wood, et al. "HybridMR: A Hierarchical MapReduce Scheduler for Hybrid Data Centers", IEEE 33rd International Conference on Distributed Computing Systems, 2013.

[37] L. Columbus, "10 Charts That Will Change Your Perspective Of Big Data's Growth", Forbes, 2020. [Online]. Available: https://www.forbes.com/sites/louiscolumbus/2018/05/23/10-charts-that-will-change-your-perspective-of-big-datas-growth/?sh=7b310adc2926. [Accessed:

12- Dec- 2020].

[38] "Executive Summary: Data Growth, Business Opportunities, and the IT Imperatives | The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things", Emc.com, 2020. [Online]. Available: https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm. [Accessed: 09- Dec- 2020].

[39] "Top 20 Big Data Statistics for 2020 - Sigma Blog", Sigma Blog, 2020. [Online]. Available: https://www.sigmacomputing.com/blog/top-20-big-data-statistics-for-2020/. [Accessed: 09- Dec- 2020].

[40] "Google Search Statistics", 2020. [Online]. Available: https://www.internetlivestats.com/google-search-statistics/. [Accessed: 11- Dec- 2020].

[41] S. Hashmi, "50+ WhatsApp Stats & Facts You Must Know in 2020", Connectiva Systems, 2020. [Online]. Available: https://www.connectivasystems.com/whatsapp-facts-stats-2020/. [Accessed: 09- Dec- 2020].

[42] "25+ Big Data Statistics - How Big It Actually Is in 2020?", TechJury, 2020. [Online]. Available: https://techjury.net/blog/big-data-statistics/#gref. [Accessed: 14- Dec- 2020].

[43] "Domo Resource - Data Never Sleeps 5.0", Domo.com, 2020. [Online]. Available: https://www.domo.com/learn/data-never-sleeps-5?aid=ogsm072517_1&sf100871281=1. [Accessed: 09- Dec- 2020].

[44] "How to Create a Business Case for Data Quality Improvement", 2020. [Online]. Available: https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement/. [Accessed: 12- Dec- 2020].

[45] P. Singh, P. Choudhury, A. Dey and P. Pramanik, "Recommender systems: an overview, research trends, and future directions", International Journal of Business and Systems Research, vol. 15, no. 1, p. 14, 2020. Available: 10.1504/ijbsr.2021.10033303.

[46] Ishwarappa and J. Anuradha, “A brief introduction on big data 5Vs characteristics and hadoop technology,” Procedia Comput. Sci., vol. 48, no. C, pp. 319–324, 2015, doi: 10.1016/j.procs.2015.04.188.

[47] Spark.apache.org. n.d. Apache Spark™ - Unified Analytics Engine For Big Data. [online] Available at: <https://spark.apache.org/> [Accessed 7 August 2020]..

[48] Storm.apache.org. n.d. Apache Storm. [online] Available at: <https://storm.apache.org/>.

[49] Khan, N., Yaqoob, I., Hashem, I., Inayat, Z., Mahmoud Ali, W., Alam, M.,

Shiraz, M. and Gani, A., 2014. Big Data: Survey, Technologies, Opportunities, and Challenges. The Scientific World Journal, 2014, pp.1-18.

[50] Han Hu, Yonggang Wen, Tat-Seng Chua and Xuelong Li, 2014. Toward Scalable Systems for Big Data Analytics: A Technology Tutorial. IEEE Access, 2, pp.652-687.

[51] DigitalOcean. n.d. Hadoop, Storm, Samza, Spark, And Flink: Big Data Frameworks Compared | Digitalocean. [online] Available at: <https://www.digitalocean.com/community/tutorials/hadoop-storm-samza-spark-and-flink-big-data-frameworks-compared#apache-storm>.

[52] L. Wang et al., "MapReduce across Distributed Clusters for Data-intensive Applications," 2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum, Shanghai, 2012, pp. 2004-2011, doi: 10.1109/IPDPSW.2012.249.