REFERENCES

Banda, M., and Ngassam, E. (2017). A data management and analytic model for business intelligence applications. *2017 IST-Africa Week Conference (IST-Africa)*.

Mollova, S., Georgieva, P,. and Kostadinov, A. (2018). Fault-tolerance of a Laboratory Computer Cluster. *2018 20th International Symposium on Electrical Apparatus and Technologies (SIELA)*.

Mollova, S., Zhekov, M., Kostadinov, A. and Georgieva, P. (2018). Laboratory model for research on computer cluster systems. 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO).

Sultana, Z. (2012). Analysis and Implementation of Cluster Computing Using Linux Operating System. IOSR Journal of Computer Engineering, 2(3), pp.06-11.

Khabibi, A.D., Satrya, G.B., Herutomo, A. (2014). Analisa Kinerja Capacity Scheduling Algorithm pada Sitem Job scheduling *Hadoop*. Bandung: Fakultas Informatika, Universitas Telkom.

Saouabi, M. and Ezzati, A. (2017). A comparative between *hadoop MapReduce* and *Apache Spark* on HDFS. *Proceedings of the 1st International Conference on Internet of Things and Machine Learning - IML '17*.

Gopalani, S. and Arora, R. (2015). Comparing *Apache Spark* and Map Reduce with Performance Analysis using *K-Means*. *Internatiodnal Journal of Computer Applications*, 113(1), pp.8-11.

Jianchao, T., Shuqiang, Y., Chaoqiang, H. and Zhou, Y. (2016). Design and Implementation of Scheduling Pool Scheduling Algorithm Based on Reuse of Jobs in *Spark*. *2016 IEEE First International Conference on Data Science in Cyberspace (DSC)*.

Chen, H. and Wang, F. (2015). *Spark* on entropy: A reliable & efficient scheduler for low-latency parallel jobs in heterogeneous cloud. *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*.

Matthews, R. (1997). The Science of Murphy's Law. *Scientific American*, 276(4), pp.88-91.

Hadjar, K. and Jedidi, A. (2019). A New Approach for Scheduling Tasks and/or Jobs in *Big Data* Cluster. *2019 4th MEC International Conference on Big Data and Smart City (ICBDSC)*.

Samadi, Y., Zbakh, M. and Tadonki, C. (2016). Comparative study between *Hadoop* and *Spark* based on *HiBench* benchmarks. *2016 2nd International Conference on Cloud Computing Technologies and Applications (CloudTech)*.

Cheng, D., Zhou, X., Lama, P., Wu, J. and Jiang, C. (2017). Cross-Platform Resource Scheduling for *Spark* and *MapReduce* on *YARN*. *IEEE Transactions on Computers*, 66(8), pp.1341-1353.

Wani, M. and Jabin, S. (2017). *Big Data*: Issues, Challenges, and Techniques in Business Intelligence. *Advances in Intelligent Systems and Computing*, pp.613-628.

Sas.com. (2019). *What is Big Data and why it matters*. [online] Available at: https://www.sas.com/en_id/insights/big-data/what-is-big-data.html [Accessed 21 Oct. 2019].

Spark .Apache.org. (2019). *Apache Spark TM - Unified Analytics Engine for Big Data*. [online] Available at: http://*Spark .Apache*.org/. [Accessed 1 Jun. 2019].

Buyya, R. (1999). *High performance cluster computing*. Upper Saddle River, N.J.: Prentice Hall PTR.

Yeo, C., Buyya, R., Pourreza, H., Eskicioglu, R., Graham, P. and Sommers, F. (n.d.). Cluster Computing: High-Performance, High-Availability, and High-Throughput Processing on a Network of Computers. *Handbook of Nature-Inspired and Innovative Computing*, pp.521-551.

Pavković, N., Skala, T. and Vidić, V. (2013). Automatic Enlarge and Deployment of Computer Cluster Using Dual-Boot Approach. *Automatika*, 54(2), pp.242-251.

Paquette, D. (2019). *Best Practice: Always 3 Nodes Minimum in a Cluster | Scale Computing*. [online] Scale Computing. Available at: https://www.scalecomputing.com/blog/best-practice-always-3-nodes-minimum-in-a-cluster [Accessed 26 Oct. 2019].

Mercier, M., Glesser, D., Georgiou, Y. and Richard, O. (2017). *Big Data* and *HPC* collocation: Using *HPC* idle resources for *Big Data* analytics. *2017 IEEE International Conference on Big Data (Big Data)*.

Vugt, S. (2014). *Pro Linux High Availability Clustering*. 1st ed. Berkeley, CA: Apress.

Moniruzzaman, A., Waliullah, M. and Rahman, M. (2015). A High Availability Clusters Model Combined with Load Balancing and Shared Storage Technologies for Web Servers. *International Journal of Grid and Distributed Computing*, 8(1), pp.109-120.

Utrera, G., Corbalan, J., and Labarta, J. (2014). Scheduling parallel jobs on multicore clusters using CPU oversubscription. *The Journal of Supercomputing*, 68(3), pp.1113-1140.

Liu, J., Ding, N,. and Xu, J. (2010). The Research of the Double Master/Slave System Architecture in the MPI Parallel Environment. *2010 International Conference on Electrical and Control Engineering*.

Talwani, S. and Chana, I. (2017). Fault tolerance techniques for scientific applications in cloud. *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*.

Almasi, G., and Gottlieb, A. (1989). *Highly parallel computing*. Redwood City: Benjamin Cummings.

Patterson, D. (2013). *Computer organization and design*. San Francisco: Elsevier Science & Technology.

Flynn, M. (1972). Some Computer Organizations and Their Effectiveness. *IEEE Transactions on Computers*, C-21(9), pp.948-960.

K Kazemi, U. (2017). Clustering methods in *Big Data*. Journal of Embedded Systems and Processing, 2(3), pp.1-5.

Amazon Web Services, Inc. (2019). What is *Apache Spark*? | Introduction to *Apache Spark* and Analytics | AWS. [online] Available at: https://aws.amazon.com/big-data/what-is-*Spark* / [Accessed 22 Sep. 2019].

Sharma, M., Chauhan, S. and Kishore, K. (2016). A review: Map reduce and *Spark* for *Big Data* analytics. *International Journal of Advanced Technology in Engineering and Science*, 4(6), pp.42-50.

Gautam, J., Prajapati, H., Dabhi, V. and Chaudhary, S. (2015). A survey on job scheduling algorithms in *Big Data* processing. *2015 IEEE International Conference on Electrical, Computer*

, and Communication Technologies (ICECCT).

Usama, M., Liu, M. and Chen, M. (2017). Job schedulers for *Big Data* processing in *Hadoop* environment: testing real-life schedulers using benchmark programs. *Digital Communications and Networks*, 3(4), pp.260-273.

White, T. (2015). *"Hadoop: The Definitive Guide, 4th Edition"*. 4th ed. O'Reilly Media, Inc.

Gavali, V., Mane, T. Jumnake, G., and M, S. (2018). A *Big Data*: Tools, System,s and Benchmarks. *2018 International Conference on Information, Communication, Engineering, and Technology (ICICET)*.

Ahn, H., Kim, H., and You, W. (2018). Performance Study of *Spark* on *YARN* Cluster Using *HiBench*. *2018 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*.

Kay, J. and Lauder, P. (1988). A fair share scheduler. *Communications of the ACM*, 31(1), pp.44-55.

Jain, R., Chiu, D., and Hawe, W. (1984). A quantitative measure of fairness and discrimination for resource allocation in shared computer system. Hudson, Mass.: Eastern Research Laboratory, Digital Equipment Corp.

Schmidt, J., And Osebold, R. (2017). Environmental management systems as a driver for sustainability: state of implementation, benefits, and barriers in german construction companies. Journal of Civil Engineering and Management, 23(1), Pp.150-162.