

REFERENCES

- Artini, M., Atzori, C., Bardi, A., La Bruzzo, S. and Manghi, P., 2014, September. *TagTick: A tool for annotation tagging over solr indexes*. In Digital Libraries (JCDL), 2014 IEEE/ACM Joint Conference on (pp. 407-408). IEEE.
- Bai, J., 2013, July. *Feasibility analysis of big log data real time search based on Hbase and ElasticSearch*. In Natural Computation (ICNC), 2013 Ninth International Conference on (pp. 1166-1170). IEEE.
- Chang, B.R., Tsai, H.F., Wang, Y.A. and Chen, C.Y., 2015, November. *Realization of secondary indexing to NoSQL database with intelligent adaptation*. In Technologies and Applications of Artificial Intelligence (TAAI), 2015 Conference on (pp. 449-452). IEEE.
- Chen, D., Chen, Y., Brownlow, B.N., Kanjamala, P.P., Arredondo, C.A.G., Radspinner, B.L. and Raveling, M.A., 2016. *Real-Time or Near Real-Time Persisting Daily Healthcare Data into HDFS and ElasticSearch Index inside a Big Data Platform*. IEEE Transactions on Industrial Informatics.
- Dean, J. and Ghemawat, S., 2008. *MapReduce: simplified data processing on large clusters*. Communications of the ACM, 51(1), pp.107-113.
- Fan, J., Li, X., Liu, C.H., Buell, J., Lu, G. and Lu, L., 2014, June. *Diagnosing Virtualized Hadoop Performance from Benchmark Results: An Exploratory Study*. In Big Data (BigData Congress), 2014 IEEE International Congress on (pp. 578-585). IEEE.
- Gupta, S. and Rani, R., 2016, August. *A comparative study of elasticsearch and CouchDB document oriented databases*. In Inventive Computation Technologies (ICICT), International Conference on (Vol. 1, pp. 1-4). IEEE.
- Holmes, A., 2012. *Hadoop in practice*. Manning Publications Co..
- Howe, D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., Hill, D.P., Kania, R., Schaeffer, M., St Pierre, S. and Twigger, S., 2008. *Big data: The future of*

- biocuration*. Nature, 455(7209), pp.47-50.
- Kuč, R. and Rogoziński, M., 2013. *Mastering ElasticSearch*. Packt Publishing Ltd.
- Kuč, R. and Rogozinski, M., 2016. *ElasticSearch server*. Packt Publishing Ltd.
- Kotsiantis, S.B., Zaharakis, I. and Pintelas, P., 2007. *Supervised machine learning: A review of classification techniques*.
- Lam, C., 2010. *Hadoop in action*. Manning Publications Co..
- Li, X.M. and Wang, Y.Y., 2015, September. *Design and Implementation of an Indexing Method Based on Fields for Elasticsearch*. In Instrumentation and Measurement, Computer, Communication and Control (IMCCC), 2015 Fifth International Conference on (pp. 626-630). IEEE.
- Mai, J.E., 1999. *Deconstructing the indexing process*. In Advances in Librarianship (pp. 269-298). Emerald Group Publishing Limited.
- Mattmann, C. and Zitting, J., 2011. *Tika in action*. Manning Publications Co..
- Moreira, L.J. and Silva, L.A., 2016, July. *Data classification combining Self-Organizing Maps and Informative Nearest Neighbor*. In Neural Networks (IJCNN), 2016 International Joint Conference on (pp. 706-713). IEEE.
- Paro, A., 2015. *ElasticSearch cookbook*. Packt Publishing Ltd.
- Thacker, U., Pandey, M. and Rautaray, S.S., 2016, March. *Performance of elasticsearch in cloud environment with nGram and non-nGram indexing*. In Electrical, Electronics, and Optimization Techniques (ICEEOT), International Conference on (pp. 3624-3628). IEEE.
- Tian, F. and Chen, K., 2011, July. *Towards optimal resource provisioning for running mapreduce programs in public clouds*. In Cloud Computing (CLOUD), 2011 IEEE International Conference on (pp. 155-162). IEEE.
- Vavilapalli, V.K., Murthy, A.C., Douglas, C., Agarwal, S., Konar, M., Evans, R., Graves, T., Lowe, J., Shah, H., Seth, S. and Saha, B., 2013, October. *Apache hadoop yarn: Yet another resource negotiator*. In *Proceedings of the 4th annual Symposium on Cloud Computing* (p. 5). ACM.

- Vellaipandiyan, S. and Srikrishnan, V., 2014, December. An approach to discover the best-fit factors for the optimal performance of Hadoop map reduce in virtualized environment. In Computational Intelligence and Computing Research (ICCIC), 2014 IEEE International Conference on (pp. 1-5). IEEE.
- Vokorokos, L., Uchnár, M. and Leščišin, L., 2016, November. Performance optimization of applications based on non-relational databases. In Emerging eLearning Technologies and Applications (ICETA), 2016 International Conference on (pp. 371-376). IEEE
- White, T., 2012. *Hadoop: The definitive guide*. " O'Reilly Media, Inc."
- Zhou, P., Lei, J. and Ye, W., 2011. Large-scale data sets clustering based on mapreduce and hadoop. *Journal of Computational Information Systems*, 7(16), pp.5956-5963.