

## DAFTAR PUSTAKA

- Ahmed, N., Barczak, A. L., Rashid, M. A., & Susnjak, T. (2021). Spark introduced a new data abstraction technique called resilient distributed dataset (RDDs) [3] that improves multiple applications' performances. Its application execution time is an essential factor in measuring real-time processing where the optimum . *Big Data and Cognitive Computing*, 65.
- Airflow, A. (n.d.). DAGs. Retrieved from Apache Airflow Website: <https://airflow.apache.org/docs/apache-airflow/stable/core-concepts/dags.html>
- Chen, H., Chiang, R. H., & Storey, V. C. (2021). Business Intelligence and Analytics: From Big Data to Big Impact. *MIS Quarterly*, 1165-1188.
- Day, M.-Y., Cheng, T.-K., & Li, J.-G. (2018). AI Robo-Advisor with Big Data Analytics for Financial Services. *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. Barcelona: IEEE.
- Fan, W. (2015). Data Quality: From Theory to Practice. *ACM SIGMOD Record*, 7-18.
- García, S., Ramírez-Gallego, S., Luengo, J., Benítez, J. M., & Herrera, F. (2016). Big data preprocessing: methods and prospects. *Big Data Analytics volume 1*, 1-9.
- Hongxun, T., Hinggang, W., Kun, Z., Mingtai, S., Haosong, L., Zhongping, X., . . . Yaqi, C. (2018). Data Quality Assessment for On-line Monitoring and Measuring System of Power Quality Based on Big Data and Data Provenance Theory. *2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)*. Chengdu: IEEE.
- Juneja, A., & Das, N. N. (2019). Big Data Quality Framework: Pre-Processing Data in Weather Monitoring Application. *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*. Faridabad: IEEE.
- Mishra, D. D., Pathan, S., & Murthy, C. (2018). Apache Spark Based Analytics of Squid Proxy Logs. *2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS)*. Indore: IEEE.

- MongoDB. (n.d.). *Schema Validation*. Retrieved from MongoDB Website: <https://www.mongodb.com/docs/manual/core/schema-validation/>
- Nagle, T., Redman, T. C., & Sammon, D. (2017, September 11). *Only 3% of Companies' Data Meets Basic Quality Standards*. Retrieved from hbr.org: <https://hbr.org/2017/09/only-3-of-companies-data-meets-basic-quality-standards>
- Palmer, M. (2022, August 3). *What is Google Cloud Composer?* Retrieved from zuar.com: <https://www.zuar.com/blog/what-is-google-cloud-composer/>
- Pointer, I. (2023, March 30). *What is Apache Spark? The big data platform that crushed Hadoop*. Retrieved from infoworld.com: <https://www.infoworld.com/article/3236869/what-is-apache-spark-the-big-data-platform-that-crushed-hadoop.html#:~:text=Show%20More-,Apache%20Spark%20defined,with%20other%20distributed%20computing%20tools>
- Ramanan, B., Drabeck, L., Woo, T., Cauble, T., & Rana, A. (2020). ~PB&J~ - Easy Automation of Data Science/Machine Learning Workflows. *2020 IEEE International Conference on Big Data (Big Data)* (pp. 361-371). Atlanta: IEEE.
- Rao, S. (n.d.). *6 Key Data Quality and Data Trustability Metrics You Should Automate on Azure and Snowflake*. Retrieved from Firsteigen.com: <https://firsteigen.com/blog/6-key-data-quality-metrics-you-should-be-tracking/>
- Shah, S., Amannejad, Y., & Krishnamurthy, D. (2021). Diaspore: Diagnosing Performance Interference in Apache Spark. *IEEE Access*.
- Sobotik, T. (2021, November 2). *How to ensure data quality with Great Expectations*. Retrieved from medium.com: <https://medium.com/snowflake/how-to-ensure-data-quality-with-great-expectations-271e3ca8b4b9>
- Spark, A. (n.d.). *Cluster Mode Overview*. Retrieved from Apache Spark Website: <https://spark.apache.org/docs/latest/cluster-overview.html>
- Stedman, C. (2022, December). *Data Quality*. Retrieved from techtarget.com: <https://www.techtarget.com/searchdatamanagement/definition/data-quality>
- Suleykin, A., Panfilov, P., & Bakhtadze, N. (2019). Industrial Track: Architecting Railway KPIs Data Processing with Big Data Technologies. *2019 IEEE International Conference on Big Data (Big Data)*. Los Angeles: IEEE.

- Tan, Q. (2018). Application of MongoDB Technology in NoSQL Database in Video. *Proceedings of the 8th International Conference on Management and Computer Science (ICMCS 2018)* (pp. 104-108). Shenyang: Atlantis Press.
- Taylor, D. (2023, July 15). *What is MongoDB? Introduction, Architecture, Features & Example*. Retrieved from guru99.com: <https://www.guru99.com/what-is-mongodb.html>
- Vaughan, J. (2020, November). *Google Bigquery*. Retrieved from techtarget.com: <https://www.techtarget.com/searchdatamanagement/definition/Google-BigQuery>
- Yu, Q., & Tang, Y. (2020). Cloud Accounting-based SME Investment Decission in the Big Data Era. *2020 5th International Conference on Mechanical, Control and Computer Engineering (ICMCCE)*. Harbin: IEEE.
- Zukowski, M. (2019). Cloud-based SQL Solutions for Big Data. In S. Sakr, & A. Y. Zomaya, *Encylopedia of Big Data Technologies* (pp. 1-7). Springer Cham.
- Zuo, X. (2023). Research on Data Quality Improvement Program Based on Big Data Application. *Research on Data Quality Improvement Program Based on Big Data Application*. Chongqing: IEEE.