REFERENCES

[1] A. Ismail, H. L. Truong, and W. Kastner, "Manufacturing process data analysis pipelines: a requirements analysis and survey," *Journal of Big Data*, vol. 6, no. 1, pp. 1–26, 2019. [Online]. Available: https://doi.org/10.1186/s40537-018-0162-3

[2] F. Pervaiz, A. Vashistha, and R. Anderson, "Examining the challenges in development data pipeline," *COMPASS 2019 - Proceedings of the 2019 Conference on Computing and Sustainable Societies*, pp. 13–21, 2019.

[3] C. Dehury, P. Jakovits, S. N. Srirama, V. Tountopoulos, and G. Giotis, *Data pipeline architecture for serverless platform*. Springer International Publishing, 2020, vol. 1269 CCIS. [Online]. Available: http://dx.doi.org/10.1007/978-3-030-59155-7_18

[4] N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "Data lifecycle challenges in production machine learning: A survey," *SIGMOD Record*, vol. 47, no. 2, pp. 17–28, 2018.

[5] A. Akoglu and G. Vargas-Solar, "Putting Data Science Pipelines on the Edge," pp. 1–13, 2021. [Online]. Available: http://arxiv.org/abs/2103.07978

[6] A. Raj, J. Bosch, H. H. Olsson, and T. J. Wang, "Modelling Data Pipelines," *Pro- ceedings - 46th Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2020*, pp. 13–20, 2020.

[7] M. Helu, T. Sprock, D. Hartenstine, R. Venketesh, and W. Sobel, "Scalable data pipeline architecture to support the industrial internet of things," *CIRP Annals*, vol. 69, no. 1, pp. 385–388, 2020.

[8] N. Konstantinou and N. W. Paton, "Feedback driven improvement of data preparation pipelines," *Information Systems*, vol. 92, 2020.

[9] E. G. Renart, D. Balouek-Thomert, and M. Parashar, "An edge-based framework for enabling data-driven pipelines for IoT systems," *Proceedings - 2019 IEEE 33rd International Parallel and Distributed Processing Symposium Workshops, IPDPSW 2019*, pp. 885–894, 2019.

[10] U. Zdun, E. Ntentos, K. Plakidas, A. El Malki, D. Schall, and F. Li, "On the design and architecture of deployment pipelines in cloud-and service-based computing-a model-based qualitative study," *Proceedings - 2019 IEEE International Conference on Services Computing, SCC 2019 - Part of the 2019 IEEE World Congress on Services*, pp. 141–145, 2019.

[11] E. G. Renart, D. Balouek-Thomert, and M. Parashar, “Edge Based Data-Driven Pipelines (Technical Report),” 2018. [Online]. Available: http://arxiv.org/abs/1808.01353

[12] H. Sebei, M. A. Hadj Taieb, and M. Ben Aouicha, “Review of social media analytics process and Big Data pipeline,” *Social Network Analysis and Mining*, vol. 8, no. 1, 2018. [Online]. Available: https://doi.org/10.1007/s13278-018-0507-0

[13] A. F. Conceição, J. V. Sánchez, B. G. Dos Santos, D. Vieira, and V. Rocha, “Pipeline architecture for mobile data analysis,” *International Conference on In- formation Networking*, no. February 2015, pp. 492–496, 2014.

[14] D. Weitzel, S. McKee, B. P. Bockelman, J. Thiltges, M. Babik, and I. Vukotic, “The Service Analysis and Network Diagnosis Data Pipeline,” *Proceedings of INDIS 2021: 8th Workshop on Innovating the Network for Data-Intensive Sci- ence, Held in conjunction with SC 2021: The International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–11, 2021.

[15] G. Vargas-Solar, M. S. Hassan, and A. Akoglu, “JITA4DS: Disaggregated Execu- tion of Data Science Pipelines Between the Edge and the Data Centre,” *Journal of Web Engineering*, vol. 21, no. 1, pp. 1–26, 2021.

[16] V. Zachariou, C. E. Bauer, D. K. Powell, and B. T. Gold, “Ironsmith: An automated pipeline for QSM-based data analyses,” *NeuroImage*, vol. 249, no. October 2021, p. 118835, 2022. [Online]. Available: https://doi.org/10.1016/j.neuroimage.2021.118835

[17] S. N. Mousavi, F. Chen, M. Abbasi, M. R. Khosravi, and M. Rafiee, “Efficient pipelined flow classification for intelligent data processing in IoT,” *Digital Com- munications and Networks*, vol. 8, no. 4, pp. 561–575, 2022.

[18] S. Gakhar, J. Cahoon, W. Le, X. Li, K. Ravichandran, H. Patel, M. Friedman, B. Haynes, S. Qiao, A. Jindal, and J. Leeka, “Pipemizer: An Optimizer for Ana- lytics Data Pipelines,” *Proceedings of the VLDB Endowment*, vol. 15, no. 12, pp. 3710–3713, 2022.

[19] Q. ting Lin, W. Yang, X. Zhang, Q. gang Li, Y. feng Liu, Q. Yan, and L. Sun, “Systematic and benchmarking studies of pipelines for mammal WGBS data in the novel NGS platform,” *BMC Bioinformatics*, vol. 24, no. 1, pp. 1–17, 2023. [Online]. Available: https://doi.org/10.1186/s12859-023-05163-w

[20] P. Anbukarasu, S. Nanisetty, G. Tata, and N. Ray, “Tiny-HR: Towards an interpretable machine learning pipeline for heart rate estimation on edge

devices," vol. 2022, no. August, pp. 1–10, 2022. [Online]. Available: http://arxiv.org/abs/2208.07981

[21] L. Bürgy, M. Weigert, G. Hatzopoulos, M. Minder, A. Journé, S. J. Rahi, and P. Gönczy, "CenFind: a deep-learning pipeline for efficient centriole detection in microscopy datasets," *BMC bioinformatics*, vol. 24, no. 1, p. 120, 2023.

[22] E. Solovieva and H. Sakai, "PSReliP: an integrated pipeline for analysis and visualization of population structure and relatedness based on genome-wide genetic variant data," *BMC bioinformatics*, vol. 24, no. 1, p. 135, 2023. [Online]. Available: https://doi.org/10.1186/s12859-023-05169-4

[23] A. Witayangkurn, A. Arai, and R. Shibasaki, "Development of Big Data-Analysis Pipeline for Mobile Phone Data with Mobipack and Spatial Enhancement," *IS-PRS International Journal of Geo-Information*, vol. 11, no. 3, 2022.

[24] M. Hamdani, Y. Aklouf, and H. A. Bouarara, "A Parallel Fuzzy Load Balancing Algorithm for distributed Nodes over a Cloud System," *International Journal of Grid and High Performance Computing*, vol. 14, no. 1, 2022.

[25] D. Farrell and J. E. Nielsen, "DataPipeline: Automated importing and fitting of large amounts of biophysical data," *Journal of Computational Chemistry*, vol. 33, no. 29, pp. 2357–2362, 2012.

[26] T. Um, B. Oh, B. Seo, M. Kweun, G. Kim, and W. Y. Lee, "Fast Flow: Ac- celerating Deep Learning Model Training with Smart Offloading of Input Data Pipeline," *Proceedings of the VLDB Endowment*, vol. 16, no. 5, pp. 1086–1099, 2023.

[27] P. Song, Y. Qie, C. Hao, Y. Li, Y. Zhao, Y. Hao, H. Liu, and Y. Qi, "Resource- Saving Customizable Pipeline Network Architecture for Multi-Signal Processing in Edge Devices," *Sensors*, vol. 22, no. 15, pp. 1–17, 2022.

[28] J. Yin, Y. Li, D. Robinson, and C. Yu, "RESPECT: Reinforcement Learning based Edge Scheduling on Pipelined Coral Edge TPUs," 2023. [Online]. Available: http://arxiv.org/abs/2304.04716

[29] I. Moutsopoulos, E. C. Williams, and I. I. Mohorianu, "bulkAnalyseR: an acces-sible, interactive pipeline for analysing and sharing bulk multi-modal sequencing data," *Briefings in bioinformatics*, vol. 24, no. 1, pp. 1–7, 2023.

[30] H. Wang, C. Imes, S. Kundu, P. A. Beerel, S. P. Crago, and J. P. Walters, "QuantPipe: Applying Adaptive Post-Training Quantization for Distributed Transformer Pipelines in Dynamic Edge Environments," 2022. [Online]. Available: http://arxiv.org/abs/2211.04515

[31] A. Q. Khan, N. Nikolov, M. Matskin, R. Prodan, D. Roman, B. Sahin, C. Bussler, and A. Soylu, “Smart Data Placement Using Storage-as-a-Service Model for Big Data Pipelines,” *Sensors*, vol. 23, no. 2, pp. 1–20, 2023.

[32] Y. Hu, C. Imes, X. Zhao, S. Kundu, P. A. Beerel, S. P. Crago, and J. P. N. Walters, “Pipeline Parallelism for Inference on Heterogeneous Edge Computing,” 2021. [Online]. Available: http://arxiv.org/abs/2110.14895

[33] A. Goel, C. Tung, X. Hu, G. K. Thiruvathukal, J. C. Davis, and Y. H. Lu, “Effi- cient Computer Vision on Edge Devices with Pipeline-Parallel Hierarchical Neu- ral Networks,” *Proceedings of the Asia and South Pacific Design Automation Conference, ASP-DAC*, vol. 2022-Janua, pp. 532–537, 2022.

[34] N. Skatchkovsky and O. Simeone, “Optimizing Pipelined Computation and Communication for Latency-Constrained Edge Learning,” *IEEE Communications Letters*, vol. 23, no. 9, pp. 1542–1546, 2019.

[35] S. Redyuk, S. Schelter, Z. Kaoudi, and V. Markl, “DORIAN in action: Assisted Design of Data Science Pipelines,” *Proceedings of the VLDB Endowment*, vol. 15, no. 12, pp. 3714–3717, 2022.

[36] L. Reimann and G. Kniesel-Wünsche, “An Alternative to Cells for Selective Execution of Data Science Pipelines,” 2023. [Online]. Available: http://arxiv.org/abs/2302.14556

[37] B. Sudharsan, J. G. Breslin, and M. I. Ali, “RCE-NN: A five-stage pipeline to execute neural networks (CNNs) on resource-constrained IoT edge devices,” *ACM International Conference Proceeding Series*, 2020.

[38] F. Moyón, R. Soares, M. Pinto-Albuquerque, D. Mendez, and K. Beckers, “In- tegration of Security Standards in DevOps Pipelines: An Industry Case Study,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Arti- ficial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12562 LNCS, pp. 434–452, 2020.

[39] F. Ullah, A. J. Raft, M. Shahin, M. Zahedi, and M. A. Babar, “Security support in continuous deployment pipeline,” *ENASE 2017 - Proceedings of the 12th International Conference on Evaluation of Novel Approaches to Software Engineering*, pp. 57–68, 2017.

[40] S. J. Qin, “Process data analytics in the era of big data,” *AIChE Journal*, vol. 60, no. 9, pp. 3092–3100, 2014. [Online]. Available: https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.14523

[41] N. Konstantinou, M. Koehler, E. Abel, C. Civili, B. Neumayr, E. Sallinger, A. A. Fernandes, G. Gottlob, J. A. Keane, L. Libkin, and N. W. Paton, “The vada architecture for cost-effective data wrangling,” in *Proceedings of the 2017 ACM International Conference on Management of Data*, ser. SIGMOD ’17. New York, NY, USA: Association for Computing Machinery, 2017, p. 1599–1602. [Online]. Available: https://doi.org/10.1145/3035918.3058730