Lam, C. (2011). *Hadoop in Action*. Manning Publications Co.: Stamford, CT. (Lam, 2011).

Hadoop in Action" is a comprehensive guide that introduces readers to Hadoop and its ecosystem. The book provides a solid foundation in understanding the basics of Hadoop, including its architecture, file system (HDFS), and core components like MapReduce. It demonstrates how Hadoop can be leveraged to process large datasets efficiently. The book includes case studies that showcase Hadoop in real-world applications:

Converting 11 million image documents from the New York Times archive.

Mining data at China Mobile.

Recommending the best websites at StumbleUpon.

Building analytics for enterprise search in IBM’s Project ES2.

Furthermore, the book covers advanced topics such as setting up a Hadoop cluster, data processing with Hive and Pig, and introduces other components of the Hadoop ecosystem like HBase, ZooKeeper, and Sqoop. This book equips readers with the necessary knowledge to get started with Hadoop and understand its significance in the realm of Big Data. This source helps to justify the importance of Hadoop in Big Data analytics and offers a thorough explanation of what a distributed file system is.

The Past, Present, and Future of Machine Learning APIs (Atakan Cetinsoy et al., 2016).

This paper discusses the evolution of ML APIs, tracing their journey from the era when researchers relied on pen and paper, to the present, where IT systems deploy these ML models. It also addresses current challenges, emphasizing the treatment of data before implementing an ML model. The authors highlight solutions such as Machine Learning as a Service (e.g., Azure ML, AWS) and REST APIs (e.g., BigML), showing how these technologies are valuable because they provide evidence of large companies creating and using APIs for modelling their ML models.

Deep Machine Learning and Neural Networks: An Overview (Mishra and Gupta, 2017).

According to Mishra and Gupta (2017), deep learning surpasses traditional ML models in its ability to perceive text and images. Neural Networks, as a crucial component of deep learning, are discussed in depth—specifically, ANNs and CNNs, which are key to this paper. It concludes that NNs are among the more popular techniques for solving deep learning problems.

Accelerating Relational Databases by Leveraging RemoteMemory and RDMA (Li et al., 2016).

Li et al. (2016) studied the crucial role of memory in RDBMS, especially comparing physical memories (SSD or HDD) with cloud memory (remote direct access, RDMA). This paper examines four scenarios to demonstrate that remote memory hosted in servers outperforms traditional physical memory systems. It is relevant to note that the state of the art of RDBMS has dramatically improved thanks to cloud-based solutions

**NoSQL: The Future of Big Data Analytics and Comparison with RDBMS (Arshad et al., 2023),**

This article compares NoSQL technologies with traditional RDBMS in the context of Big Data analytics. The paper describes NoSQL as "Not Only SQL" and categorizes these databases into key-value stores, document databases, wide-column stores, and graph databases, with Hadoop classified as a wide-column store. It outlines the evolution of Big Data from megabytes and gigabytes to terabytes and petabytes, constantly challenging the industry to develop new storage solutions to meet escalating demands. The nature of Big Data is also elucidated in terms of its volume, variety, velocity, and variability.

Furthermore, the paper details the ACID properties of RDBMS—atomicity, consistency, isolation, and durability—and compares them with the CAP theorem of NoSQL, which emphasizes strong consistency, high availability, and partition tolerance. The authors conducted a survey among relevant IT companies, revealing a preference for NoSQL technologies when managing Big Data. The study concludes that applications dealing with Big Data tend to perform better in NoSQL environments.

***How to mention the sources in the paper:***

In a literature review, it's standard to mention sources through in-text citations rather than writing down the full name of the paper each time you refer to it. The Harvard referencing style, for example, would have you include the author’s surname and the year of publication in parentheses within the text. This method allows your narrative to flow more smoothly while still giving credit to the original authors and enabling readers to find the detailed reference in your bibliography or reference list.

Here’s a brief example to illustrate:

Incorrect: "In the literature review, writing down the name of the paper each time for reference might disrupt the flow of reading."

Correct: "According to Smith (2020), integrating in-text citations within the narrative of a literature review enhances readability and allows for a smoother flow of information."

Remember, the exact format for your in-text citations and reference list entries will depend on the specific guidelines provided by your instructor or the preferred citation style of your academic discipline.

**Paper References:**

Alwosheel, A., van Cranenburgh, S. and Chorus, C.G., 2018. Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. Journal of Choice Modelling. https://doi.org/10.1016/j.jocm.2018.07.002

Arnold, B. and T., 2017. kerasR: R Interface to the Keras Deep Learning Library. *The Journal of Open Source Software*, 2(14), 296. <https://doi.org/10.21105/joss.00296>.

Arshad, M., Brohi, M.N., Soomro, T.R., Ghazal, T.M., Alzoubi, H.M. and Alshurideh, M., 2023. NoSQL: Future of Big Data Analytics Characteristics and Comparison with RDBMS. In: *The Effect of Information Technology on Business and Marketing Intelligence Systems*, pp.1927-1951. doi: 10.1007/978-3-031-12382-5\_106.

Barry-Straume, J., Tschannen, A., Engels, D.W. and Fine, E., 2018. An Evaluation of Training Size Impact on Validation Accuracy for Optimized Convolutional Neural Networks. *SMU Data Science Review*, 1(4), Article 12. https://scholar.smu.edu/datasciencereview/vol1/iss4/12

Bughin, J., Seong, J., Manyika, J., Chui, M., & Joshi, R. (2018). *Notes from the AI frontier: Applications and value of deep learning*. McKinsey Global Institute.

Candel, C.J.F., Sevilla Ruiz, D. and García-Molina, J.J., 2022. A unified metamodel for NoSQL and relational databases. *Information Systems, 104*, p.101898. <https://doi.org/10.1016/j.is.2021.101898>.

Cetinsoy, A., Martin, F.J., Ortega, J.A. and Petersen, P., 2016. The Past, Present, and Future of Machine Learning APIs. In*: PAPIs 2015 - Proceedings of the 2015 Conference on Predictive APIs and Apps*, vol. 50, JMLR: Workshop and Conference Proceedings, pp.43-49.‌

Chollet, F. (2018). *Deep Learning with Python*. Shelter Island, NY: Manning Publications. pp. 253-259.

Halevy, A., Norvig, P., & Pereira, F. (2009). *The Unreasonable Effectiveness of Data*. IEEE Intelligent Systems, 24(2), 8-12.

Holmes, A. (2012). *Hadoop in Practice*. Manning Publications Co.: Shelter Island, NY.

Khan, S.H., Hayat, M. and Porikli, F., 2018. Regularization of deep neural networks with spectral dropout. *Neural Networks.* https://doi.org/10.1016/j.neunet.2018.09.009

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification with deep convolutional neural networks.* Advances in Neural Information Processing Systems, 25.

Lam, C. (2011). *Hadoop in Action*. Manning Publications Co.: Stamford, CT.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep learning. Nature,* 521(7553), 436-444.

Li, F., Das, S., Syamala, M. and Narasayya, V.R., 2016. Accelerating Relational Databases by Leveraging Remote Memory and RDMA. In: *Proceedings of the 2016 International Conference on Management of Data - SIGMOD ’16.* https://doi.org/10.1145/2882903.2882949.

Mishra, C. and Gupta, D.L., 2017. Deep Machine Learning and Neural Networks: An Overview. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 6(2), pp.66-73. <https://doi.org/10.11591/ijai.v6.i2.pp66-73>.

Neeta Awasthy and Nikhila Valivarthi (2023). Evolution of Hadoop and Big Data Trends in Smart World. *Springer eBooks*, pp.99–127. https://doi.org/10.1007/978-3-031-13577-4\_6.

Palanisamy, S. and SuvithaVani, P., 2020. A survey on RDBMS and NoSQL Databases: MySQL vs MongoDB. In: *2020 International Conference on Computer Communication and Informatics (ICCCI -2020)*, 22-24 January 2020, Coimbatore, India. IEEE.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., & Dennison, D. (2015). *Hidden technical debt in Machine learning systems*. In Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2 (NIPS'15).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need.* Advances in Neural Information Processing Systems, 30.