This literature review is based on a topic paper that explores how technologies like RDBMS, HDFS, and APIs are shaping the constantly evolving field of deep learning, particularly in the area of Neural Networks. To develop this paper, 16 articles and 2 books were reviewed, contributing to its creation. The sources have been thematically classified into two categories:

Storage Solutions: Comparing HDFS and RDBMS.

Neural Networks and Machine Learning APIs.

1. Storage Solutions: Comparing HDFS and RDBMS.

Chuck Lam (2011) provides a solid foundation in the basics of Hadoop, including its architecture, file system (HDFS), and core components like MapReduce. He demonstrates how Hadoop can be leveraged to process large datasets efficiently also shows Hadoop in real worlds applications for large organisations such as the New Your Times, China Mobile, StubmbleUpon and IBM. This source helps justify the importance of Hadoop as a storage solution and its emergence by providing computational capabilities over large amounts of data. Hadoop then came into play to tackle the challenges posed by Big Data for large corporations, a focus of Neeta Awasthy and Nikhila Valivarthi (2023) study. The index building for Google Search involves massive datasets and performs statistical analysis in indexing through large-scale batch processing. Facebook manages two primary clusters that store approximately 12 terabytes of data each. Additionally, eBay maintains substantial clusters designed to manage the data of 180 billion active users. To this paper, the value of this study lies in how it demonstrates the real-world applications of HDFS. A key component for RDBMS is memory and its performance role it plays that is what Li et al., (2016) studied. They compared 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. Palanisamy and Suvitha Vani (2020) conducted a survey comparing the concepts of NoSQL and RDBMS, including their limitations, and also addressed the advantages and types of NoSQL databases. Their research concluded that a NoSQL database is an excellent choice when the data is structured, and the volume is not large, while NoSQL is preferable for unstructured data or for structured data with the potential for rapid growth. Essential to this paper is the way this article summarizes the points where NoSQL overtakes RDBMS. Arshad et al., (2023) compare 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. Candel, Sevilla Ruiz, and Garcia-Molina (2022) are determined to prove that a unified metamodel for NoSQL and relational databases represents the future in the field of databases, especially as NoSQL technologies have gained popularity recently. In their conclusion, the authors present a metamodel named U-Schema, where both NoSQL and RDBMS are optimized. Pertinent to this paper is an understanding of the current state of the art concerning RDBMS.

1. Machine Learning APIs, and Neural Networks.

Halevy, Norvig, and Pereira (2009) argue that large volumes of data can be more valuable than complex algorithms in developing artificial intelligence systems. They maintain that large-scale data can improve language processing and that simple algorithms can outperform complex ones when coupled with massive datasets. How this study fits into the paper is significant because it demonstrates an early stage of what is now known as an API; the authors used web-based data to train their models. The evolution of ML API (Atakan Cetinsoy et al., 2016), traces 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. A big breakthrough happened when Krizhevsky, Sutskever, and Hinton (2012) explain the development of a deep convolutional neural network (CNN) that outperforms prior models in classifying high-resolution images in the ImageNet challenge. The CNN features 60 million parameters and 650,000 neurons across five convolutional and three fully connected layers. This study is one of the first to introduce the term 'dropout' as a method to prevent overfitting. Continuing with the development of deep learning LeCun, Bengio, and Hinton (2015), emphasizing the ability of deep learning to create computational models that recognize complex patterns through multiple processing layers. The authors describe how deep learning has improved areas like speech recognition, visual recognition, object detection, drug discovery, and genomics. All these advancements are thanks to the use of backpropagation algorithms. It also mentions that deep convolutional networks are great for image, video, audio processing, text processing, and speech processing. A new neural network architecture, known as the *Transformer,* represents an evolution from recurrent and convolutional neural networks and was introduced by Vaswani et al. (2017). The *Transformer* employs attention mechanisms, which enable better parallelization, reduce training times, and enhance performance on machine learning translation tasks. This paper concludes that the Transformer is a superior model to its predecessors in terms of translation models. 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. Chollet (2018) provides great examples of best practices for deep learning, particularly the use of Keras as an API for modelling CNNs. Chollet demonstrates how Keras streamlines data processing in a CNN structure, from processing IMDB reviews to constructing network layers. This example was instrumental to this paper, showcasing Keras robustness and ease of use in the context of neural networks. Lee and Song (2019) focus on examining parameter estimation procedures for deep neural networks, as well as the structures of CNN models, ranging from basic to advanced techniques. The authors are also determined to demonstrate the critical steps in CNNs that enhance image classification performance on the CIFAR-10 dataset using Keras. Their conclusions suggest that utilizing multiple stacks of convolutional layers along with batch normalization can lead to improved predictions. Another important finding that is relevant to this paper is that Keras, due to its popularity as a neural network API, enables individuals to quickly familiarize themselves with deep learning methodologies. Jin, Chollet, Song, and Hu (2023) present AutoKeras, an Automated Machine Learning (AutoML) library designed to simplify the application of deep learning. It offers a solution for challenges in model selection and hyperparameter tuning, thereby making deep learning accessible to those with limited technical expertise. The library provides a user-friendly interface, assisting inexperienced users in addressing machine learning problems with minimal coding required. This study represents a significant step towards democratizing deep learning technology and aligns with the promising future of ML APIs discussed in this paper. Barry-Straume, Tschannen, Engels, and Fine (2018) assess how varying training set sizes influence the validation accuracy of CNNs. The study determines the optimal data volume required to achieve maximum accuracy during model validation. Furthermore, the research indicates that larger datasets can significantly enhance the predictive capabilities of CNNs. This study is crucial to the present paper as it introduces the concept of validation. In the implementation phase, validation was conducted to assess overfitting. Initially, the neural network and the specific CNN in focus here were prone to overfitting; however, through validation, this was effectively mitigated. Alwosheel, van Cranenburgh, and Chorus (2018) examine the appropriate sample size for ANNs in discrete choice modelling. They fill a gap in empirical guidelines by establishing a rule-of-thumb based on Monte Carlo analyses of both synthetic and real data. Their research suggests that a dataset size fifty times the number of weights in the ANN is more effective than the traditionally used benchmark of a dataset only ten times the weight count. This research illuminates a section of this paper, where, in phase 2, an overfitted NN was corrected by increasing the dataset size, thus resolving the overfitting issue. Khan, Hayat, and Porikli (2019) introduce 'Spectral Dropout,' an enhancement of the dropout technique newly introduced for the ImageNet challenge in 2012 by Krizhevsky, Sutskever, and Hinton (2012). Spectral Dropout augments traditional CNNs with a decorrelation transform and tackles overfitting by mitigating weak and noisy Fourier domain coefficients of network activations. The research concludes that this method's efficacy surpasses current regularization methods and increases network training speed. This paper is crucial in addressing overfitting in CNN phase 1, highlighting that since 2012, researchers have been using and improving dropout, leading to its implementation in CNN phase 2.

In conclusion, this literature review has identified two significant gaps:

* The application of HDFS and RDBMS in NN contexts.
* The needs of non-technical users who must model NNs using datasets large enough to overwhelm conventional applications. All literature reviewed was directed towards technical users proficient in advanced programming techniques for modelling NNs.

This paper aims to address the gaps mentioned above.