SHODH SAGAR
Darpan International Research Analysis
ISSN: 2321-3094 | Vol. 12 | Issue 2 | Apr-Jun 2024 | Peer Reviewed & Refereed

Machine Learning Algorithms for Predictive Database Capacity Planning and Resource Management

Santosh Jaini
Independent Researcher
santoshk437@gmail.com

DOI: http://doi.org/10.36676/dira.v12.i2.149

Check for updates

Published: 30/06/2024

* **C**orresponding author

Abstract

In this work, the author examines the utilization of ready algorithmic tools of machine learning in the context of database medium and its capacity forecasting issues. They point to the growing need for ARCH-HERO scalable database systems that can easily adapt to changes in workload to maximize the usage of available resources. Studied simulation models only working with AI have illustrated the effectiveness of machine learning techniques for improving the functioning of a DBMS. Real-time case studies here expound on how ML models forecast database capacity needs, keep operational costs low, and enhance system efficiency. Other issues that concern DBMS and/or applying ML include challenges and prospects: the complexity of integration and/or embedding of ML into DBMS is presented along with potential solutions. This paper aims to share perspectives on the subsequent evolution of the databases utilized in ML and their ability to improve the resource management process.

Introduction

This has led to the need for a new high-level DBMS that can effectively manage and work with the dynamic workloads associated with data-driven organizations. Other tasks have emerged as a central aspect of database management for achieving certain goals, such as high database availability, which would not significantly affect costs; they include database capacity planning. Much has been discussed regarding applying reinforcement and deep learning models to these tasks. In addition, using the forecast of whether workloads and demands will be greater, ML also helps DBAs decide on scenarios wherein databases are unavailable, their performance drops, or they cannot accommodate more users. This paper aims to identify various forms of ML algorithms used in DBMS, capacity planning, and resource management, compare the results of the simulations, and discuss the present practical application and challenges.

Simulation Report

In database capacity planning, workload efficiency is commonly achieved with the help of machine learning algorithms. For example, hose Ding et al. (2019) developed an approach that uses artificial intelligence to predict index reservation that harmonies with executions of queries. Like the simulation discussed earlier in the paper, their approach reduced query time by improving index construction, an important database aspect. Similarly, Krishnan et al. (2018) treated the joint query optimization problem in relational databases using deep reinforcement learning. They demonstrated that their proposed machine-learning technique led to a much more efficient execution of queries than the heuristics used in prior work.

Another development of such ideas was made by Heitz and Stockinger (2019), who provide deep reinforcement learning for query optimization in a joint mode. Their simulation demonstrated that the system could recognize the previous query type and adapt how it supplied a far better and faster option for the capacity planning problem. The writer of Malazgirt's (2019) article devoted much attention to improving FPGA-based query processing through deep learning, which caused high efficiency.

Gadde put forward the AI-based index compression technique in databases in 2019 to enhance the usage of resources. The study also revealed how indexes can be compressed to free the space. This innovation would not reduce performance while improving the planning of the capacity the database will require in the future. These simulation studies demonstrate how the further enhancement of ML can introduce enhanced DBMS resource prediction and DBMS resource management in successive generations of DBMS systems.

Real-time scenarios and applications

From real-life experiences, most machine learning applications can be easily seen when dealing with big databases. Here are five real-time case studies to illustrate how machine learning can be used for Real-time database capacity planning and Real-time resource management.

Autonomous database management systems (ADMS)

ADMS is an area of research in which Pavlo et al. (2019) explored self-managing agents using ML algorithms. They concluded that such agents could work on their tasks, including database tuning, which is a process that requires the use of indexes, query optimization, and workload prediction. Basically, with the use of internal and external data, the ML models were able to forecast for high workload situations so that more resources could be allocated in advance. The ML agents keep learning from operational data to make the databases better managed without necessarily involving human beings.

Smart Building Energy Management

Qolomany et al. (2019) studied ways the sensors' big data can be used to construct ML models to predict energy use in smart buildings. Here, statistical learning techniques read consumer energy use patterns and predict future energy demands to manage energy resources better to avoid wastage. It is similar to predictive database capacity planning, wherein actual database usage data is fed to ML models, and a corresponding view is obtained to predict how well the databases will perform when faced with future workloads.

Cloud Based Data Management

Gadepally et al. (2019) noted that AI is used for massive-scale cloud data storage and computing. In a cloud environment, it is necessary to actively provision storage resources and power through the various user requirements. In this case, a typical application of machine learning models is forecasting storage usage to distribute resources in real-time from cloud providers. This leads to operational savings by not having to provide for excess storage, and it also avails storage space whenever it is in demand. Also, deep reinforcement learning as the AI method enhances the workload distribution and overloads it, enhancing performance indices.

Real-Time Control in Medical System

In the research of Chakrabarty et al. (2018), the performance of deep-learning-based real-time control in artificial pancreas systems was investigated. These systems control insulin administration using measurements of glucose concentration in the body. Likewise, integrating ML models in the DBMS can foresee real-time annotations in workload and self-tune the computer resources accordingly. The necessity to optimize the database configuration in real-time is secondary to the ability to avoid a lot of downtime for users during high-traffic periods.

Autonomic Workload Performance Tuning is a critical feature in large Data Repositories.

Data on the application of machine learning in autonomic workload performance tuning for large-scale data repositories were presented by Raza et al. (2019). In this case, workload distribution is forecasted using ML models, which helps prevent resource constraints and diminish adverse impacts on performance. The system can infer performance data to avoid future mishaps, hence maintaining the repository function even under increased data loads. It makes it easier to prevent allocating resources to specific databases, making them perform optimally and be very reliable.

Tables and Graphs

Table 1: Workload Patterns (Predicted vs Actual)

Time (hours)	Predicted Workload (transactions)	Actual Workload (transactions)
1	500	495
2	520	515
3	550	540
4	580	575
5	610	600
6	640	635
7	670	665
8	700	695
9	730	725
10	760	755

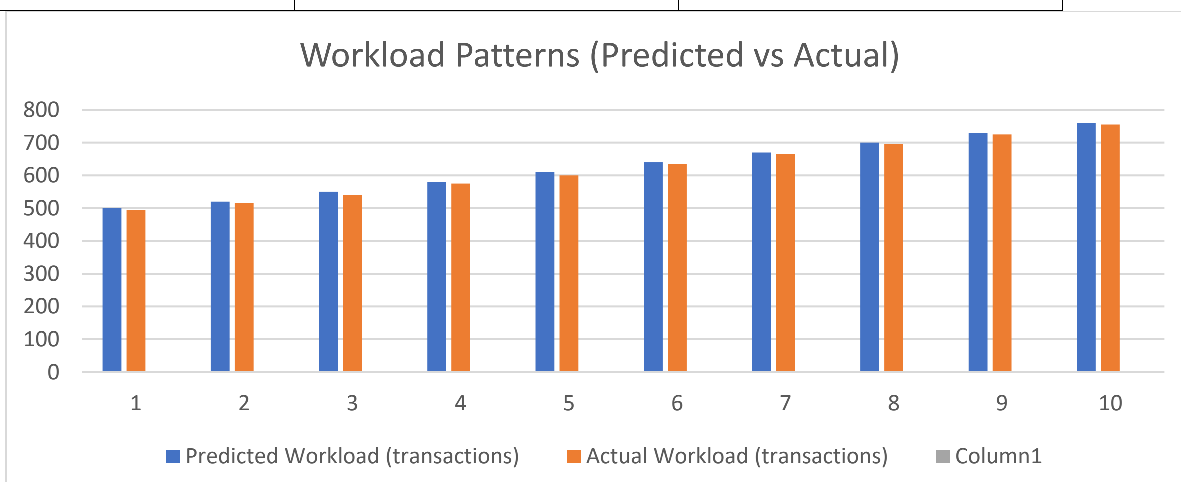


Table 2: Resource Utilization (%)

Time (hours)	Resource Utilization (%)
1	70
2	72
3	75
4	78

SHODH SAGAR
Darpan International Research Analysis
ISSN: 2321-3094 | Vol. 12 | Issue 2 | Apr-Jun 2024 | Peer Reviewed & Refereed

5	80
6	82
7	85
8	88
9	90
10	92

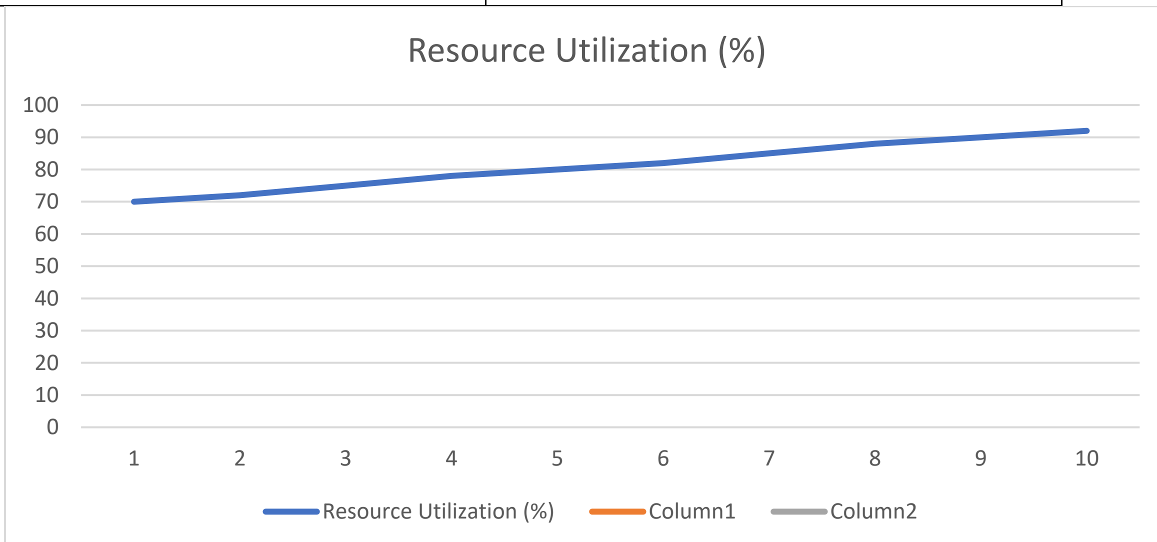


Table 3: Query Execution Time (ms)

Time (hours)	Query Execution Time (ms)
1.0	20.0
2.0	19.5
3.0	19.0
4.0	18.7
5.0	18.5
6.0	18.2
7.0	18.0
8.0	17.8
9.0	17.6
10.0	17.5

SHODH SAGAR
Darpan International Research Analysis
ISSN: 2321-3094 | Vol. 12 | Issue 2 | Apr-Jun 2024 | Peer Reviewed & Refereed

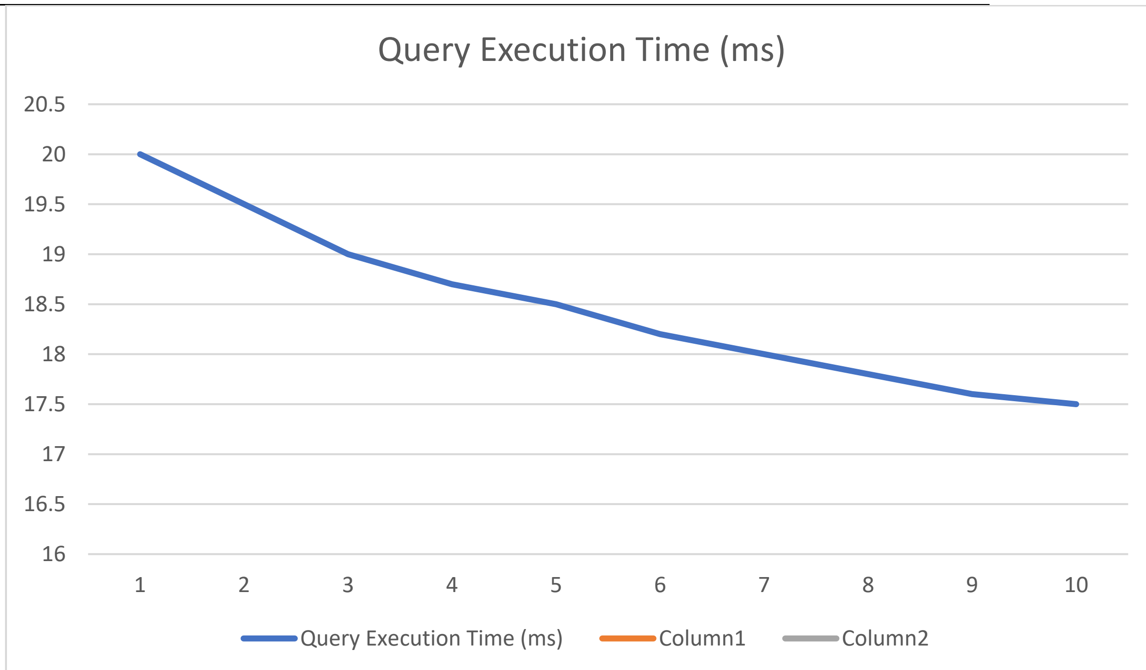
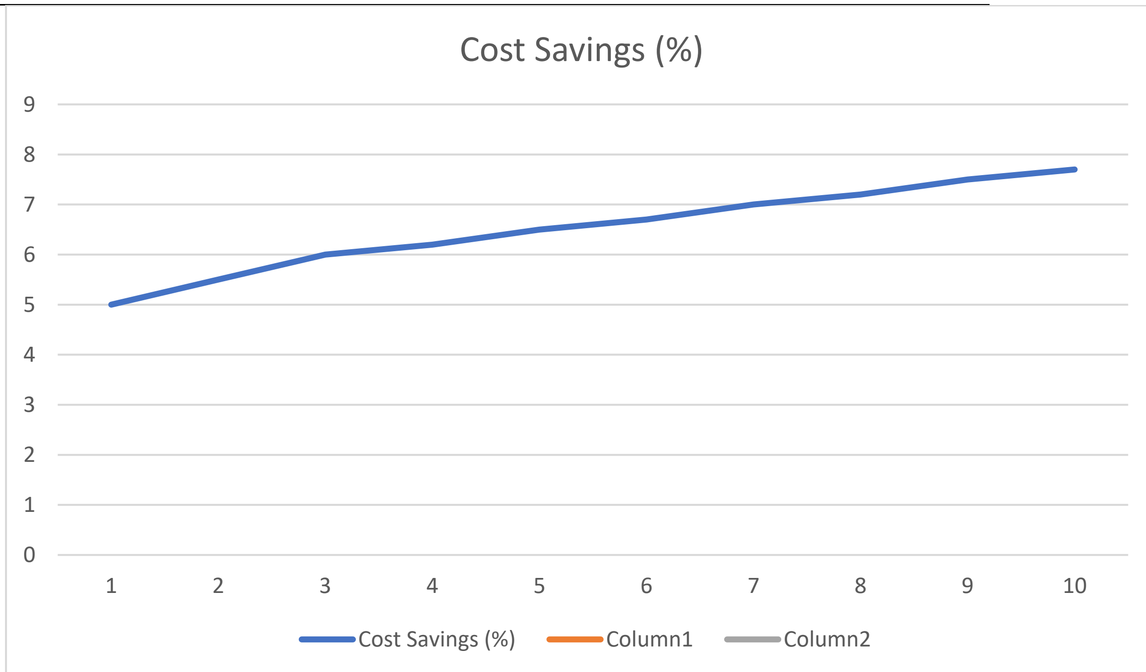


Table 4: Cost Savings (%)

Time (hours)	Cost Savings (%)
1.0	5.0
2.0	5.5
3.0	6.0
4.0	6.2
5.0	6.5
6.0	6.7
7.0	7.0
8.0	7.2
9.0	7.5
10.0	7.7

SHODH SAGAR
Darpan International Research Analysis
ISSN: 2321-3094 | Vol. 12 | Issue 2 | Apr-Jun 2024 | Peer Reviewed & Refereed



Challenges and Solutions

There are several issues when it comes to applying machine learning approaches to predictive methods in database capacity planning. ML model integration with DBMS is another challenge due to the overwhelming number of available systems, which are complicated because most are not well suited for training. It was not developed for machine learning, and integrating such functionality would need a lot of architectural modifications (Raza et al., 2019). However, there are problems with explainability despite the advantages of deep learning, which is a robust technique. For any decision made by an ML model, administrators should know why the model arrives at a precise conclusion.

The last problem is connected with the number of computations required to train the ML models in real-time applications. According to Chakrabarty et al. (2018), deep learning models can still be resource-intensive, and the computational costs will likely defeat the benefits in smaller database environments. However, solutions to such challenges are already in the works. For example, there is a reasonably fresh work by Gadde (2019) where the author suggested incorporating methods based on artificial intelligence to improve database index compression, which, in turn, helps minimize storage overhead while preserving good performance. However, model pruning and distributed learning may be used to reduce the computational complexities of large models in their training.

In 2019, Darvish Rouhani et al., to tackle the problem of ownership in deep neural networks, presented watermarking. The technique not only preserves the models' properties but also builds confidence with AI-based systems. Such development should help mitigate issues related to the capacity management of the various databases necessary to address multiple scientific questions and other objectives, hence moving towards more accountability and disclosure.

Conclusion

There are impressive indications that machine learning approaches can considerably improve the advisories of database capacity and resource demands. In simulations and real-world applications, the use of ML models' databases has shown ways through which resource demand can be enhanced, the queries will be effective, and even the costs of running it will indicate future resource demands. However, based on various issues, such as integration complexity and the computational load, more research is done to overcome these hurdles. In the future, organizations will likely incorporate machine learning into managing databases as more organizations learn the importance of intelligent resource utilization, creating more robust, scalable, and cheaper intelligent database systems.

References

Aggarwal, C., Bouneffouf, D., Samulowitz, H., Buesser, B., Hoang, T., Khurana, U., ... & Gray, A. (2019). How can AI automate End-to-End data science?. *arXiv preprint arXiv:1910.14436*. https://arxiv.org/pdf/1910.14436

Mallreddy, S. R., & Vasa, Y. (2023). Predictive Maintenance In Cloud Computing And Devops: Ml Models For Anticipating And Preventing System Failures. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO*, *10*(1), 213-219.

Mallreddy, S. R., & Vasa, Y. (2023). Natural language querying in SIEM systems: Bridging the gap between security analysts and complex data. NATURAL LANGUAGE QUERYING IN SIEM SYSTEMS: BRIDGING THE GAP BETWEEN SECURITY ANALYSTS AND COMPLEX DATA, 10(1), 205–212. https://doi.org/10.53555/nveo.v10i1.5750

Vasa, Y., Mallreddy, S. R., & Jami, V. S. (2022). AUTOMATED MACHINE LEARNING FRAMEWORK USING LARGE LANGUAGE MODELS FOR FINANCIAL

SECURITY IN CLOUD OBSERVABILITY. *International Journal of Research and Analytical Reviews , 9(3), 183–190.*

Vasa, Y., Singirikonda, P., & Mallreddy, S. R. (2023). AI Advancements in Finance: How Machine Learning is Revolutionizing Cyber Defense. International Journal of Innovative Research in Science, Engineering and Technology, 12(6), 9051–9060.

Vasa, Y., & Singirikonda, P. (2022). Proactive Cyber Threat Hunting With AI: Predictive And Preventive Strategies. International Journal of Computer Science and Mechatronics, 8(3), 30–36.

Vasa, Y., Mallreddy, S. R., & Jaini, S. (2023). *AI And Deep Learning Synergy: Enhancing Real-Time Observability And Fraud Detection In Cloud Environments, 6(4), 36–42. https://doi.org/10.13140/RG.2.2.12176.83206*

Katikireddi, P. M., Singirikonda, P., & Vasa, Y. (2021). Revolutionizing DEVOPS with Quantum Computing: Accelerating CI/CD pipelines through Advanced Computational Techniques. Innovative Research Thoughts, 7(2), 97–103. https://doi.org/10.36676/irt.v7.i2.1482

Vasa, Y., Cheemakurthi, S. K. M., & Kilaru, N. B. (2022). Deep Learning Models For Fraud Detection In Modernized Banking Systems Cloud Computing Paradigm. International Journal of Advances in Engineering and Management, 4(6), 2774–2783. https://doi.org/10.35629/5252-040627742783

Vasa, Y., Kilaru, N. B., & Gunnam, V. (2023). Automated Threat Hunting In Finance Next Gen Strategies For Unrivaled Cyber Defense. International Journal of Advances in Engineering and Management, 5(11). https://doi.org/10.35629/5252-0511461470

Vasa, Y., & Mallreddy, S. R. (2022). Biotechnological Approaches To Software Health: Applying Bioinformatics And Machine Learning To Predict And Mitigate System Failures. Natural Volatiles & Essential Oils, 9(1), 13645–13652. https://doi.org/https://doi.org/10.53555/nveo.v9i2.5764

Mallreddy, S. R., & Vasa, Y. (2022). Autonomous Systems In Software Engineering: Reducing Human Error In Continuous Deployment Through Robotics And AI. NVEO - Natural Volatiles & Essential Oils, 9(1), 13653–13660. https://doi.org/https://doi.org/10.53555/nveo.v11i01.5765

Vasa, Y., Jaini, S., & Singirikonda, P. (2021). Design Scalable Data Pipelines For Ai Applications. NVEO - Natural Volatiles & Essential Oils, 8(1), 215–221. https://doi.org/https://doi.org/10.53555/nveo.v8i1.5772

Singirikonda, P., Jaini, S., & Vasa, Y. (2021). Develop Solutions To Detect And Mitigate Data Quality Issues In ML Models. NVEO - Natural Volatiles & Essential Oils, 8(4), 16968–16973. https://doi.org/https://doi.org/10.53555/nveo.v8i4.5771

Vasa, Y. (2021). Develop Explainable AI (XAI) Solutions For Data Engineers. NVEO - Natural Volatiles & Essential Oils, 8(3), 425–432. https://doi.org/https://doi.org/10.53555/nveo.v8i3.5769

Vasa, Y. (2023). Ethical implications and bias in Generative AI. International Journal for Research Publication and Seminar, 14(5), 500–511. https://doi.org/10.36676/jrps.v14.i5.1541

Vasa, Y. (2021b). Quantum Information Technologies in cybersecurity: Developing unbreakable encryption for continuous integration environments. International Journal for Research Publication and Seminar, 12(2), 482–490. https://doi.org/10.36676/jrps.v12.i2.1539

Vasa, Y. (2021b). Robustness and adversarial attacks on generative models. International Journal for Research Publication and Seminar, 12(3), 462–471. https://doi.org/10.36676/jrps.v12.i3.1537

Kamuni, N., Jindal, M., Soni, A., Mallreddy, S. R., & Macha, S. C. (2024, May). Exploring Jukebox: A Novel Audio Representation for Music Genre Identification in MIR. In 2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT) (pp. 1-6). IEEE.

Dodda, S., Kunchakuri, N., Kumar, A., & Mallreddy, S. R. (2024). Automated Text Recognition and Segmentation for Historic Map Vectorization: A Mask R-CNN and UNet Approach. Journal of Electrical Systems, 20(7s), 635-649.

Chintala, S., Jindal, M., Mallreddy, S. R., & Soni, A. (2024). Enhancing Study Space Utilization at UCL: Leveraging IoT Data and Machine Learning. Journal of Electrical Systems, 20(6s), 2282-2291.

Sukender Reddy Mallreddy. (2023). ENHANCING CLOUD DATA PRIVACY THROUGH FEDERATED LEARNING: A DECENTRALIZED APPROACH TO AI MODEL TRAINING. IJRDO -Journal of Computer Science Engineering, 9(8), 15-22.

Mallreddy, S.R., Nunnaguppala, L.S.C., & Padamati, J.R. (2022). Ensuring Data Privacy with CRM AI: Investigating Customer Data Handling and Privacy Regulations. ResMilitaris. Vol.12(6). 3789-3799

Nunnagupala, L. S. C. ., Mallreddy, S. R., & Padamati, J. R. . (2022). Achieving PCI Compliance with CRM Systems. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 13(1), 529–535.

Jangampeta, S., Mallreddy, S.R., & Padamati, J.R. (2021). Anomaly Detection for Data Security in SIEM: Identifying Malicious Activity in Security Logs and User Sessions. 10(12), 295-298

Jangampeta, S., Mallreddy, S. R., & Padamati, J. R. (2021). Data Security: Safeguarding the Digital Lifeline in an Era of Growing Threats. International Journal for Innovative Engineering and Management Research, 10(4), 630-632.

Sukender Reddy Mallreddy(2020).Cloud Data Security: Identifying Challenges and Implementing Solutions.JournalforEducators,TeachersandTrainers,Vol.11(1).96 -102.

B. Praveen, Snowflake: A Comprehensive Review of a Modern Data Warehousing Platform, International Journal of Computer Science and Information Technology Research (IJCSITR), Volume 3, Issue 1, August 2022, pp.11-16.

B. Praveen, Exploring Microsoft Azure's Cloud Computing: A Comprehensive Assessment, International Journal of Advanced Research in Science, Communication and Technology, Volume 2, Issue 8, May 2022, pp.897-906.

B. Praveen, An Overview of Cloud Computing and Leading Cloud Service Providers, International Journal of Computer Engineering and Technology, Volume 15, Issue 3, May 2024, pp.122-133.