

Performance Comparison of MicroSD and eMMC Storage in a Single-Node Hadoop Environment

Muhammad Arfah Asis^{1✉}, Lutfi Budi Ilmawan², Nur Ikhwan Alfiansyah³,
Rahman Ramadhan⁴

^{1, 2, 3, 4} Informatics Engineering, Faculty of Computer Science, Universitas Muslim Indonesia, Indonesia

✉ **Corresponding Author** : Muhammad Arfah Asis (e-mail: muhammad.arfah.asis@umi.ac.id)

Article Information	ABSTRACT
Article History Received : October 6, 2024 Revised : November 2, 2024 Accepted : January 2, 2025	This study analyzes the performance comparison between eMMC and MicroSD storage in a single-node Hadoop environment, focusing on data processing efficiency using the Terasort and TestDFSIO benchmarks. In this experiment, four different data sizes, namely 500MB, 1GB, 1.5GB, and 2GB, were tested to evaluate how well each storage type handles data processing. The test results show that eMMC consistently outperforms MicroSD across all tested dataset sizes. The larger the data size processed, the more significant the performance comparison between the two storage types. At a data size of 2GB, eMMC is almost four times faster than MicroSD, showing a very clear advantage in processing efficiency. In addition, the results of the TestDFSIO test support this finding. In the test, eMMC shows a write speed that is 50% higher than MicroSD, and a read speed that is almost twice as fast at a data size of 10GB. This performance difference confirms that eMMC has a better capacity to handle large data, which is an important factor in applications that require intensive processing. The findings emphasize that eMMC offers better performance and stability than MicroSD, making it a more suitable choice for applications requiring high speed and efficiency in Hadoop environments. This research is expected to provide valuable insights for developers and researchers considering optimal storage solutions for big data processing.
Keywords: <i>Single Board Computer; Hadoop; Storage Performance; MicroSD; eMMC</i>	

INTRODUCTION

In the ever-growing digital era, data has become one of the most valuable assets in decision-making and business strategies in various sectors, including business, health, education, and government. Companies and organizations increasingly rely on data analysis to gain insights that can improve operational efficiency, product development, and customer service. As the volume and complexity of data (big data) increases, the need for efficient, reliable, and high-performance storage solutions is increasingly pressing (Adnan et al., 2019).

Big data has become a major focus in the field of information and communication technology. Big data refers to a very large, diverse, and rapidly changing volume of data (Oliveira & Bollen, 2023), which requires new techniques and technologies for collection, storage, processing, and analysis (Susanto, 2024). Hadoop, an open-source framework designed for large-scale data storage and processing, has become a leading solution in handling Big data (Ma et al., 2023). Hadoop allows the distribution of data and processing tasks across multiple computers, which significantly increases the efficiency and speed of data processing (Elkawagy & Elbeh, 2020).

One of the technologies popularly used by researchers and developers in the context of using Hadoop is the Single Board Computer (SBC), which is known for its small size, low power consumption, and affordable price (Adnan et al., 2020; Qureshi, 2024). SBC have been implemented in various applications, such as mini servers (Godinho et al., 2023; Mantik & Jaya, 2021), Internet of Things (IoT) systems (Godinho et al., 2023; Ramadhani et al., 2020), computing clusters (Abiola et al., 2024; Hoffmann et al., 2024), and educational projects (Ariza & Baez, 2021; Kuss et al., 2018; Patel & Hamirani, 2021). Although SBC have great potential, their ability to handle more demanding data storage tasks, especially in Big Data scenarios, still needs to be further explored (Lee et al., 2021).

Data storage on SBC usually uses two types of media, namely Micro Secure Digital (MicroSD) Card and Embedded Multi Media Card (eMMC). MicroSD is a more common choice due to its wide availability, ease of use, and low cost. However, MicroSD often have limitations in terms of read/write speed and endurance, especially when used for intensive data storage tasks. On the other hand, eMMC offers higher read/write speeds and better reliability than MicroSD, but at a higher cost (Thomas-Brans et al., 2024). Choosing the right storage media is critical to ensure optimal performance of SBC in data storage applications.

The use of eMMC and MicroSD for big data storage on SBCs has been observed across various sectors. For example, MicroSD cards are commonly used in IoT systems for data logging (Dwi Agustin, 2022; Heru Sandi & Fatma, 2023; Sutikno et al., 2021), where high-speed access is not crucial, but large volumes of data are continuously recorded, such as in environmental monitoring stations. However, for applications requiring faster processing and larger datasets, like edge computing in surveillance systems or localized analytics, eMMC storage is preferred due to its higher speed and durability, making it more suitable for handling substantial data workloads over extended periods (Godinho et al., 2023). This choice enables SBCs to manage big data tasks more effectively in resource-constrained environments, highlighting the importance of selecting storage that aligns with the demands of data-intensive applications on SBCs.

Combining Hadoop capabilities with SBC offers great potential to create cost-effective and efficient data storage systems (Adnan et al., 2019). Small and medium-sized organizations, which often have limited resources, can leverage this combination to meet their data storage needs without having to spend a lot on more expensive hardware. However, the use of SBC with MicroSD and eMMC storage media in a Hadoop environment raises several important aspects that need to be studied. This study needs to reveal the data storage performance of SBC when using Hadoop, identify the factors that affect the performance, and compare the performance between MicroSD and eMMC in intensive data storage tasks. By conducting benchmarking using Hadoop, we can systematically measure the data storage performance of SBC and obtain important observation data. This benchmarking involves testing the performance of various storage configurations to determine the read/write speed, latency, and endurance of MicroSD and eMMC when used on SBC.

This study aims to analyze the data storage performance of SBC using SD Cards and eMMC in a Hadoop environment. This study will provide new insights into how SBC can be optimized for more demanding data storage tasks, as well as offer practical solutions for organizations looking for efficient and cost-effective data storage alternatives. Thus, this research is expected to contribute significantly in the field of information and communication technology, as well as encourage further innovation in the use of more affordable and efficient computing technology.

RESEARCH METHODS

This study employs an experimental approach to compare the performance of two types of storage, MicroSD and eMMC, in a single-node Hadoop environment. Two primary tests were conducted: Terasort and TestDFSIO, both of which are part of Hadoop benchmarking, where Hadoop implements the MapReduce methodology (Kim et al., 2018; Qureshi, 2024). MapReduce is a programming model that enables large-scale data processing by dividing tasks into two main

stages: Map, which maps input data into key-value pairs, and reduce, which combines the results of the Map phase to produce the final output (Jayshree, 2022; Jing et al., 2017; Tahsir Ahmed Munna et al., 2018).

In the Terasort test, each storage was evaluated for processing data sizes of 500MB, 1GB, 1.5GB, and 2GB. To ensure result accuracy, each data size was tested 10 times. This test aims to measure the time taken by each storage medium to sort data in a Hadoop environment, which involves intensive read and write operations. In the TestDFSIO test, measurements were conducted 10 times, consisting of 5 write tests and 5 read tests. This test was performed using a 1GB file size per file with a total of 10 files, equivalent to 10GB of data. The focus of this test is to evaluate the read and write performance of each storage medium when handling large files in parallel within a Hadoop environment. The results of both tests were analyzed to compare the performance of MicroSD and eMMC storage, aiming to determine which storage medium is more efficient in handling data processing tasks in a single-node Hadoop environment.

This study utilized the Orange Pi 5 Pro with 8GB of RAM, a Single Board Computer (SBC) known for its efficiency and flexibility in small-scale computing applications. The Orange Pi 5 Pro is equipped with an 8-core 64-bit processor with a clock speed of up to 2.4GHz, which is powerful enough to handle Hadoop operations on a single node, while also supporting a variety of storage media. The storage media used in this study were the Sandisk Extreme Pro 64GB MicroSD and Orange Pi 64GB eMMC.

RESULTS AND DISCUSSION

In the Terasort test, experiments were conducted on four different data sizes: 500MB, 1GB, 1.5GB, and 2GB. Each data size was tested 10 times to obtain an average result. The time required to complete each test is presented in a diagram that illustrates the performance comparison between MicroSD and eMMC. The results of the testing on the Orange Pi with MicroSD storage are shown in Table 1.

Table 1. Terasort Test Results Using MicroSD

Experiment No.	Execution Time (second)			
	500 MB	1 GB	1.5 GB	2 GB
1	58	132	331	506
2	58	118	344	515
3	57	120	294	499
4	53	142	293	512
5	55	174	253	494
6	55	187	334	471
7	52	223	304	556
8	54	211	341	549
9	52	212	278	548
10	60	217	346	545

Table 1 presents the execution times (in seconds) for processing different data sizes using MicroSD storage. For the 500MB dataset, the execution times were consistent, ranging from 52 to 60 seconds, indicating reliable performance when handling smaller data sizes. As the data size increased to 1GB, however, the execution times exhibited more variation, ranging between 118 and 223 seconds. This suggests that larger datasets place greater demands on the system, resulting in more pronounced fluctuations in performance.

Similarly, for the 1.5GB dataset, execution times ranged from 253 to 346 seconds, further emphasizing the trend of increasing variability and processing duration as the data size grew. The 2GB dataset produced the longest execution times, spanning from 471 to 556 seconds, which represents a substantial rise in processing time compared to the smaller datasets. This significant increase underscores the challenges and resource constraints associated with managing larger

datasets in a single-node Hadoop environment. Overall, the results reveal a clear relationship between data size and execution time, with both processing time and performance variability rising markedly for datasets larger than 1GB. If visualized in a graph, these results can be seen more clearly in Figure 1.

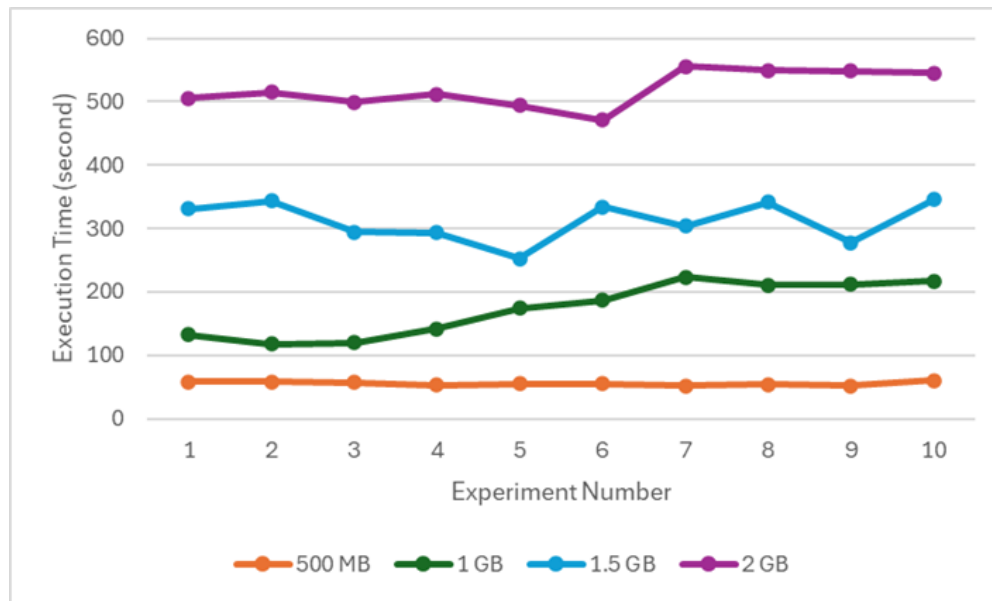


Figure 1. Terasort Execution Time using MicroSD

As shown in Figure 1, only the 500MB data size exhibits stable performance. The results of the tests using eMMC storage can be seen in Table 2.

Table 2. Terasort Test Results Using eMMC

Experiment No.	Execution Time (second)			
	500 MB	1 GB	1.5 GB	2 GB
1	43	71	107	146
2	41	68	102	141
3	40	68	94	136
4	42	60	92	137
5	42	65	87	140
6	41	65	97	135
7	44	62	98	139
8	43	66	103	132
9	43	67	99	135
10	40	67	89	128

The table 2 presents the execution times (in seconds) for processing different data sizes, namely 500MB, 1GB, 1.5GB, and 2GB over 10 experimental trials each. The data was processed in a single-node Hadoop environment, and these results reflect the performance using eMMC storage. For the 500MB dataset, execution times remained quite consistent, ranging from 40 to 44 seconds, demonstrating steady performance for smaller data sizes. As the data size increased to 1GB, the execution times varied slightly, ranging from 60 to 71 seconds, indicating a manageable increase in processing time as the dataset grew.

For the 1.5GB dataset, execution times ranged from 87 to 107 seconds, showing a more noticeable increase in time as the data size became larger, but still with relatively consistent performance across the trials. The 2GB dataset required the longest execution times, ranging from 128 to 146 seconds, which demonstrates a significant increase in processing time compared to smaller datasets.

These results show that as the dataset size increases, the execution time follows a predictable upward trend, with larger datasets taking more time to process. However, the eMMC storage performed efficiently across all data sizes, with minimal fluctuations in execution times. This relationship between data size and execution time can be visualized in Figure 2 for a clearer comparison across different data sizes.

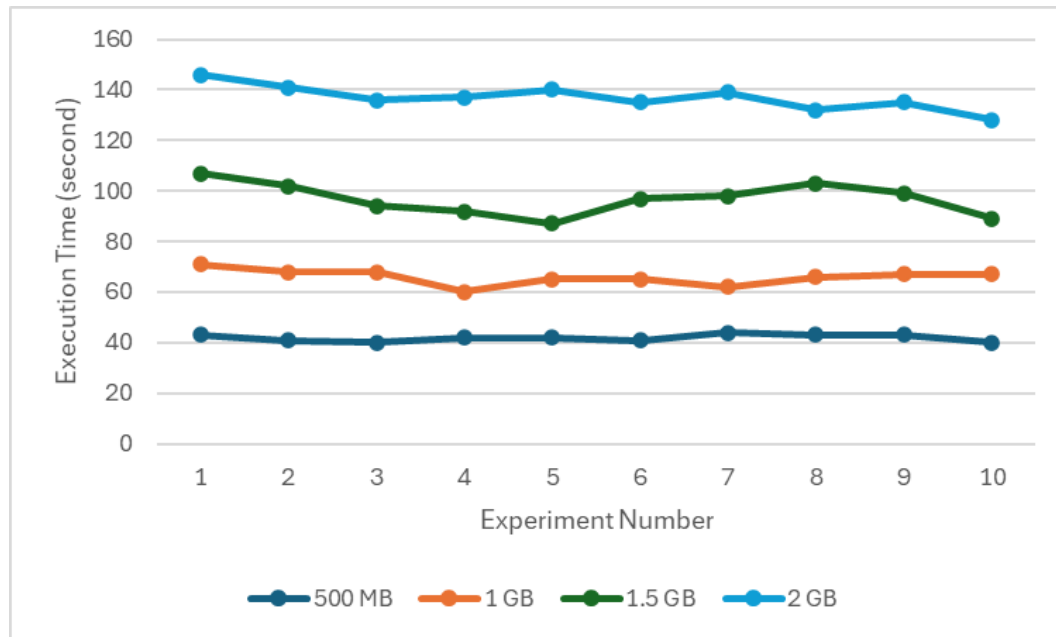


Figure 2. Terasort Execution Time using eMMC

As shown in Figure 2, the Orange Pi 5 Pro operates more consistently when using eMMC compared to MicroSD. The execution time comparison can be seen in Table 3.

Table 3. Comparison of Terasort Execution Times Between MicroSD and eMMC

Data Size	Execution Time (second)	
	MicroSD	eMMC
500MB	55.4	41.9
1 GB	173.6	65.9
1.5 GB	311.8	96.8
2 GB	519.5	136.9

Table 3 shows that eMMC consistently delivers faster execution times than MicroSD for all data sizes. For a 500MB dataset, eMMC is 13.5 seconds faster than MicroSD. At 1GB, the difference in execution times increases significantly to 107.7 seconds. For 1.5GB, eMMC is ahead by 215 seconds, and at 2GB, eMMC is 382.6 seconds faster than MicroSD.

This difference indicates that the larger the data size, the larger the difference in execution time between MicroSD and eMMC, with eMMC continuing to show much superior performance. Especially at 2GB, eMMC is almost four times faster than MicroSD. These results can be seen in a more understandable graphical form in Figure 3.

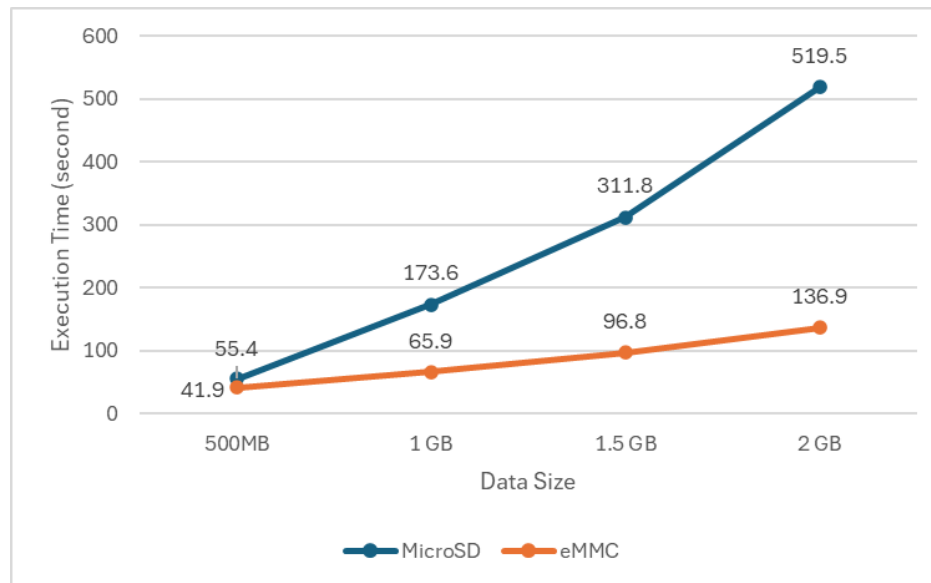


Figure 3. Terasort Execution Time using eMMC

In Figure 3, the execution time required for each type of storage is clearly visible. This graph shows that the execution time using eMMC for 2GB of data is faster than the execution time for 1GB of data when using MicroSD. After the Terasort test, the next step is to perform the TestDFSIO test. The results of the TestDFSIO test can be seen in Table 4.

Table 4. Comparison of TestDFSIO Between MicroSD and eMMC

Storage	Data Size (MB)	Number of files	Total processed (MB)	Write		Read	
				Time (s)	Speed (MB/s)	Time (s)	Speed (MB/s)
MicroSD	1024	10	10240	312.39	32.78	134.21	76.30
eMMC	1024	10	10240	206.86	49.50	72.82	140.62

Table 4 shows the results of the TestDFSIO test that measures the read and write performance of MicroSD and eMMC storage, with a data size of 1GB per file and a total of 10 files (equivalent to 10GB of data). In the write test, MicroSD took 312.39 seconds to process a total of 10GB of data, with a write speed of 32.78 MB/s. In contrast, eMMC performed much better, taking only 206.86 seconds to write the same data with a write speed of 49.50 MB/s. This shows that eMMC has much higher write efficiency than MicroSD, with a speed increase of up to 50% in this test.

In the read test, MicroSD took 134.21 seconds to read a total of 10GB of data with a read speed of 76.30 MB/s. Meanwhile, eMMC only took 72.82 seconds, with a much higher read speed of 140.62 MB/s. The eMMC read speed in this test is almost twice that of MicroSD. Overall, eMMC shows better read and write speeds than MicroSD, both in terms of execution time and data transfer speed, with significant differences in both tests.

CONCLUSION

From the results of this study, it can be concluded that eMMC consistently demonstrates better performance compared to MicroSD in both the Terasort and TestDFSIO tests. In the Terasort test, eMMC produced faster execution times across all data sizes, with the time difference becoming more pronounced as the data size increased. For instance, at a data size of 2GB, eMMC was nearly four times faster than MicroSD, highlighting its efficiency and capability in handling larger datasets. The TestDFSIO results further support these findings, with eMMC showing a higher write speed—206.86 seconds at a speed of 49.50 MB/s, compared to MicroSD's 312.39

seconds at a speed of 32.78 MB/s. The read speed of eMMC was also significantly better, taking only 72.82 seconds with a speed of 140.62 MB/s, nearly double that of MicroSD. Overall, this study emphasizes that eMMC is a more efficient choice for applications requiring data storage in a Hadoop environment, especially when managing large datasets. The more stable and faster performance of eMMC makes it better suited for intensive computing needs and big data processing tasks, where rapid access to and processing of large amounts of data are crucial. These findings underscore the importance of selecting storage media that can handle the demands of big data effectively, with eMMC emerging as a more suitable option for these environments.

ACKNOWLEDGEMENTS

This research was funded by the Rector of Universitas Muslim Indonesia (UMI) Makassar through the PUF (Penelitian Unggulan Fakultas) program in 2024. The authors would like to express their gratitude to UMI for the financial support that made this research possible.

REFERENCES

- Abiola, S., Mengel, S., & Sill, A. (2024). Use of low-end single board computer clusters to prototype cluster administration and benchmarking. *Practice and Experience in Advanced Research Computing 2024: Human Powered Computing*, 1–3. <https://doi.org/10.1145/3626203.3670618>
- Adnan, A., Tahir, Z., & Asis, M. A. (2019). Performance Evaluation of Single Board Computer for Hadoop Distributed File System (HDFS). *2019 International Conference on Information and Communications Technology (ICOIACT)*, 624–627. <https://doi.org/10.1109/ICOIACT46704.2019.8938434>
- Adnan, Tahir, Z., Yohannes, C., & Ariel. (2020). Performance Evaluation of Mini Single Board Computer in Hadoop Big Data Cluster. *IOP Conference Series: Materials Science and Engineering*, 875(1), 012037. <https://doi.org/10.1088/1757-899X/875/1/012037>
- Ariza, J. Á., & Baez, H. (2021). Understanding the role of single-board computers in engineering and computer science education: A systematic literature review. *Computer Applications in Engineering Education*, 10(7), cae.22439. <https://doi.org/10.1002/cae.22439>
- Dwi Agustin, R. (2022). Rancang Bangun Alat Monitoring Pasang Surut Air Laut Berbasis IoT dengan NodeMCU ESP8266 dan HC-SR04. *Jurnal Ilmu Dan Inovasi Fisika*, 6(2), 147–157. <https://doi.org/10.24198/jiif.v6i2.40345>
- Elkawagy, M., & Elbeh, H. (2020). High Performance Hadoop Distributed File System. *International Journal of Networked and Distributed Computing*, 8(3), 119. <https://doi.org/10.2991/ijndc.k.200515.007>
- Godinho, A., Rosado, J., Sá, F., & Cardoso, F. (2023). IoT single board computer to replace a home server. *2023 18th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6. <https://doi.org/10.23919/CISTI58278.2023.10212031>
- Heru Sandi, G., & Fatma, Y. (2023). PEMANFAATAN TEKNOLOGI INTERNET OF THINGS (IOT) PADA BIDANG PERTANIAN. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(1), 1–5. <https://doi.org/10.36040/jati.v7i1.5892>
- Hoffmann, R. B., Griebler, D., Righi, R. da R., & Fernandes, L. G. (2024). Benchmarking parallel programming for single-board computers. *Future Generation Computer Systems*, 161, 119–134. <https://doi.org/10.1016/j.future.2024.07.003>
- Jayshree, J. (2022). Big Data Analytics : Use of Hadoop Mapreduce. *Great Britain Journals Press*, 22(2). <https://journalspress.uk/index.php/LJRCST/article/view/865>
- Jing, W., Tong, D., Wang, Y., Wang, J., Liu, Y., & Zhao, P. (2017). MaMR: High-performance

- MapReduce programming model for material cloud applications. *Computer Physics Communications*, 211, 79–87. <https://doi.org/10.1016/j.cpc.2016.07.015>
- Kim, J., Kancharla, A., Seol, J., Park, I., & Park, N. (2018). Optimized Common Parameter Set Extraction Framework by Multiple Benchmarking Applications on a Big Data Platform. *International Journal of Networked and Distributed Computing*, 6(4), 195. <https://doi.org/10.2991/ijndc.2018.6.4.1>
- Kuss, F. S., Castilho, M. A., Peres, L. M., & Silva, F. (2018). Aulacast: A single board computer platform to support teaching. *CSEDU 2018 - Proceedings of the 10th International Conference on Computer Supported Education*, 1, 366–373. <https://doi.org/10.5220/0006776803660373>
- Lee, E., Oh, H., & Park, D. (2021). Big Data Processing on Single Board Computer Clusters: Exploring Challenges and Possibilities. *IEEE Access*, 9, 142551–142565. <https://doi.org/10.1109/ACCESS.2021.3120660>
- Ma, C., Zhao, M., & Zhao, Y. (2023). An overview of Hadoop applications in transportation big data. *Journal of Traffic and Transportation Engineering (English Edition)*, 10(5), 900–917. <https://doi.org/10.1016/j.jtte.2023.05.003>
- Mantik, J., & Jaya, A. C. (2021). Single-Board Computer For Affordable Personal Data Storage Server. *Jurnal Mantik*, 5(2).
- Oliveira, R. A. de, & Bollen, M. H. J. (2023). Deep learning for power quality. *Electric Power Systems Research*, 214, 108887. <https://doi.org/10.1016/j.epsr.2022.108887>
- Patel, B. A. M., & Hamirani, H. R. (2021). Affordable Educational Laptop With Physical Computing Using Single Board Computer. *2021 International Conference on Computer Communication and Informatics (ICCCI)*, 1–4. <https://doi.org/10.1109/ICCCI50826.2021.9402691>
- Qureshi, B. (2024). Optimizing Hadoop Scheduling in Single-Board-Computer-Based Heterogeneous Clusters. *Computation*, 12(5), 96. <https://doi.org/10.3390/computation12050096>
- Ramadhani, R. D., Thohari, A. N. A., & Nugraha, N. A. (2020). Sistem Keamanan Ruang Berbasis Internet of Things Menggunakan Single Board Computer. *Jurnal Nasional Informatika Dan Teknologi Jaringan*, 1(2), 0–5.
- Susanto, F. L. (2024). Implementation of Big Data Analytics and its Challenges in Digital Transformation Era: A Literature Review. *IJIIS: International Journal of Informatics and Information Systems*, 7(2), 90–99. <https://doi.org/10.47738/ijis.v7i2.204>
- Sutikno, T., Satrian Purnama, H., Pamungkas, A., Fadlil, A., Mohd Alsofyani, I., & Jopri, M. H. (2021). Internet of things-based photovoltaics parameter monitoring system using NodeMCU ESP8266. *International Journal of Electrical and Computer Engineering (IJECE)*, 11(6), 5578. <https://doi.org/10.11591/ijece.v11i6.pp5578-5587>
- Tahsir Ahmed Munna, M., Muhammad Allayear, S., Mohtashim Alam, M., Shah Mohammad Motiur Rahman, S., Samadur Rahman, M., & Mesbahuddin Sarker, M. (2018). Simplified Mapreduce Mechanism for Large Scale Data Processing. *International Journal of Engineering & Technology*, 7(3.8), 16. <https://doi.org/10.14419/ijet.v7i3.8.15211>
- Thomas-Brans, F., Fukami, A., Clement, Q., Heckmann, T., & Sauveron, D. (2024). Case of study for in situ memory reading on damaged MultiMedia Card. *Forensic Science International: Digital Investigation*, 48, 301698. <https://doi.org/10.1016/j.fsidi.2024.301698>