UNIJOURN

International Journal of Sustainable Transportation Technology
Vol. 6, No. 1, April 2023, 7-11

Development of Digital Twin Platform for Electric Vehicle Battery System

Putu Handre Kertha Utama[1], Irsyad Nashirul Haq[1,2*], Edi Leksono[1], Muhammad Iqbal Juristian[1], Ghulam Azka Alim[1], Justin Pradipta[1]

[1]Engineering Physics, Faculty of Industrial Technology, Institut Teknologi Bandung, Indonesia
[2]National Center for Sustainable Transportation Technology, Indonesia

**Email: irsyad.n@itb.ac.id*

Abstract

The battery system in electric vehicles needs proper monitoring and control to ensure reliable, efficient, and safe operation. Recent advancement in cyber-physical technology has brought the emerging digital twin concept. This concept opens a new possibility of real-time condition monitoring and fault diagnosis of the battery system. Although it sounds promising, the concept implementation still faces many challenges. One of the challenges is the availability of a platform to develop digital twins, which involves data pipelines and modeling tools. The data pipeline will include the acquisition, storing, and extract-transform-load (ETL) with high velocity, volume, value, variety, and veracity data, known as big data. The modeling tools must provide applications to build the high-fidelity model, one of the required elements of the digital twin. Based on those urgencies, this paper proposes a platform that facilitates a digital twinning of the battery system in an electric vehicle. The platform is built on the open-source framework CDAP, equipped with a data pipeline and modeling tools. It has run several performance tests with different computation resource configurations and workloads. Doubling the processing power can reduce 12% of computation time while increasing memory size by four times only reduces 10% of computation time. The result shows that the processing power affects the performance digital twin platform more than the memory size.

Keywords

Digital twin; Electric vehicle; CDAP; Big data; Battery system

1 Introduction

Electric vehicle (EV) demand has been increasing over the years as EVs present contribute significant benefits from an environmental perspective [1]–[3]. One critical component of an EV is a battery system, which supplies energy for EV operation. The battery system needs proper monitoring and control to perform safe, reliable, and efficient functions [4], [5]. The onboard battery management system (BMS) is usually responsible for those roles [1], [4], [6]. Challenges will appear as the number of cells within the system increases [7]–[9]. Precisely monitoring and estimating the battery system state of such a large number of cells will require huge storage space and intensive computation power.

The emergence of a new concept called digital twins became a solution for those challenges. A digital twin is a digital counterpart of a physical entity (an object or a process) in the digital domain [10], [11]. A physical entity is represented in the digital world by high-fidelity models that mimic the state of the physical entity [10]–[12]. Apart from a digital model that only snapshots of the physical object at a time, a digital twin presents real-time states of a physical object. This concept also allows the physical object to modify its real-time behavior concurrently based on feedback generated by its digital twin counterpart [1]. It opens a new way to utilize the Internet of Things (IoT), big data, and artificial intelligence (AI) technologies to revolutionize EV BMS technologies [4], [13].

The battery system digital twin can be helpful in monitoring and making the appropriate control decision [4], [14]. The digital part within the digital twin handles data storage and computations, while onboard battery management handles control actions [15], [16]. Developing a digital twin requires five dimensions: physical, virtual, connection, data, and service, to be present [17]. From those dimensions, data plays a significant role in representing physical entities. Data of the physical entities gathered from the sensor could be entering a big-data territory [18]. Furthermore, the data need to process to produce knowledge about the state of the physical entities. It demands an analytical platform to develop models. Hence, a digital twin platform with a data pipeline and modeling tools is required.

Received August 8, 2022, Revised August 31, 2022, Accepted for publication January 3, 2023

Although much research has been dedicated to battery system digital twins [1], [4], [19], [20], only a few explain the data pipeline. This paper aims to provide a deeper analysis of the data pipeline and build a platform. The platform will enable the digital twinning of the electric vehicle battery system. In the next section, Section II, we explain the architecture of the digital twin platform and the method for testing the platform. Section III discusses the test result and concludes our findings in Section IV.

2 Methodology

2.1 Digital Twin Platform

The digital twin platform becomes the necessary infrastructure to develop a digital twin. It provides the pipeline to manage gathered data and modeling tools to generate a high-fidelity model. In this paper, the platform is built on an open-source CDAP framework shown in Figure 1. The platform components are virtual machines that run on a server. Table 1 gives detailed specifications of the server and each virtual machine/instance in Figure 1.

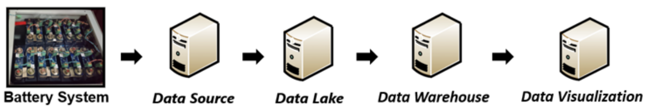


Figure 1 Digital twin data flow

The physical device of the Data Source in the field can be an industrial personal computer (PC) or an embedded system. The Data Source task is to gather all the data from the sensors in the vicinity. It stores the operation data of the battery system in the form of SQL and continuously sends recent data to the Data Lake. As the name suggests, the Data Lake pools the acquired data and stores it in the form of a parquet. We chose this data format because data stored in parquet occupy less space [21] and execute queries faster than the standard row format data [22].

Table 1 Digital twin resources configuration

Config.	Server	Data Lake	Data Warehouse	Data Visualization
LAN IP	192.168.1.86	192.168.1.45	192.168.1.44	192.168.46
CPU	24CPU Intel® Xeon® @1.8GHz	8vCPU	4vCPU	2vCPU
Cores	24	8	4	4
Memory	128GB	64GB	8GB	16GB
Storage	1TB	120GB	120GB	120GB

The Data Lake will also be responsible for creating a battery model to estimate the State of Charge (SoC). The battery model is a deep neural network (DNN) based on [23]. Data gathered in the Data Lake will be utilized for training the DNN model. After that, the Data Lake will send the estimation result to Data Warehouse. The last stage is Data Visualization. This virtual machine act as a human-machine interface to give the user information about the updated battery SoC. The Data visualization will display the battery SoC by Application Programming Interface (API) provided by the Data Warehouse.

2.2 Platform Functional and Performance Test

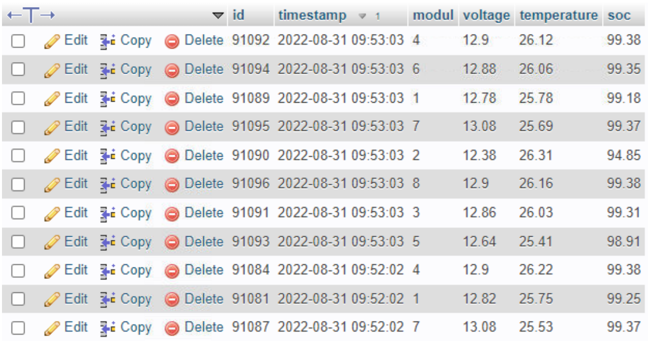
To ensure the function and performance of the platform, it needs to go through several tests. The tests refer to The Performance Measurement Framework for Cloud Computing (PMFCC) and ISO 25010 [24]. It divides into two parts. First, services delivery that includes (1) the platform can deliver the intended tasks and (2) the presence of null values and duplicates in the Data Warehouse. Data stored in the Data Warehouse should not contain null values and duplicates because this data will send to users via Data Visualization. Second, the performance efficiency test will include (1) computation resource utilization and (2) execution time. The latter will show the relationship between execution time and computation resources. It may reveal the factor that affects the platform performance and potential bottleneck.

For the first test, the platform should estimate the SoC of a battery system. Data preparation and estimation programs are built based on Python programming languages. It begins with the Data Lake taking the battery operation data. Battery data in parquet transform to data-frame with Pandas module in Python. Those data are processed before it feeds to the DNN battery model. The DNN will estimate the SoC of the battery, transform data to SQL format and send the result to Data Warehouse. In the second test, we varied both data size and resource configurations. The platform task is to estimate the SoC of different operation data sizes. We also configure the Data Lake resources (CPU cores and RAM size) with different combinations.

3 Manufacturing Process

3.1 Functional Test

In functional tests, the platform successfully carried out the task of estimating the battery SoC. Data can flow throughout the data pipeline from the Data Lake to the Data Warehouse. Results show null values and duplicates were not present in the Data Warehouse because the Data Lake already cleaned null values of battery operation data from the Data Source before being further used. A Unique ID is also assigned for each battery operation data to avoid duplication shown in Figure 2.



				id	timestamp	modul	voltage	temperature	soc
☐	Edit	Copy	Delete	91092	2022-08-31 09:53:03	4	12.9	26.12	99.38
☐	Edit	Copy	Delete	91094	2022-08-31 09:53:03	6	12.88	26.06	99.35
☐	Edit	Copy	Delete	91089	2022-08-31 09:53:03	1	12.78	25.78	99.18
☐	Edit	Copy	Delete	91095	2022-08-31 09:53:03	7	13.08	25.69	99.37
☐	Edit	Copy	Delete	91090	2022-08-31 09:53:03	2	12.38	26.31	94.85
☐	Edit	Copy	Delete	91096	2022-08-31 09:53:03	8	12.9	26.16	99.38
☐	Edit	Copy	Delete	91091	2022-08-31 09:53:03	3	12.86	26.03	99.31
☐	Edit	Copy	Delete	91093	2022-08-31 09:53:03	5	12.64	25.41	98.91
☐	Edit	Copy	Delete	91084	2022-08-31 09:52:02	4	12.9	26.22	99.38
☐	Edit	Copy	Delete	91081	2022-08-31 09:52:02	1	12.82	25.75	99.25
☐	Edit	Copy	Delete	91087	2022-08-31 09:52:02	7	13.08	25.53	99.37

Figure 2 Data warehouse database

3.2 Performance Test

For performance tests, we set the Data Lake in two configurations, specified in Table 2. The Data Lake was deployed to estimate SoC with different data sizes in parquet format, ranging from a thousand rows up to a million rows, each containing 11 columns, as shown in Table 3. Figure 3 depicts the relation between the execution time and data size for this test. Configuration 2 can execute the task faster than Configuration 1 for small data sizes. However, the opposite things happened for data sizes above 477 KB. Trendlines in Figure 3 suggest that the execution time increases linearly with data size. Correlation is strong, with an R-square of 0.99.

Table 2 Data Lake configuration

Configuration	vCPU (core)	RAM (GB)
Configuration 1	8	64
Configuration 2	4	16

Table 3 Data test

Data	Number of Row	Data Size (KB)
1	1,000	6.20
2	10,000	47.9
3	100,000	477
4	500,000	2,400
5	1,000,000	4,900

During tests, we also measure the utilization of CPU and RAM. Figure 4 and Figure 5 show the utilization of CPU and RAM for Data Lake Configuration 1 and Configuration 2. RAM usage for Configuration 2 is nearly maximum at a data size of 4.900 KB (Figure 4). It causes Configuration 2 to need more time than Configuration 1 (Figure 2). RAM is the cause of the bottleneck for Configuration 2 while still having enough CPU for computation.

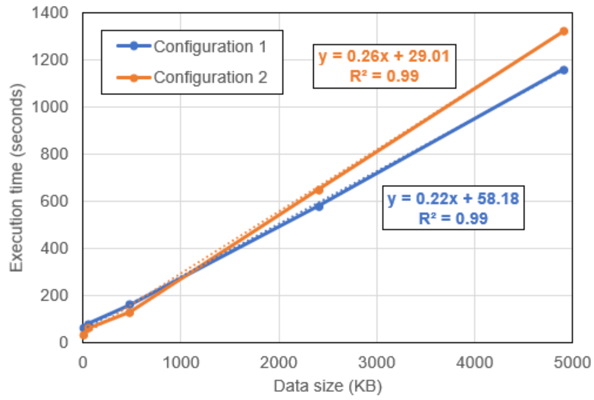


Figure 3 Relation between execution time and data size

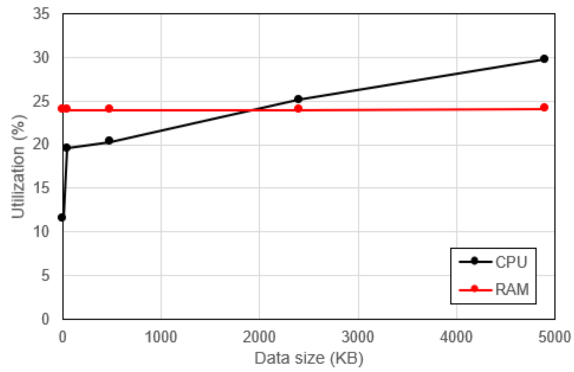


Figure 4 Computation resources utilization for configuration 1

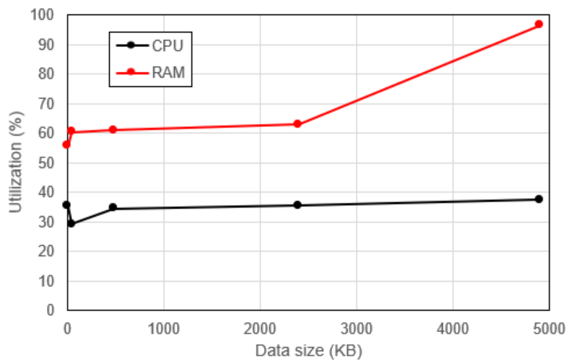


Figure 5 Computation resources utilization for configuration 2

The above results indicate that RAM size might limit the platform performance, so we configure the Data Lake with different RAM (Table 4). Each configuration was deployed to estimate the SoC of 4,900 KB battery data. Figure 6 shows that increasing RAM size can reduce execution time. This result shows that the digital twin platform is affected by memory size availability. Strong exponential relation was present, indicated by the R2 value of 0.99. This exponential relation between execution time and RAM size suggests an optimum number size of RAM. At some point, adding more will only slightly reduce the execution time.

Table 4 Data Lake for different ram sizes

vCPU (core)	RAM (GB)
8	18
8	16
8	32
8	64

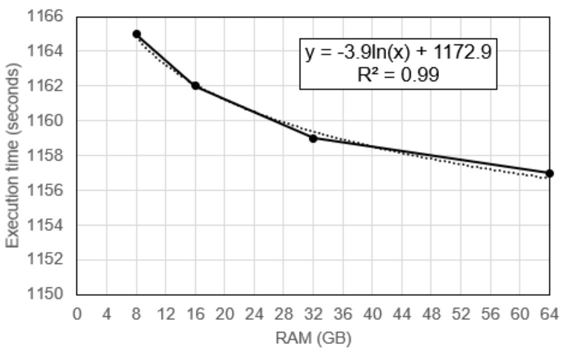


Figure 6 Execution time for different RAM size configuration

Further, we also configured the Data Lake with different CPU cores (Table 5). As Figure 7 suggests, adding more computation power (CPU cores) can reduce execution time. Although the trendline correlation (R2) was not as strong as in previous results, it suggests that upsizing computation power makes task execution faster than expanding memory sizes.

Table 5 Data Lake for different CPU core

vCPU (core)	RAM (GB)
1	64
2	64
4	64
8	64

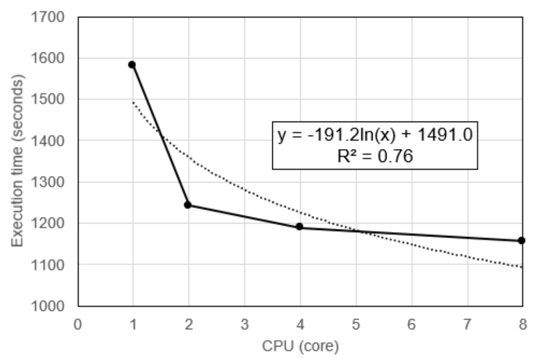


Figure 7 Execution time for different CPU core configuration

We summarize the result of the performance test in Table 6. Doubling the CPU cores of Configuration 2 reduces computation time by more than quadrupling RAM size. Notice that the Data Lake with 8 CPU cores and 64 GB RAM is only slightly better than the 16 GB RAM with the same number of cores. It means the platform performance is affected by computation power.

Table 6 Performance test summary

vCPU (core)	RAM (GB)	Execution Time (seconds)	Condition
4	16	1320	Configuration 2
8	16	1162	Doubling CPU core of Configuration 2
4	64	1189	Quadrupling RAM size of Configuration 2
8	64	1157	Doubling CPU core and Quadrupling RAM size of Configuration 2 (Configuration 1)

4 Conclusion

The result of performance tests suggests that SoC estimation is a computation-intensive process. Neural network SoC estimation was a heavily repeating calculation pattern because of its multiplication between the input and weight matrix. Another way to improve execution time is parallel computing via GPU. Adding GPU to the existing digital twin platform will be our future research.

Although computation resources affect the digital twin platform performance more than RAM or memory sizes, adequate RAM is also necessary to ensure optimal computation time. At this point, we only manage to collect four data points for the CPU variation test. More data points are required to have a better R2 value. Hence, we will continue our test with different variations of CPU and RAM to get more data points.

Acknowledgment

This research was partially supported by the ITB Research Grant, Lembaga Pengelola Dana Pendidikan, the Ministry of Research and Technology/National Research and Innovation Agency of Indonesia, and PT Pertamina RTI.

References

[1] G. Bhatti, H. Mohan, and R. Raja Singh, "Towards the future of smart electric vehicles: Digital twin technology," *Renew. Sustain. Energy Rev.*, vol. 141, p. 110801, 2021, doi: https://doi.org/10.1016/j.rser.2021.110801.

[2] H. E. O. Farias and L. N. Canha, "Battery Energy Storage Systems: Impact Analysis on Different Loads with Distributed Generation," in *2019 IEEE PES Innovative Smart Grid Technologies Conference - Latin America (ISGT Latin America)*, 2019, pp. 1–5, doi: 10.1109/ISGT-LA.2019.8895361.

[3] J. Cao, X. Chen, R. Qiu, and S. Hou, "Electric vehicle industry sustainable development with a stakeholder engagement system," *Technol. Soc.*, vol. 67, p. 101771, 2021, doi: https://doi.org/10.1016/j.techsoc.2021.101771.

[4] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte, and D. U. Sauer, "Digital twin for battery systems: Cloud battery

management system with online state-of-charge and state-of-health estimation," *J. Energy Storage*, vol. 30, p. 101557, 2020, doi: https://doi.org/10.1016/j.est.2020.101557.

[5] K. Alamin, Y. Chen, E. Macii, M. Poncino, and S. Vinco, "A Machine Learning-based Digital Twin for Electric Vehicle Battery Modeling," *ArXiv*, vol. abs/2206.0, 2022.

[6] I. N. Haq *et al.*, "Development of battery management system for cell monitoring and protection," in *2014 International Conference on Electrical Engineering and Computer Science (ICEECS)*, 2014, pp. 203–208, doi: 10.1109/ICEECS.2014.7045246.

[7] F. Yang, W. Li, C. Li, and Q. Miao, "State-of-charge estimation of lithium-ion batteries based on gated recurrent neural network," *Energy*, vol. 175, pp. 66–75, 2019, doi: https://doi.org/10.1016/j.energy.2019.03.059.

[8] C. Piao, Z. Huang, L. Su, and S. Lu, "Research on outlier detection algorithm for evaluation of battery system safety," *Adv. Mech. Eng.*, vol. 2014, 2014, doi: 10.1155/2014/830402.

[9] M. I. Karmawijaya, I. N. Haq, E. Leksono, and A. Widyotriatmo, "Development of Big Data Analytics Platform for Electric Vehicle Battery Management System," in *2019 6th International Conference on Electric Vehicular Technology (ICEVT)*, 2019, pp. 151–155, doi: 10.1109/ICEVT48285.2019.8994013.

[10] R. Minerva, G. M. Lee, and N. Crespi, "Digital Twin in the IoT Context: A Survey on Technical Features, Scenarios, and Architectural Models," *Proc. IEEE*, vol. 108, no. 10, pp. 1785–1824, 2020, doi: 10.1109/JPROC.2020.2998530.

[11] B. R. Barricelli, E. Casiraghi, and D. Fogli, "A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications," *IEEE Access*, vol. 7, pp. 167653–167671, 2019, doi: 10.1109/ACCESS.2019.2953499.

[12] N. Bazmohammadi *et al.*, "Microgrid Digital Twins: Concepts, Applications, and Future Trends," *IEEE Access*, vol. 10, pp. 2284–2302, 2022, doi: 10.1109/ACCESS.2021.3138990.

[13] M. M. Rathore, S. A. Shah, D. Shukla, E. Bentafat, and S. Bakiras, "The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities," *IEEE Access*, vol. 9, pp. 32030–32052, 2021, doi: 10.1109/ACCESS.2021.3060863.

[14] A. Fuller, Z. Fan, C. Day, and C. Barlow, "Digital Twin: Enabling Technologies, Challenges and Open Research," *IEEE Access*, vol. 8, pp. 108952–108971, 2020, doi: 10.1109/ACCESS.2020.2998358.

[15] Z. Wang *et al.*, "A Digital Twin Paradigm: Vehicle-to-Cloud Based Advanced Driver Assistance Systems," in *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, 2020, pp. 1–6, doi: 10.1109/VTC2020-Spring48590.2020.9128938.

[16] Q. Lin, J. Wang, R. Xiong, W. Shen, and H. He, "Towards a smarter battery management system: A critical review on optimal charging methods of lithium ion batteries," *Energy*, vol. 183, pp. 220–234, 2019, doi: https://doi.org/10.1016/j.energy.2019.06.128.

[17] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital Twin in Industry: State-of-the-Art," *IEEE Trans. Ind. Informatics*, vol. 15, no. 4, pp. 2405–2415, 2019, doi: 10.1109/TII.2018.2873186.

[18] A. Rasheed, O. San, and T. Kvamsdal, "Digital Twin: Values, Challenges and Enablers From a Modeling Perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020, doi: 10.1109/ACCESS.2020.2970143.

[19] S. Singh, M. Weeber, and K. P. Birke, "Implementation of Battery Digital Twin: Approach, Functionalities and Benefits," *Batteries* , vol. 7, no. 4. 2021, doi: 10.3390/batteries7040078.

[20] L. Merkle, M. Pöthig, and F. Schmid, "Estimate e-Golf Battery State Using Diagnostic Data and a Digital Twin," *Batteries* , vol. 7, no. 1. 2021, doi: 10.3390/batteries7010015.

[21] P. Pirzadeh, M. Carey, and T. Westmann, "A performance study of big data analytics platforms," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 2911–2920, doi: 10.1109/BigData.2017.8258260.

[22] Á. A. Isla, "HDFS File Formats: Study and Performance Comparison," 2018.

[23] P. H. K. Utama, H. H. Husniyyah, I. N. Haq, J. Pradipta, and E. Leksono, "State of Charge (SoC) Estimation of Battery Energy Storage System (BESS) Using Artificial Neural Network (ANN) Based on IoT- Enabled Embedded System," in *2021 International Conference on Instrumentation, Control, and Automation (ICA)*, 2021, pp. 77–82, doi: 10.1109/ICA52848.2021.9625697.

[24] L. E. Bautista Villalpando, A. April, and A. Abran, "Performance analysis model for big data applications in cloud computing," *J. Cloud Comput.*, vol. 3, no. 1, p. 19, 2014, doi: 10.1186/s13677-014-0019-z.