Implementation of Data Abstraction Layer Using Kafka on SEMAR Platform for Air Quality Monitoring

Yohanes Yohanie Fridelin Panduman[#], Mochamad Rifki Ulil Albaab[#], Adnan Rachmat Anom Besari[+],
Sritrusta Sukaridhoto[+,*], Anang Tjahjono[#], Rizqi Putri Nourma Budiarti[&]

[#] *Electrical Engineering Department, Politeknik Elektronika Negeri Surabaya, Surabaya, 60111, Indonesia*
E-mail: yohanyfp@pasca.student.pens.ac,id, mochrifkiulila@pasca.student.pens.ac.id, anang_tj@yahoo.co.id

[+] *Informatics and Computer Department, Politeknik Elektronika Negeri Surabaya, Surabaya, 60111, Indonesia*
E-mail: anom@pens.ac.id, [*]dhoto@pens.ac.id (corresponding author)

[&]*Department of Engineering, Nahdlatul Ulama University of Surabaya, Surabaya, 60237, Indonesia*
E-mail: rizqi.putri.nb@unusa.ac.id

Abstract—**Urbanization and fast-growing industries causing air quality in urban areas to be bad and even tend to be dangerous. In addition, the largest percentage of energy emissions come from the transportation sector, specifically on road transportation. Therefore, the need for a quality detection system that is capable of distributing and displaying large data information in real-time cannot be resolved by the system currently used by the government. This research offers a solution to the implementation of data abstraction in cloud computing which is built using the concept microservice architecture and integrated with mobile-based sensors to detect air quality in real-time. This solution consists of integrated cloud computing services using Smart Environment Monitoring and Analytical in Real-time (SEMAR) and Vehicles as Mobile Sensor Networks (VaaMSN) to detecting air quality. SEMAR was built with microservice references consist of data abstraction, communication, data analytical with business analytics proccess, data storage with Big data service and also real-time visualization in maps, chart, and table through dasboard website. Through the experiments that we did show that the microservice of data abstraction layer can be installed at the SEMAR stage indicating that the average delay in sending information is around 0.09 ms (90μs), this indicates that the system can be said to be real-time. With specific and real-time locations in data visualization, the government can use this method as an new alternative method of air quality.**

Keywords— **cloud computing; air quality; Kafka; data abstraction; internet of things.**

I. INTRODUCTION

The transportation development increase the urban people health. In 2010, Jakarta's population was 9,607,787, and 57.8% of the population experienced the effects of air pollution. These people suffer from various diseases related to air pollution, causing an increase in health costs up to Rp. 38.5 trillion / USD54 billion [2]. Also, Indonesia is the 6th largest emitter of greenhouse gases in the world (IEA 2015) where 40% of Energy Emission Percentages come from the transportation sector, and 90% of these transportation emissions come from road transportation.

The current government uses several air quality sensor that installed on the air monitoring station in a fixed position to measure air condition. However, the system has weaknesses like difficult to maintenance, high costs because it requires a lot of devices, only cover a small area and need more effort to maintain the station. The system does not use the Big Data environment but uses a general database. Therefore it does not apply the concept of data abstraction for data distribution and machine learning methods for air quality analysis. The air quality parameters in Indonesian air pollution rules are Particulate (PM_{10}), Carbon monoxide (CO), Ozone (O_3), Sulfur dioxide (SO_2) and Nitrogen dioxide (NO_2) [3].

There are several studies on real-time environmental monitoring used cloud computing that implements with big data concept, one of which is Smart Environment Monitoring and Analytics in Real-time System (SEMAR) [4-8]. SEMAR is used to detect river water quality using portable devices integrated with water quality sensors [4] and also ROV (small robot submarine) that are installed with water quality sensors [5], [6]. This is an alternative solution to assist governments in monitoring water environments in urban areas. Water quality samples detected using a water quality sensor were installed in the ROV, then sent into

Hadoop Big Data platform in server. The next development is adding an Internet of Things (IoT) device that is integrated with the SEMAR Platform with Big Data analyzers for real-time water quality monitoring [7], [8]. The latest development, the SEMAR Platform, is integrated with Mobile sensor devices to detect air quality conditions [9]. However, on the SEMAR extension cloud computing platform, it has not implemented the concept of data abstraction for data distribution. Therefore, SEMAR platform can be further developed with server devices built in clustering concept [10], [11] to support the microservice architecture concept [12]. Thus, the data abstraction concept [13] could be added to increase the speed time of processing data and reduce costs.

This paper consists of four sections. The Introduced section in section 1 present this research background problem and presents the previous study and related works about this research. The system details and method in this research presented in Section 2. Then Section 3 is a section that presents the research results and also contains discussion of that results of research.

A. Architecture Microservices

Previous research and work explained that cloud computing was developed using the microservice architecture for computing on IoT. In 2017 Long Sun et al. conducted a study on microservice architecture showed that the use of microservice architecture increased scalability, adaptability, and interoperability stronger so that it could provide good support for developing cloud services [14]. Then Mario Villamizar, in his research, showed that the cost of microservice architectural infrastructure could be reduced by 70%. Experiments are carried out by comparing servers built using microservice architecture and other infrastructure [15].

B. Data Abstraction

Previous research and work about data abstraction, which is used to improve cloud computing service. In 2016, Shi et al. showed the function of data abstraction that used to collect data and combine data representations with being used in the application layer [16]. Through research conducted by Rindra Wiska et al. [17], explained that Kafka could be used to distribute large amounts of data for monitoring air quality, where Kafka has a function as data abstraction to collect sensor data to data storage.

C. Air Quality Detection

Previous work and studies have developed cloud computing technology that is integrated with air monitoring systems for real-time data detection, research conducted by KSE Phala researchers show the integration of wireless communication sensors with an Air Quality Monitoring System (AQMS) device for detection in real-time in 2016 [14], [18]. Other research that published by L. Kang, et al. explained about cloud services are integrated with smart buses to detect environmental conditions such as air quality and road damage [15], [19]. This system collects urban environment sensing data which is obtained by sensor devices to database servers and displays data in electronic map sites.

II. Materials and Method

The system in this research is present in Figure 1 which consists of 2 major parts, namely VaaMSN as a device for detecting air quality parameters consisting of GPS sensors and Smart Hub Car as edge computing devices that are connected with an air quality detection system. The next part is SEMAR Cloud Computing is cloud computing built using microservice architecture which consists of microservice of Communication, microservice of Abstraction Data, microservice of Data Analysis, microservice of Data Storage, and microservice of Visualization. In this system, a smart HUB car is used to distribute sensor data to the SEMAR Cloud Computing server.
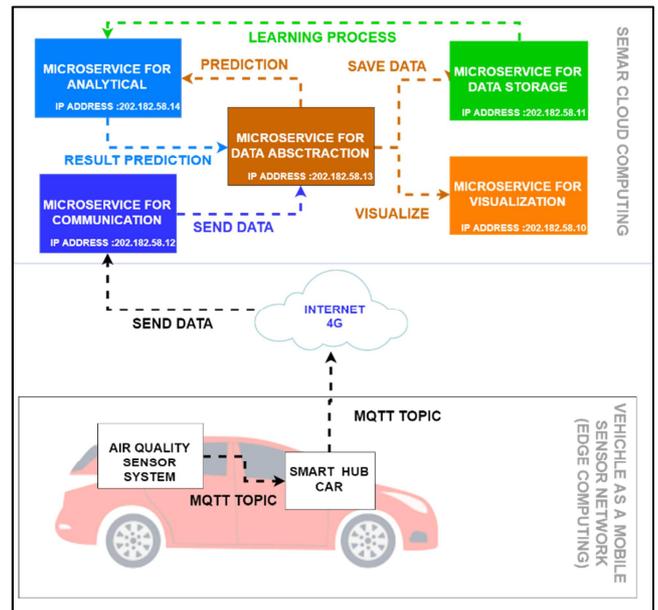

Fig.1 System Design of Research

A. System of Air Quality Detection Sensor

We use several sensors such as Sharp GP2Y1010AU0F (Particulate), MQ131 (Ozone), MiCS 2714 (Nitrogen dioxide), MQ135 (Sulfur dioxide) and MQ7 (Carbon monoxide) to detect air quality parameters. The sensors are installed on the microcontroller as an embedded system. This embedded system is used to collect data sensors and convert them into ug / m3 air quality units.
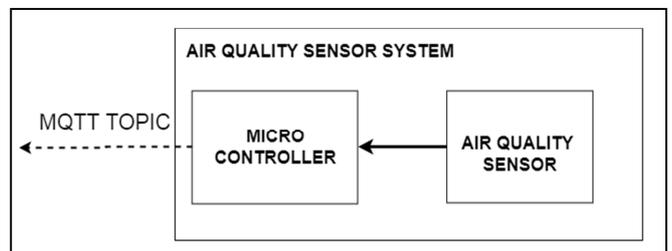

Fig.2 Design of Detection Sensor System for Air Quality.

The detailed system of air quality detection sensors is presented in figure 2. This part system is built based on the concept of wireless sensor networks. Through the MQTT communication protocol, the converted data is sent to other devices to be processed to the server using a WiFi network.

The air quality detection sensor system takes information periodically in 5 seconds.

B. System of Smart HUB Car

This part system as the primary device used to processing and sending the sensor data of air quality obtained by air quality sensor devices to cloud computing. Smart HUB car consists of SBC, WiFi 4G, and GPS. Smart HUB car connected with air quality sensor systems through a Wifi connection. Through MQTT communication The air quality sensor devices sent sensor data with "sensor_device" topic MQTT, then the data received in SCB is added with location data from GPS sensor and sent to cloud computing using topic "air sensors" through MQTT communication.
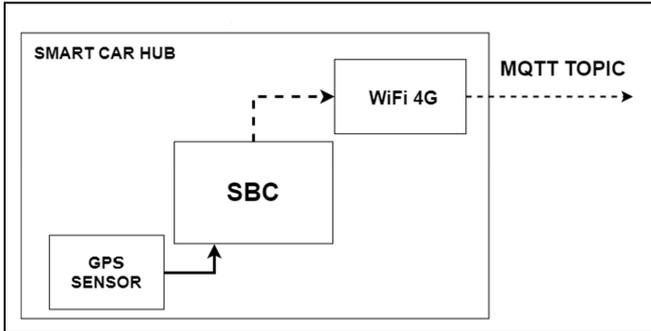

Fig.3 System Design of Smart HUB Car.

Figure 2 shows a detailed system of this Smart HUB Car. The result of this part of the system is sending air quality data to SEMAR by formatting the line 'information of air quality, location, and time of taking the data'.

C. Microservice for Communication

The part of this system is utilized to the administration of the information stream between SEMAR and VaaMSN. This framework utilized for dispersed information from VaaMSN to microservice for data abstraction. This microservice depended on the IP address '202.182.58.12'.
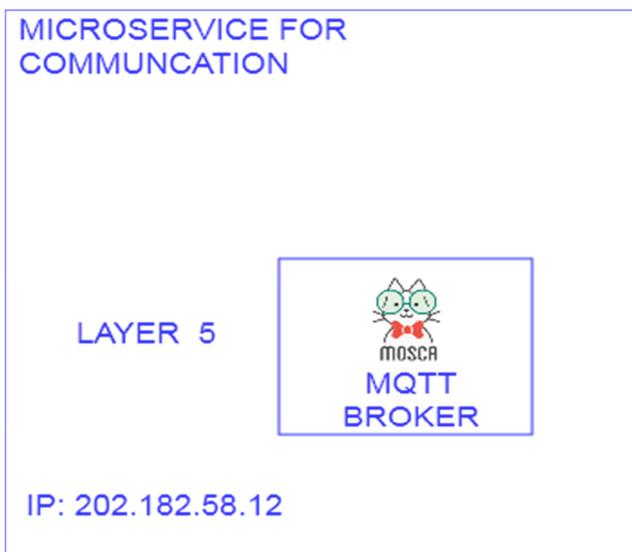

Fig.4 Block system of communication.

The substance of this system involving broker of MQTT service [20] was demonstrated in figure 4. This service is utilized as the primary service of communication through MQTT and built on the IP address 202.182.58.12:4001.

D. Microservice for Data Abstraction

This microservice is utilized for overseeing information stream in distributed computing. This framework utilized for circulated information to other microservice in SEMAR cloud computing. This data abstraction service was installed on IP address '202.182.58.13'.
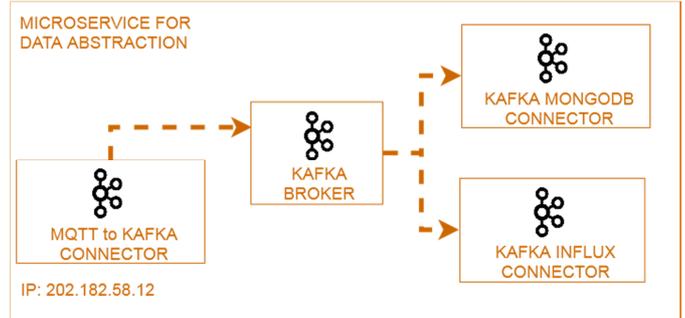

Fig.5 The design system of data abstraction.

Figure 5 demonstrates the substance of Microservice for data abstraction comprising of broker of Kafka service, MQTT to Kafka connector, Kafka MongoDB connector, and Kafka InfluxDB connector. The Kafka service is utilized as the primary communication administration with the IP address '202.182.58.13:9092'.

Information from MQTT will be sent to the broker of Kafka service utilizing MQTT to Kafka broker using the topic 'air sensor.' The Kafka broker will process the information and disseminate the information to the forecasting framework on microservice Information Examination through the point 'Kafka_airsenso.' After the process of prediction is finished, the result data will be sent back to Kafka Broker for further distribution to other subsystems with the "airsensor_analytical" Kafka topic. Kafka MongoDB connector and Kafka Influx connector will get information using the "airsensor_analytical" Kafka topic and forward it to IP address '202.182.58.11' as a data storage service and to '202.182.58.11' as a data visualization service.

Two connector applications that have been constructed utilized python language. This connector work comprises of getting information from the consumer of Kafka and sending information to different microservice. The principal connector is utilized for conveying information to the MongoDB database server utilizing a REST Application Protocol Interface and the other connector is utilized for sending information to the InfluxDB on microservice for visualization.

E. Microservice For Visualization

This part system used to visualize the information of air quality that consists of several services like data storage in time series, UI interfaces, and administrator panel (dashboard) and visualization for the public. This microservice depends on IP address 202.182.58.10 to imagine information about the air conditions obtained from

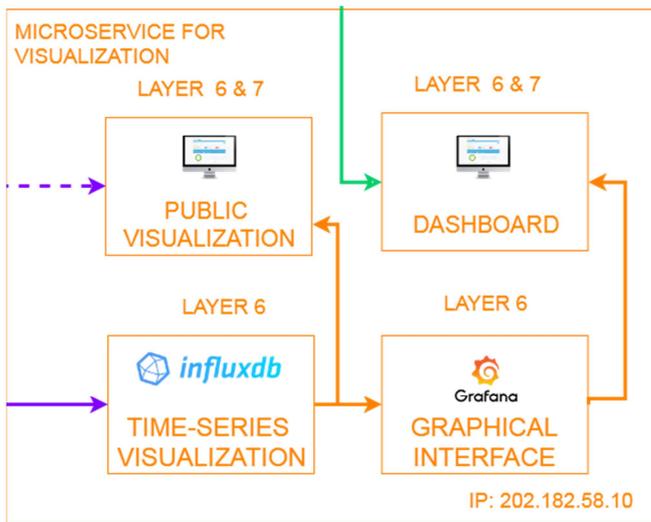VaaMSN. The substance of these systems has appeared in Figure 6.

Fig.6 The visualization design system

InfluxDB is the storage of time series data that we use in this micro-service to imagine continuous air quality information [21]. InfluxDB administration can be found at IP address '202.182.58.10:8087'. Grafana is a UI interface service that is used to represent data from influxDB in time-based groupings. This service dashboard is installed at the IP address '202.182.58.10:3002'. The last part is the public visualization and dashboard of the SEMAR cloud computing page, which is used as the UI to view air quality information.

The visualization procedure begins by sending information to InfluxDB utilizing the 'write point ()' process in Python. Information got on Influxdb is added to the 'time' section which contains the information recovery time because InfluxDB is a period arrangement database. Besides, Grafana recovers information from InfluxDB according to the parameters desired by the client and produces a graphical interface as tables, chart, and maps. Information plans are "{timestamp, sensor id, sensor_pm10, sensor_co, sensor_so2, sensor_o3, sensor_no, gps_longitude, gps_latitude, label_index}". We constructed three kinds of perception, 1) Guide show area purpose of air quality information, 2) Chart indicates information in line graph dependent on time arrangement, 3) Table show information of air present the sensors data with location, index air quality result from prediction process and also time take the data.

Public representation page recovers air quality information put away on influxDB, notwithstanding this framework has included WebSocket correspondence framework utilized for perception process in an ongoing when this page is active or opened by the user.

F. Microservice For Data Storage

This is a piece of microservices that goes about as primary storage for air quality information transmitted by VaaMSN. This subsystem comprises MongoDB No SQL and REST API web administration that keeps running on the Node.js, this REST API utilized as the principle approach to control information on MongoDB. Microservice for data storage based on IP address '202.182.58.11'. Air quality

information as JSON strings gotten by 'airsensor_analytical' Kafka topic from Kafka Brokers that will be disseminated to the MongoDB connector, by the information sent to RestAPI MongoDB at the IP address '202.182.58.11:3001' and spared to MongoDB.
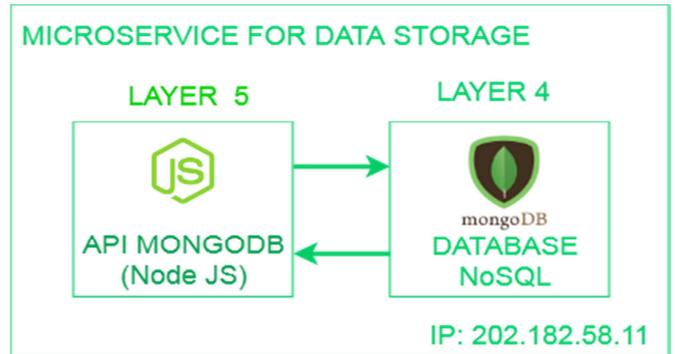

Fig.7 The data storage block system

As appeared in Figure 7, information put away on this microservice is utilized in the learning procedure to discover new information models that will be used to accelerate the information examination process. Information is likewise utilized for business analytics. The Air Quality Index got from forecast system are put away in this data with a plan comprising of the current timestamp, data information of air quality, location, air quality index label.MongoDB was picked because it has strong security, stable execution, operational ease, and best versatility off another NoSQL platform.

G. Microservice For Analytical

This microservice comprises of 1) Process of Learning, 2) Classification for Real-time prediction, and 3) Service for Business Analytical. Microservice for information examination was based on IP address "202.182.58.14". This part system procedure starts with air quality information put away in MongoDB information stockpiling to be utilized during the time spent making order models with AI calculations. The continuous arrangement process assumes a job in anticipating the air quality file from the information sent by the Kafka Broker utilizing the information show utilized has been manufactured. Business Analytics is utilized to show synopsis information at specific occasions and districts with yield as air quality conditions, this framework utilizes factual examination of information acquired.
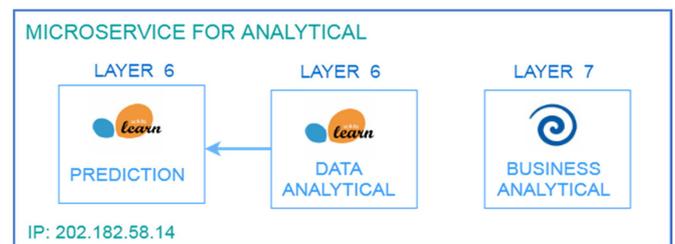

Fig.8 The analytical service block system

The present Republic of Indonesia government utilizes Air Pollution Standard Index Range (ISPU) to decide the

quality class [3] has appeared in table 1. Along these lines, the ISPU rule is utilized as a kind of perspective in this research. Parameters of the ISPU incorporate Ozone (O_3), Particulate (PM_{10}), Nitrogen dioxide (NO_2) Sulfur dioxide (SO_2), and Carbon monoxide (CO). In table 1 at 'Range' column alludes to the list detailing of air quality formulation values

TABLE I
ISPU (STANDART INDEX OF AIR POLLUTION)

A gap of Index Label	Category
0 until 50	Good
51 until 100	Average (Moderate)
101 until 199	Not Healthy
200 until 299	Very Not Healthy
more 300	Dangerous (Hazardous)

In the preparation procedure of the dataset, we utilized Scikit-learn [22] for leading the preparation procedure of the dataset. Support Vector Machine [23] and Decision Tree are utilized as arrangement calculations for this learning process. The best algorithm between the two would be chosen by looking at the aftereffects of the calculation. Support Vector Machine (SVM) is an algorithm utilized for classification and regression [24]. The classification function for decision toward becoming

$$f(w^T \phi(x) + b) = sign\left(\sum_{i=1}^{l} y_i \alpha_i K(x_i, x) + b \right) \qquad (1)$$

The decision tree is an AI calculation that utilizes tree decisions, for example, trees and the conceivable consequent impact [25], which includes the consequences of occasions, asset expenses, and utilities. The Decision Tree algorithm is a standout amongst the best classifiers while thinking about order precision; this calculation considers the arrangement work which incorporates the reliant quality (variable) given by the estimation of the autonomous characteristic (input) (variable). Gini Index is determined dependent on the equation underneath, where the likelihood of the ith class for c target classes of a given characteristic is Pi, in the meantime, Pi is the likelihood of class I [25].

$$Gini = 1 - \sum_{i=1}^{c}(P_i)^2 \qquad (2)$$

The preparation of machine learning process separates the dataset into two sections, 70% is utilized for the preparation procedure for the training process and 30% for the testing procedure, testing is utilized to decide the precision of every calculation. Order models are developed from preparing informational collections. While the estimation of the fabricated grouping model execution utilized test set. The strategy utilized is known as the holdout technique.

Scikit-learn is a library with python which is utilized for the ongoing characterization procedure and learning process. Air quality information sent from edge computing with the topic MQTT 'sensor' was sent to Kafka Broker and circulated through the Kafka topic 'Kafka air sensor,' at that point the information is changed over into JSON Array to be utilized noticeable all around quality grouping expectation process.

The aftereffect of this procedure is the forecast of the air quality file in numbers 0 to 4 speaking to the classifications in Table 1, going from good, average (moderate), not healthy, very not healthy, and dangerous hazardous). These outcomes are put away in the 'label' variable and went into the JSON Array information that was gotten beforehand so the information contains 'air quality information, location, the present time and the air quality index label.' Consolidated information is sent back through the Kafka topic "airsensor_analytical" for use by other sub-frameworks.

The procedure of Extract Transform Load (ETL) in the Pentaho apparatus [26] is utilized for business analytical frameworks in this examination. The Pentaho instrument takes air quality information as per the parameters sent by the client to the open representation interface; these parameters incorporate 'location,' 'midpoint,' 'information inclusion territory' and 'information of data time.' The following procedure is recovering information put away in MongoDB information stockpiling utilizing the 'MongoDB input' hinder by entering these parameters on the square. At that point, the chose information from MongoDB is gotten will be prepared to utilize measurable calculations with cools with the most elevated recurrence.
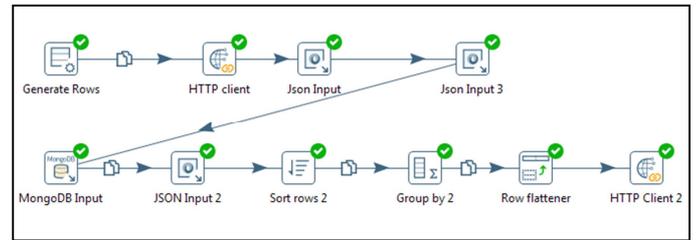

Fig.9 Pentaho block process

III. RESULT AND DISCUSSION

In the results section, we present the results of the implementation of data abstraction using Kafka on micro-service architectures in cloud computing as well as some of the experiments we have done. The experiments conducted included the transmission of sensor information to distributed computing utilizing MQTT service, information circulation in distributed computing utilizing Kafka service, imagining information progressively, showing resume information produced from business investigation forms at explicit areas and times and contrasting frameworks and the present air quality identification frameworks.

A. Communication Testing using MQTT

Communication testing using MQTT serves to test the time delay on MQTT communication that used to send the air quality information from SBC on VaaMSN to distributed computing. SBC will send data with formatting the line 'information of air quality, location and time of taking the data'. In favor of distributed computing, the information will be gotten by MQTT Brokers and sent to the microservice for data abstraction as an information reflection layer on this examination. Tests are done to compute the time expected to complete the procedure.

Figure 10 demonstrates the defer pattern for this testing perform that used 250 information acquired from the time of delivery and information taking time on the server. By

studying mediocre data, it might be seen that communication through the MQTT service execution has a typical delay of 0.04ms (40μs). According to investigations coordinated by V. Altukhov [27] showed that the best strength of suspension for a consistent system is around 600μs, so it can be said that data can be represented in real-time.
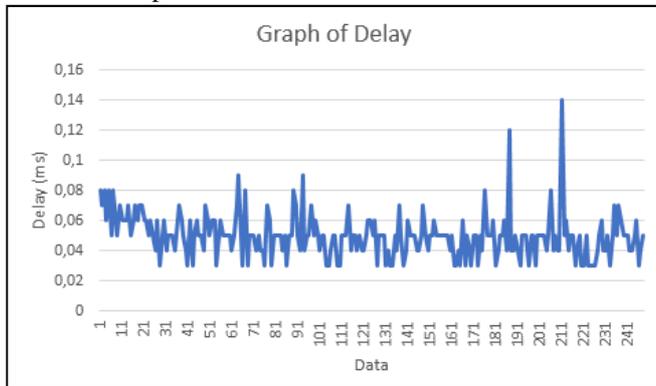

Fig.10 Graph MQTT communication delay

B. Kafka Communication Performance

Kafka communication performance serves to test the time delay on Kafka communication as information deliberation that used to appropriate information between microservices in distributed computing. Kafka broker sends a line of data comprising of 'location, air quality sensor information, and time of taking the data' to microservice of scientific. The consequence of the system will be sent back to Kafka Broker and dispersed the information to InfluxDB and MongoDB. The test is done to compute the time expected to process sending information from the broker of Kafka to the Kafka connector.
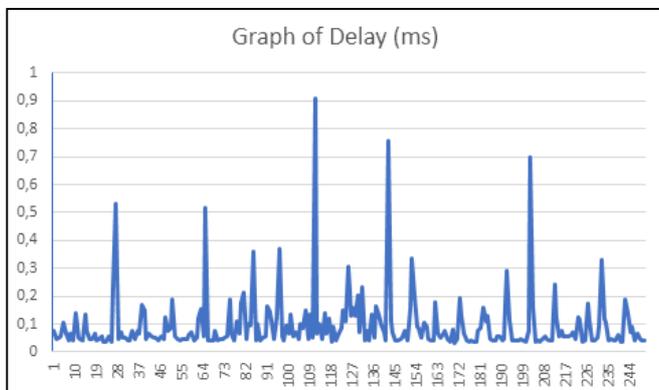

Fig.11 Chart of Kafka Communication Delay

Figure 11 illustrates the pattern of delays in the correlation of delivery time and time of receipt of air quality information. where testing information is taken several times. By ensuring normal information errors, the possibility of Kafka's communication execution has a normal delay of 0.09 ms (90μs). From this experiment present that this communication framework can be represented the data in real-time.

C. Visualization Data in Real-time

Examinations are utilized to check the reconciliation of all part frameworks in SEMAR. Examinations are done by sending sensor information from VaaMSN to the server

where delivery is carried out at regular intervals after the forecast procedure is complete, the information is sent toInfluxDB's service as a data storage in time series. The results can be seen at the page: 'http: 202.182.58.10/visual/udara'.

Fig.12 Graph display on visualization for the public.

Figure 12 demonstrates the presence of the information table on the visualization for public page; the information showed is air quality information as indicated by the ISPU rules got amid the immediate sensor explore. This test demonstrates that the framework can convey information sent by VaaMSN and show information on the visualization for public page.
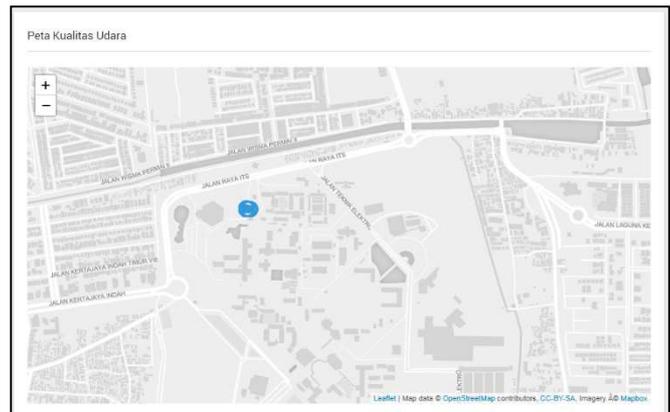

Fig.13 Map visualization for the public.

Figure 13 demonstrates a guide perspective on the air quality information during the immediate sensor test.
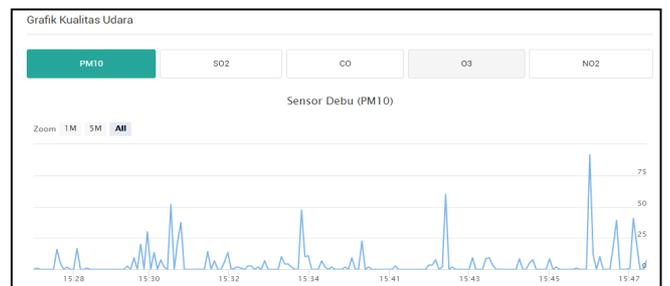

Fig.14 Chart visualization for the public.

Figure 14 demonstrates a graphical showcase of PM_{10} air quality information acquired amid a live sensor explore in line diagram dependent on time arrangement.

D. Business Analysis Processes

This test was utilized to test the framework business analytical utilizing Pentaho instruments as per those portrayed in segment Microservice for Analytical.

Fig.15 Notification on visualization for public

Figure 15 exhibits the results obtained by the customer with the parameters sent to Pentaho platform is' {longitude: 112.780892932216918, latitude: -7.28962051088299, time_start: 07/01/2018 12:00, time_end: 09/1/2018 12:00, region : 1000} '. Figure 14 exhibits the marker of air quality data obtained according to the parameters sent by the customer. The blue marker repeat, which addresses 'moderate' air quality condition, so the eventual outcomes of the business examination process are according to the present data conditions.
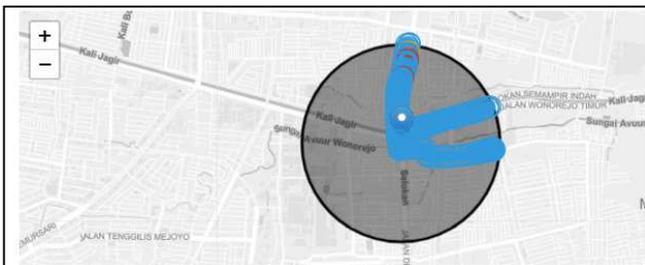

Fig.16 Bussines Analytical result area

E. Comparing This System with Other System for Air Quality

This analysis was done to consider the present air quality estimation systems used by the government with the air quality estimation methods available in this research. The test was finished by taking air quality data around the region of one of the air quality estimation stations in the city of Surabaya. In this experiment, we picked air quality information collected by air quality station in Wonorejo (Surabaya) to be compared with our systems.
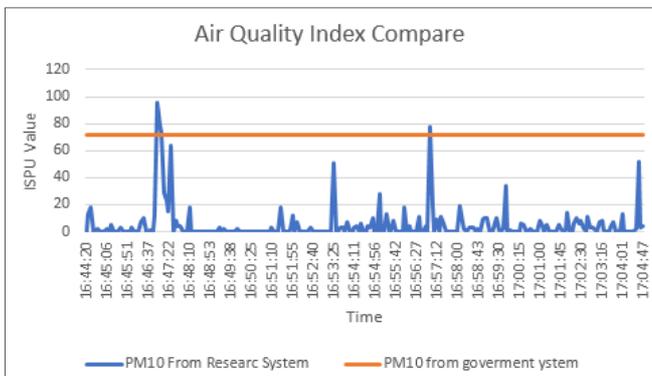

Fig.17 Chart of comparing system result based on time

This analysis was done to consider the present air quality estimation systems used by the government with the air

quality estimation methods available in this research. The test was finished by taking air quality data around the region of one of the air quality estimation stations in the city of Surabaya. In this experiment, we picked air quality information collected by air quality station in Wonorejo (Surabaya) to be compared with our systems. The data compared is the air quality data obtained by the system from the government, which is obtained at almost the same time as the data taken by our system.
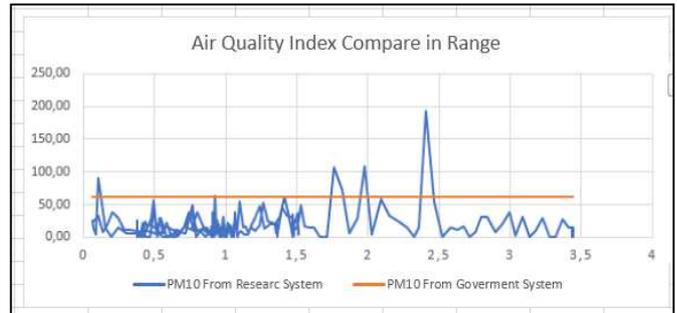

Fig.18 Chart of comparing system result based on distance

Figure 18 demonstrates the correlation of the estimation of ISPU esteem seen from the separation of estimation with the middle point being at location {- 7.312210, 112.788935} which is at air quality station in Wonorejo (Surabaya). The dark area presents a coverage area of the government air quality system with a central point on the air quality station in Wonorejo (Surabaya) at the location {- 7.312210, 112.788935} and its range is about 5 km from the center. This area coverage area by regulations issued by the government regarding the distance between air quality measurement systems. Meanwhile, air quality measurements using the VaaMSN system are displayed in the form of colored dots, where the air quality conditions are in the "good" category so that they are blue. The results of this experiment indicate that the quality observation framework used in this system presents definite information from air quality data in a particular region, different from the estimation framework used by the government where the information displayed has a very wide area and not point by point for certain regions.

Fig.19 Maps of comparing system result

IV. Conclusion

The examination to check the data abstraction layer can be implemented to the cloud computing service that was built in a microservice architecture. The communication testing from gadget sensors to distributed computing through MQTT it has a normal delay of 0.04ms (40μs) and Kafka communication suspension reaches up to 0.09 ms (90μs), but this framework can be said to be constant, so it tends to say that the air quality data shown by this system can represent real environmental conditions at a particular time. In visualization, testing shows that the microservice framework for visualization can present information that sent by VaaMSN and the data can be obtained for to visualized on the IP address 202.182.58.10. The examination results show that the air quality inspection framework in this system can be an elective technique for continuously estimating the quality of air condition so that the government can utilize this system.

Acknowledgment

References

[1] Liu, Jen-Hao, et al. "Developed urban air quality monitoring system based on wireless sensor networks." Sensing technology (ICT), *2011 fifth international conference on. IEEE*, 2011.

[2] Satya Widya Yudha. Air Pollution in Indonesia Challenges and Imperatives for Change. *Singapore: The National Bureau of Asian Research*.2016.

[3] Head of Environmental Impact Control Agency, Perhitungan Dan Pelaporan Serta Informasi Indeks Standard Pencemar Udara [Calculation And Reporting As well as Information on Air Pollution Standard Index]. Indonesia: of Environmental Impact Control Agency.1997.

[4] Sukaridhoto, S., Pramadihanto, D., Alif, M., Yuwono, A., & Funabiki, N. (2015, May). "A design of radio-controlled submarine modification for river water quality monitoring." *In Intelligent Technology and Its Applications (ISITIA)*, 2015 International Seminar on (pp. 75-80). IEEE.

[5] Berlian, M. H., Sahputra, T. E. R., Ardi, B. J. W., Dzatmika, L. W., Besari, A. R. A., Sudibyo, R. W., & Sukaridhoto, S. (2016, September). "Design and implementation of smart environment monitoring and analytics in real-time system framework based on internet of underwater things and big data." *In Electronics Symposium (IES)*, 2016 International (pp. 403-408). IEEE.

[6] Sukaridhoto, Sritrusta, et al. "Design and development of a portable, low-cost COTS-based water quality monitoring system." *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*. IEEE, 2016.

[7] Arridha, R., Sukaridhoto, S., Pramadihanto, D., & Funabiki, N. (2017). "Classification extension based on IoT-big data analytic for smart environment monitoring and analytic in the real-time system." *International Journal of Space-Based and Situated Computing*, 7(2), 82-93.

[8] Abdillah, Abid Famasya, et al. "Design and development of low-cost coral monitoring system for shallow water based on internet of underwater things." *Journal of Telecommunication, Electronic and Computer Engineering (JTEC)* 9.2-5 (2017): 97-101.

[9] *Panduman, Y. Y. F., Besari, A.R.A.,* Sukaridhoto, S., *Budiarti, R. P. N., Sudibyo, R. W.,* & Funabiki, N., (2018). "Implementation of Integration VaaMSN and SEMAR for Wide-Coverage Air Quality Monitoring." *TELKOMNIKA*, 16(6), 2630- 2642.

[10] Sukaridhoto, Sritrusta, Achmad Subhan KHalilullah, and Dadet Pramadihanto. "Further Investigation on Building and Benchmarking A Low Power Embedded Cluster for Education." *IPTEK Journal of Proceedings Series* 1.1 (2015).

[11] Sukaridhoto, S., Sasaki, Y., Ito, K., & Aoki, T. "Development of a compact cluster with embedded CPUs." *Proceeding of the Sixth Industrial Electronic Seminar*. 2004.

[12] Sill, Alan. "The design and architecture of microservices." *IEEE Cloud Computing 3.5* (2016): 76-80.

[13] Meehan, John, et al. "Data Ingestion for the Connected World." *CIDR*. 2017.

[14] Sun, Long, Yan Li, and Raheel Ahmed Memon. "An open IoT framework based on a microservices architecture." *China Communications 14.*2 (2017): 154-162.

[15] Villamizar, Mario, et al. "Infrastructure cost comparison of running web applications in the cloud using AWS lambda and monolithic and microservice architectures." Cluster, Cloud, and Grid Computing (CCGrid), *2016 16th IEEE/ACM International Symposium on. IEEE*, 2016.

[16] Shi, W., Cao, J., Zhang, Q., Li, Y., & Xu, L. (2016). "Edge computing: Vision and challenges." *IEEE Internet of Things Journal*, 3(5), 637-646.

[17] Wiska, Rindra et al. "Big sensor-generated data streaming using Kafka and impala for data storage in wireless sensor network for co 2 monitoring." *2016 International Workshop on Big Data and Information Security (IWBIS)*. IEEE, 2016.

[18] Phala, Kgoputjo Simon Elvis, Anuj Kumar, and Gerhard P. Hancke. "Air quality monitoring system based on ISO/IEC/IEEE 21451 standards." *IEEE Sensors Journal 16.*12 (2016): 5037-5045.

[19] Kang, L., Poslad, S., Wang, W., Li, X., Zhang, Y., & Wang, C. (2016, September). "A public transport bus as a flexible mobile smart environment sensing platform for IoT." *In Intelligent Environments (IE), 2016 12th International Conference on* (pp. 1-8). IEEE.

[20] Hunkeler, U., Truong, H. L., & Stanford-Clark, A. (2008, January). "MQTT-S—A publish/subscribe protocol for Wireless Sensor Networks." *In Communication systems software and middleware and workshops*, 2008. comsware 2008. 3rd international conference on (pp. 791-798). IEEE.

[21] Leighton, Benjamin, et al. "A best of both worlds approach to complex, efficient, time series data delivery." *International Symposium on Environmental Software Systems*. Springer, Cham, 2015.

[22] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." *Journal of machine learning research* 12.Oct (2011): 2825-2830.

[23] Cortes, Corinna, and Vladimir Vapnik. "Support-vector networks." *Machine learning* 20.3 (1995): 273-297.

[24] Nieto, PJ García, et al. "A SVM-based regression model to study the air quality at local scale in Oviedo urban area (Northern Spain): A case study." *Applied Mathematics and Computation*219.17 (2013): 8923-8937.

[25] Breiman, Leo. Classification and regression trees. Routledge, 2017.

[26] Altukhov, V., and E. Chemeritskiy. "*On real-time delay monitoring in software-defined networks.*" Science and Technology Conference (Modern Networking Technologies) MoNeTeC, 2014 First International. IEEE, 2014.

[27] Zappatore, Marco, et al. "Mobile Crowd Sensing-based noise monitoring as a way to improve learning quality on acoustics." *Interactive Mobile Communication Technologies and Learning (IMCL), 2015 International Conference on*. IEEE, 2015.