ILKOM Jurnal Ilmiah Vol. 15, No. 2, August 2023, pp. 335-343
Accredited 2nd by RISTEKBRIN No. 200/M/KPT/2020; E-ISSN 2548-7779 | P-ISSN 2087-1716

Research Article

Development of ETL (Extract, Transform and Load) Module in Indonesian Agricultural Commodities OLAP System

Aditia Yudhistira [a,1,*]; Imas Sukaesih Sitanggang [a,2]; Hari Agung Adrianto [a,3]
[a] *IPB University, Jl. Raya Dramaga, Babakan, Kec. Dramaga, Kab. Bogor, Jawa Barat, 16680 Indonesia*
[1] *aditia28aditia@apps.ipb.ac.id;* [2] *imas.sitanggang@apps.ipb.ac.id;* [3] *agung@apps.ipb.ac.id*
* *Corresponding author*

Article history: Received June 20, 2023; Revised July 10, 2023; Accepted July 10, 2023; Available online August 16, 2023

Abstract

The SOLAP system for Indonesian Agricultural Commodities is a successful development based on previous studies. Agricultural commodity data are managed in a data warehouse with a galactic schema which has 7 fact tables, namely cut flower horticulture, ornamental plant horticulture, horticulture, food crops, plantation, livestock population, and livestock production, as well as 3 dimensional tables, namely location, time, and commodity. The results of SOLAP operations on the system can be visualized in the form of crosstabs, graphs and maps. The system uses a web platform so that it can be accessed by the public. However, the SOLAP system could not update data in real time. This study aims to develop a data warehouse for Indonesian agricultural commodities SOLAP in real time by creating a scraping system. This study has succeeded in developing a data warehouse in real time on the indonesian agricultural commodity SOLAP system by creating a real time scraping system that is applied to the SOLAP server and has succeeded in making the ETL process run in real time on the SOLAP server and optimizing polygon-based spatial data visualization using the Douglas-Peucker. This study has also carried out functional testing of OLAP features and functions on the Indonesian Agricultural Commodity SOLAP system using the black box testing method. The results of this study provide accurate and real-time data on the SOLAP of Indonesian Agricultural Commodities, with the results of SOLAP feature testing achieving 100 percent pass and the data conformity test results of OLAP function as expected. In addition, the results of this study make it possible to automatically update the data according to a predetermined schedule to provide real-time information.

Keywords: Agricultural Commodities; Douglas-Peucker; Functional Testing; Real-Time; SOLAP.

Introduction

Indonesia is an agricultural country with abundant natural resource advantages and a geographical location that is considered very strategic, so most of the Indonesian population has their livelihood in the agricultural sector. According to data from the *Badan Pusat Statistik* (BPS) for August 2022, about 38.7 million people, or 29.59% of Indonesia's total population work in the agricultural sector (BPS 2022). From these data, it shows that the agricultural sector has a role in improving the economy and meeting food needs, especially for the ministry of agriculture in managing Indonesia's agricultural commodity data. Potential areas can be identified based on harvested area, production, and productivity in various sub-sectors such as horticulture, food crops, plantations, and livestock. The ministry of agriculture has facilitated the presentation of Indonesian agricultural commodity data, which can be accessed on the website https://bdsp2.pertanian.go.id/bdsp/ stated that the commodity data consists of several sub-sectors, namely horticulture, plantations, livestock and food crops in an annual period. so that data growth will produce a very large dataset [1] have successfully stored Indonesian agricultural commodity data in a data warehouse. A data warehouse is a long-term storage space for storing data collections from various sources in a scheme to assist users in making decisions. As in this study, decision-making to obtain knowledge from data stored in data warehouses used Online Analytical Processing (OLAP) [2], OLAP is used to explore the data warehouse in detail and uses a multidimensional data model [3]. OLAP aims to find the desired data quickly from large amounts of data. OLAP that supports operations on spatial data is called Spatial Online Analytical Processing (SOLAP).

A study related to SOLAP for Indonesian agricultural commodities has been successfully developed [1]. and the results of SOLAP function on the system can be visualized in the form of cross-tabs, graphs and maps. The system uses a web platform to make it accessible to the public. However, this system is limited by the large size of the GeoJSON

http://dx.doi.org/10.33096/ilkom.v15i2.1758.335-343

file, which reaches 91.30 MB. The large size of the GeoJSON file is due to the large number of coordinates that make up the polygons. The optimization of polygon-based spatial data in the visualization module of the indonesian agricultural commodities SOLAP system has been successfully carried out by [4]. In this study, the optimization of polygon-based spatial data in the visualization module of the Indonesian agricultural commodities SOLAP system was re-optimized using the Douglas-Peuckker algorithm because the spatial data in the Geospatial Information Agency contained the addition of new provinces and districts. Spatial data can be accessed on the website https://tanahair.indonesia.go.id/portal-web, while commodity data can be accessed on the website of the Ministry of agriculture of the Republic of Indonesia https://bdsp2.pertanian.go.id/bdsp/ (June 13, 2023). Agricultural commodity data is managed in a data warehouse with a galactic schema, which has 7 fact tables namely cut flower horticulture, ornamental plant horticulture, horticulture, food crops, plantation, livestock population, and livestock production, as well as 3 dimensional tables namely location, time, and commodity [5].

The creation of an Extract, Transform and Load (ETL) module for a data warehouse for Indonesian agricultural commodities was also the basis for this study in developing an ETL module to store update data and information in SOLAP for Indonesian agricultural commodities [5]. Therefore, spatial data were used, namely digital district boundary maps and non-spatial data obtained from an agricultural statistics database. The ETL modeling in this study was done using Business Process Modeling Notation (BPMN) for conceptual modeling and architectural graphs for logical design. The development of the ETL module was successfully carried out, the system test also shows that the ETL module works properly in updating the data warehouse integrated with SOLAP [5]. Although the test was declared successful and adequate, there were still differences between the data sources of the Ministry of Agriculture and the SOLAP data warehouse after the reverse test [6]. Previously, a usability test was also conducted for the Indonesian Agricultural Commodity SOLAP by [7] and useful functional tests to ensure that the data in the data warehouse is correctly captured from the data sources by [6]. In this ETL testing phase, transformations, cleansing and format changes are performed on the data, resulting in data changes. The ETL implementation is re-tested and improved to achieve data conformity in the data warehouse [6]. Although several tests have been conducted, the Indonesian agricultural commodities SOLAP has not been updated in real time. The importance of processing real-time data in a data warehouse is so that the SOLAP of Indonesian agricultural commodities can be reused by farmers, agricultural extension workers, researchers, and other parties involved in the development of the agricultural sector to support decision making in determining the appropriate commodities to be cultivated in an area..

Data warehouse testing is a very important stage in data warehouse development because analysis and decisions are made based on information generated from the data warehouse. Data warehouse testing was also performed. With three types of tests, namely fact tests, which verify the previous workloads taken by users during the needs analysis, hierarchy tests, which measure the extent to which the functional dependencies represented by the hierarchy in the conceptual schema are actually verified on the source data, and finally the conformity test, which aims to assess how well the hierarchies have been designed. The tests have successfully performed statistical tests on OLAP cubes to compare degrees of disease using statistical techniques and machine learning. The technique can produce reliable results with large and small subsets. ETL test has also been performed by [8] to determine the quality of a data warehouse, ETL test is an important and vital phase during data warehouse testing, this phase affects data quality.

Based on the limitations of previous studies, this study aims to develop a real-time data warehouse on the SOLAP system for Indonesian agricultural commodities by creating a real-time scraping system that is applied to the SOLAP server, and next goal is to make the ETL process that run in real time on the SOLAP server and perform optimization visualization on polygon-based spatial using the Douglas-Peucker algorithm. A scraping system is a program that crawls the World Wide Web in a methodical, automatic and organized way. Other names for scraping systems are ants, automatic indexers, bots, web spiders or web robots [9]. Scraping systems can retrieve data by providing a timer for the system to automatically retrieve data every few minutes [10]. The traditional ETL process is run periodically at certain intervals, such as monthly or weekly. This approach has been used by organizations for many years [8].

In this study, the tests focused on functional testing of OLAP features and functions in the Indonesian agricultural commodity SOLAP system. Functional testing of OLAP features and functions uses the black box testing method, which aims to ensure that the data input in each form matches the output generated by the SOLAP system, and to ensure that the OLAP function is consistent with the source data and data warehouse.

Method

This study was conducted through four steps, starting with the creation of a scraping system for real time data retrieval from the Ministry of Agriculture website, simplifying the polygons in all districts/cities in Indonesia, testing the SOLAP system for Indonesian agricultural commodities using black box testing, and analyzing the test results. The research stages are shown in **Figure 1**.

A. Create a scraping system to retrieve data in real time

This scraping system functions to retrieve data on the website of the ministry of agriculture of the republic of Indonesia in real time. At this stage, the system is created to optimize the stages and levels of the scraping system to achieve increased efficiency and speed in real time [11]. In previous studies, the scraping process has not been carried out on the server. The development of this study is the scraping process carried out on the server to obtain data from the website https://bdsp2.pertanian.go.id/bdsp/. To create a scraping system, researchers use python tools, this framework from scrappy is an open source and collaborative way to extract the data needed from websites in a fast, simple and repeatable process [12].

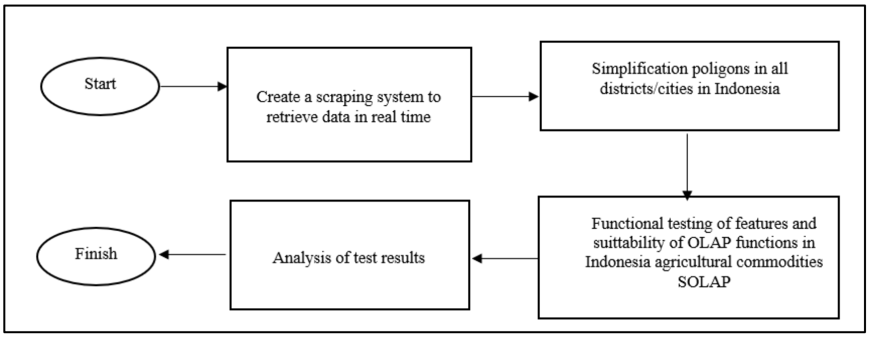


Figure 1. Research stages

B. Simplification poligons in all districts/cities in Indonesia

This stage is a simplification process performed on polygons for all administrative boundaries in Indonesia at the district/city level. The algorithm used is Douglas-Peucker. The idea of this algorithm is to reduce the number of coordinates that make up the polygon [13]. This step was done recently because there were additional provinces and districts in Indonesia [4]. Have done a comparison of algorithms simplification on vector data. The algorithm used in the study consists of Nth point, perpendicular distance, Reumann-Witkam, Opheim's, Visvalingam-Whyatt, Douglas-Peucker, distance between points, Lang, and Zhao & Saalfelds. The vector data used in this study is the polyline coastline of Hong Kong Island and Kowloon. Both areas are located in China. The number of coordinates contained in the two coastlines is as many as 500. Comparison of simplification results is measured by observing the accuracy of position coordinates and time used. The results of the study it is found that the Douglas-Peucker algorithm produces the most accuracy better when compared to other algorithms. Then the second best algorithm is the Zhao-Saafeld algorithm. The weakness of the Douglas-Peucker algorithm in this study is that it takes longer to simplify compared to other algorithms.

Subsequent studies using the Douglas-Peucker algorithm are [14]. In this study, a simplification algorithm is used which consists of four algorithms, namely Douglas-Peucker, Progressive, Angle, and Circle. The simplification of this study was carried out on railway data in China in the 1980s. The vector data of the railway tracks consists of 504 polylines with 7296 coordinates. The results of this research show that the compression ratio between the original data and the simplified data of the Douglas-Peucker algorithm reaches 85%. Meanwhile, the compression ratio of other algorithms is not better when compared with the compression ratio of Douglas-Peucker algorithm.

C. Functional testing of features and suitability of OLAP functions in Indonesia agricultural commodities SOLAP

In this testing stage, black box is used with the equivalence partition technique. Black box testing is a test based on application details such as the appearance of the application and the suitability of the function flow and the accuracy of the OLAP function in the Indonesian Agricultural Commodity SOLAP. Equivalence partition is a testing technique that is based on input values and output values generated in a function [15].

The testing of the function test of the Indonesian Agricultural Commodities SOLAP system was conducted following the black box testing approach. The test cases consist of 11 descriptions and 40 scenarios. An example of a SOLAP feature system test case is shown in **Table 3**. The test case has a positive type case, which means that the test scenario tested is to test the system in dealing with unexpected situations or unexpected problems, and vice versa for positive cases.

D. Analysis of test results

In this stage, each test result of the SOLAP testing stage for Indonesian Agricultural Commodities is analyzed. If there is a function of Indonesian Agricultural Commodity SOLAP system that is not appropriate, an error report will be generated and the cause of the error in development will be explained.

Results and Discussion

A. Create a scraping system to retrieve data in real time

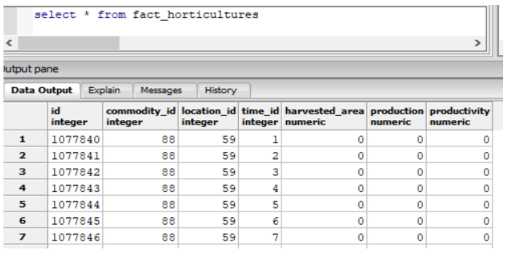
This system is built using a Python application. Which functions to collect data on Indonesian agricultural commodities from the website https://bdsp2.pertanian.go.id/bdsp/ in real time. This scraping system has been successfully implemented on the Indonesian Agricultural Commodities SOLAP server by running the scraping process and ETL automatically every 6 months. This is done because the process of updating data on the Indonesian Ministry of Agriculture website occurs every 6 months. The scraping process is run using a cronjob that is interpreted to run the scraping process and ETL at 0 hours, 0 minutes, first (1) date every 6 months (*/6).

There is a problem when scraping, namely scraping will still download existing data again, this makes the ETL process take a long time because it does the ETL process of all existing data. So, the researcher made an additional scraping script so that the downloaded data, if it is already in the database, will not be downloaded again. The following data was successfully collected by scraping the source data from 1970 to 2022 from the national and county level. With the .xls extension, the scraping results consist of 6,024 files. The display results of the source data scraping are shown in **Table 1**.

Table 1. Source data view

Subsector		Horticulture				
Indicator		Harvested area				
Level		Regency				
Province		Aceh				
Regency		Kab. Aceh Singkil				
Year		1970 - 2022				
No	**Commodity**	**Metric**	**1970**	**1971**	**1972**	**1973**
1	Adenium	m2	0	0	0	0
2	Aglaonema	m2	0	0	0	0
3	Avocado	Ha	0	0	0	0
4	Orchid	m2	0	0	0	0
5	Grape	Ha	0	0	0	0
6	Anthurium Flower	m2	0	0	0	0
7	Anthurium Leaf	m2	0	0	0	0
8	Carnation	m2	0	0	0	0
9	Apple	Ha	0	0	0	0
10	Spring Onion	Ha	0	0	0	0
11	Red Onion	Ha	0	0	0	0
12	Garlic	Ha	0	0	0	0
13	Spinach	Ha	0	0	0	0
14	Star Fruit	Ha	0	0	0	0

The data generated in the form of a data source is fed into the target data, which is the data warehouse for SOLAP of Indonesian Agricultural Commodities. The data warehouse used for SOLAP of Indonesian Agricultural Commodities is shown in **Figure 2**, and the data implemented in the data warehouse is shown in **Figure 3**.


```
select * from fact_horticultures
```

	id integer	commodity_id integer	location_id integer	time_id integer	harvested_area numeric	production numeric	productivity numeric
1	1077840	88	59	1	0	0	0
2	1077841	88	59	2	0	0	0
3	1077842	88	59	3	0	0	0
4	1077843	88	59	4	0	0	0
5	1077844	88	59	5	0	0	0
6	1077845	88	59	6	0	0	0
7	1077846	88	59	7	0	0	0

Figure 2. The example of fact table

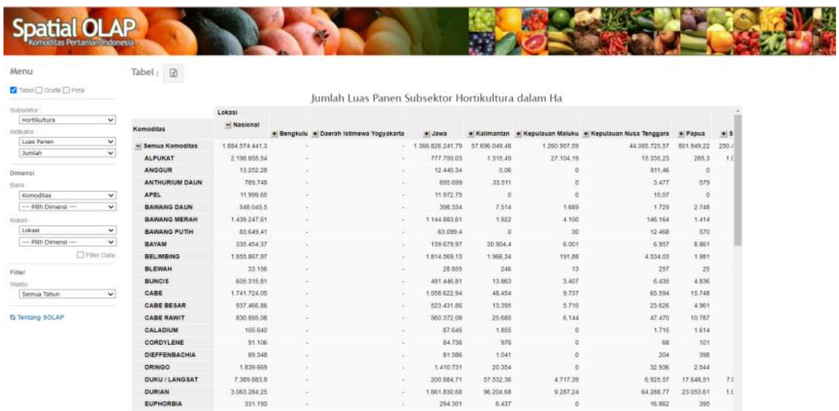


Figure 3. Crosstab in SOLAP for Indonesia Agricultural Commodities

B. Simplification polygons of districts/cities in Indonesia

This analysis uses simplified polygons with optimal threshold values, only available at the district/city level. To achieve this, we form polygons at the provincial and island levels for further analysis. To construct the provincial-level polygons, we combine all district/city polygons within each province. Similarly, we create the island-level polygons by combining all provincial-level polygons within each island. **Table 2** compares the coordinate count before and after simplification.

Table 2. Comparison of the number of coordinates that make up the polygon between before and after simplification

Number of Coordinates			
Polygon Levels	***Before simplification***	***Douglass-Peucker***	***Percentage reduction***
Regency/city	2,031,777	123,953	93,8 %
Province	1,528,078	90,475	94 %
Island	1,462,458	85,130	94%

The following is a comparison of the GeoJSON file sizes before and after the simplification shown in Table 3

Table 3. Comparison of GeoJSON file size between before and after simplification

GeoJSON file size (MB)			
Polygon Levels	***Before simplification***	***Douglass-Peucker***	***Percentage reduction***
Regency/city	113.679	7.532	93,3 %
Province	70.241	5.008	92,8 %
Island	64.536	4.754	92,2 %

The size of the GeoJSON file is determined by the number of polygon component coordinates. The shape of the polygon changes when the number of coordinates on the simplified polygon is reduced from what it was before simplification. In addition, the area of the polygon changes. Below is an example of the resulting simplification: The simplification of polygons in DKI Jakarta Province compared to the results after implementation in SOLAP can be seen in **Figure 4** and **Figure 5**, respectively.

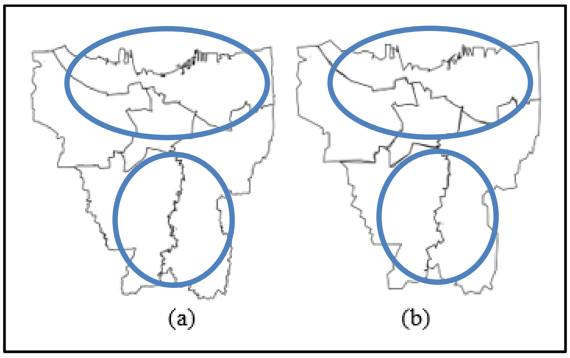


Figure 4. Comparison of simplification of polygons in DKI Jakarta province (a) before simplification (b) Douglas-Peucker with a threshold of 0.00251.

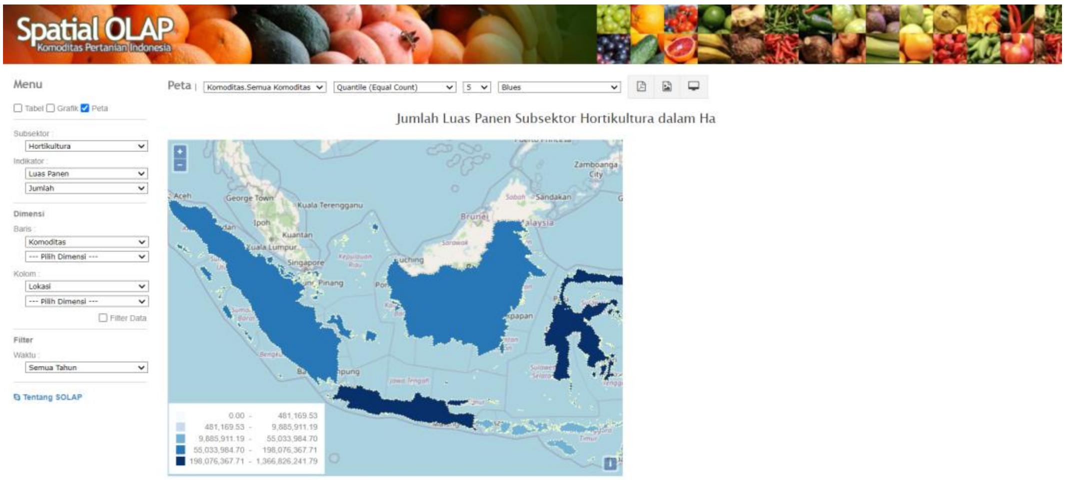


Figure 5 Simplified polygons in SOLAP

C. Testing

There are two test cases: the SOLAP feature test case and the OLAP function data suitability test. The feature test of the Indonesian Agricultural Commodities SOLAP system was tested using the black box testing approach. The 11 descriptions and 40 scenarios constitute the test cases. Table 4 shows the results of the tests performed on the Indonesian Agricultural Commodities SOLAP system, where the output is labeled as either "passed" or "failed".

Table 4. The results of testing the Indonesian Agricultural Commodities SOLAP system feature test cases

Description	ID	Scenario	Type	Expected Result	Priority	Result Case
Users access the Indonesian Agricultural Commodities SOLAP website for horticulture subsector data and data filter results.	2	Users select the value "Horticulture" in the subsector parameter, then review the data filter, select the commodity dimensions and location dimensions to filter, and then press the "right arrow" button.	Positive	Users can move the selected commodity and location dimensions to the selected data column to the right of the commodity dimension.	Medium	Passed
Users see the total area of the plantation subsector based on the table.	11	Users select the value "Plantation" in the subsector parameter, deselect other parameters, and select the values "Area" and "Quantity" in the indicator parameter. Selecting Dimension, Commodity Row and Location Column with Time Filter.	Negative	Users cannot see any display because none of the parameters are checked.	Medium	Passed
Users download and print total area results in graphical form	17	Users want to download an out-of-area amount in graphical form with a JPEG file	Positive	Users see the results of downloads outside of the area in graphical form with JPEG files.	High	Passed

An example of a feature tested in the Indonesian Agricultural Commodity SOLAP system is shown in Figure 6.

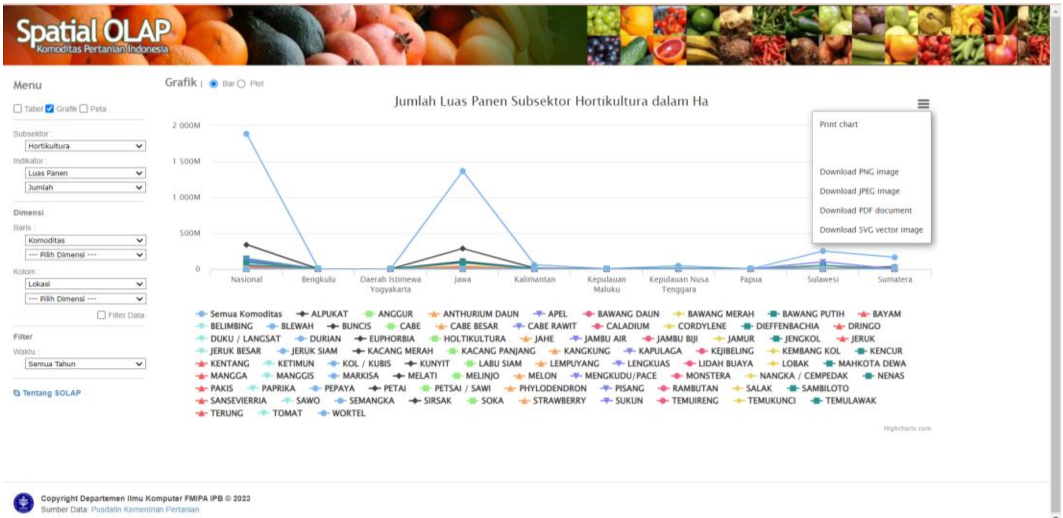


Figure 6 Example of a tested SOLAP graphic and download function

The results of testing the suitability test cases for the OLAP function data on the Indonesian Agricultural Commodity SOLAP system, can be seen in Table 5 with the output of the test results, namely passed or failed. In testing the OLAP function test cases, there are 4 descriptions and 4 test scenarios on the Indonesian Agricultural

Commodity SOLAP system and data warehouse tested, the success rate is 100%. The results of OLAP function test case testing are shown in **Table 5**.

Table 5. Testing results of OLAP function test cases

ID	Description	Table Testing	Testing atributes	Result Case
1	Summary	fact_horticultures	harvested_area	Passed
2	Average	fact_livestock_population	production	Passed
3	Maximum	fact_food_crops	Productivity	Passed
4	Minimum	fact_Plantations	acreage	Passed

The OLAP function tested on the Indonesian Agricultural Commodity SOLAP system is shown in **Figure 7**.

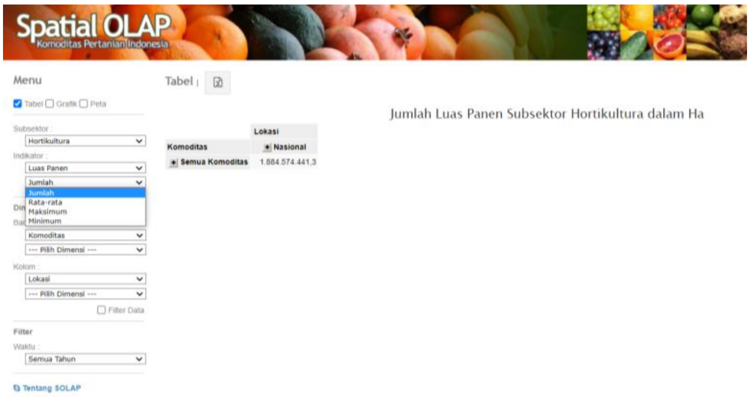


Figure 7 OLAP function

The ID 1 test case shows the results of the OLAP summary function of the harvested area of the horticulture sub-sector in Simeulue District, Aceh Province in 2016. It can be seen in **Table 6** that the total harvested area of the horticulture sub-sector in Simeulue District, Aceh Province is 260.47, which has the same number of harvested areas between the data source, data warehouse and the Indonesian Agricultural Commodity SOLAP system, which can be seen in **Figure 8** and **Figure 9**.

Table 6. Horticulture subsector data source

Subsector		Horticulture	
Indicator		Harvested area	
Level		Regency	
Province		Aceh	
Regency		Kab. Simeulue	
Tahun		1970 - 2022	
			Sum
			260.47
No	**Commodity**	**Unit**	**2016**
1	Adenium	m2	0
2	Aglaonema	m2	0
3	Avocado	Ha	0.57
4	Orchid	m2	0
5	Grape	Ha	0
6	Anthurium Flower	m2	0
7	Anthurium Leaf	m2	0
8	Carnation	m2	0
9	Apple	Ha	0
10	Spring Onion	Ha	0
11	Red Onion	Ha	0
12	Garlic	Ha	0
13	Spinach	Ha	0
14	Star Fruit	Ha	0.19
15	Adenium	Ha	0

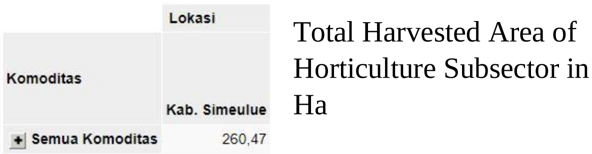


Figure 8. Results Douglas-Peucker;

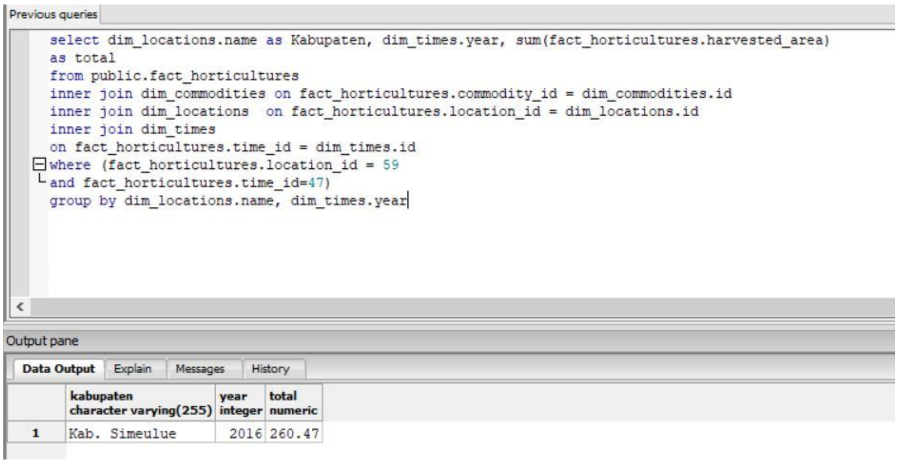


Figure 9. Results query in the data warehouse

D. Analysis of test results

At this stage, an analysis of feature testing and OLAP function data suitability testing is performed on the Indonesian Agricultural Commodity SOLAP system. Based on the test results with feature test cases and suitability test cases of OLAP function data above, the test results can be analyzed, namely

1) Testing the characteristics of the Indonesian Agricultural Commodity SOLAP system

Based on the results of testing all the features in the Indonesian Agricultural Commodity SOLAP system using 40 test scenarios, with positive 32 cases and negative 8 cases test types and there is priority testing which indicates a test case priority scale, namely high 29 cases, medium 9 cases, low 2 cases and has a 100% passed test result which means that all the features in the Indonesian Agricultural Commodity SOLAP system can function properly and can be used properly.

2) Testing the Compatibility of OLAP Function Data of Indonesian Agricultural Commodity SOLAP System

Based on the results of testing the OLAP function, testing and comparing the output has been carried out between data sources, data warehouse queries, and OLAP functions in the Indonesian Agricultural Commodity SOLAP system, including the summary, average, maximum and minimum functions, have the same results from these functions, which are declared 100% pass.

Conclusion

This research has developed from previous research, namely, the process of data collection which is done in real-time on the server, and the simplification of polygons on all administrative boundaries in Indonesia which has been implemented in the Indonesian agricultural commodity OLAP system. In addition, this study has also tested the SOLAP web-based application with black box testing. The test case is divided into 2 parts, namely the feature test and the compatibility test of OLAP function data by comparing the data source, data warehouse and OLAP functions in the Indonesian agricultural commodity SOLAP system. This test aims to evaluate the functional requirements for the performance of the designed OLAP application process. The test results show that the Indonesian agricultural commodity SOLAP application has passed the presentation 100%, which means that the built Indonesian agricultural commodity SOLAP website can help to analyze the data and can provide information according to the facts presented both in terms of features, functions and visualization.

A real-time data warehouse is expected to assist in developing the agricultural sector to support decision-making in determining the appropriate commodities to be cultivated in an area quickly and accurately. Suggestions for further research are expected to carry out non-functional testing of the Indonesian Agricultural Commodity SOLAP system to measure security, efficiency, system performance, and data integrity.

References

[1] A. I. Putri and I. S. Sitanggang, "Data Cubes Integration in Spatial OLAP for Agricultural Commodities," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 58, no. 1, 2017, doi: 10.1088/1755-1315/58/1/012034.

[2] N. P. M. Ardiyanti, A. S. Kusuma, and I. K. B. Sandika, "OLAP (On-Line Analytical Processing) Implementation for Web-Based Sales Transaction Processing (Case Study: Lilola Boutique)," *J. Teknol. Inf. dan Komput.*, vol. 4, no. 1, pp. 80–89, 2018, doi: 10.36002/jutik.v4i1.398.

[3] M. Hilman and D. Djamaludin, "Factor Analysis of ETL Process Optimization in Data Warehouses to Support Management Decision Making with Business Intelligence," *Fakt. Exacta*, vol. 11, no. 1, p. 24, 2018, doi: 10.30998/faktorexacta.v11i1.2325.

[4] A. R. Ginanjar, I. S. Sitanggang, and Annisa, "Optimization of spatial visualization module in solap for indonesian agricultural commodities," *Int. J. Geoinformatics*, vol. 16, no. 1, pp. 9–19, 2020.

[5] I. S. Sitanggang, R. Trisminingsih, F. Fuady, and H. Khotimah, "Extract, Transform, Load Module in SOLAP for Indonesia Agricultural Commodity," *3rd Int. Conf. Sustain. Inf. Eng. Technol. SIET 2018 - Proc.*, pp. 101–105, 2018, doi: 10.1109/SIET.2018.8693152.

[6] F. Ghita and R. Trisminingsih, "SOLAP Data Warehouse Testing for Indonesian Agricultural Commodities," *J. Ilmu Komput. dan Agri-Informatika*, vol. 8, no. 1, pp. 42–56, 2021, doi: 10.29244/jika.8.1.42-56.

[7] I. S Sitanggang, R. Trisminingsih, H. Khotimah, and M. Syukur, "Usability testing of SOLAP for Indonesia agricultural commodity," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 299, no. 1, 2019, doi: 10.1088/1755-1315/299/1/012054.

[8] W. S. Fana, R. Sovia, R. Permana, and M. A. Islam, "Data Warehouse Design With ETL Method (Extract, Transform, And Load) for Company Information Centre," *Int. J. Artif. Intell. Res.*, vol. 5, no. 2, pp. 132–137, 2021, doi: 10.29099/ijair.v5i2.215.

[9] A. Rahma Salsabila, M. Daffa, M. Kandias, H. Maulana, and E. D. Wahyuni, "Proceedings Of The National Seminar On Technology And Information Systems (Sitasi) 2022 Implementation Of Web Scraping Technique To Display Data English Premier League Team Implementation Of Web Scraping Technique To Showing English Premier League Team Team Data," no. September, pp. 10–11, 2022, [Online]. Available: http://sitasi.upnjatim.ac.id/%7C40.

[10] Fatmasari, Y. N. Kunang, and S. D. Purnamasari, "Web Scraping Techniques to Collect Weather Data in South Sumatera," *Proc. 2018 Int. Conf. Electr. Eng. Comput. Sci. ICECOS 2018*, pp. 385–390, 2019, doi: 10.1109/ICECOS.2018.8605202.

[11] D. Kaiying, C. Senpeng, and D. Jingwei, "On optimisation of web crawler system on Scrapy framework," *Int. J. Wirel. Mob. Comput.*, vol. 18, no. 4, pp. 332–338, 2020, doi: 10.1504/IJWMC.2020.108530.

[12] X. Tian, T. Zhang, X. Zhuang, and X. He, "Research and Implementation of Campus Network Search Engine Based on Scrapy Framework and Elasticsearch," *Proc. 32nd Chinese Control Decis. Conf. CCDC 2020*, pp. 4193–4198, 2020, doi: 10.1109/CCDC49329.2020.9164582.

[13] X. Barillot, J. Hangouët, and H. Kadri-dahmani, "Generalization of the ‘ Douglas and Peucker ’ Algorithm for Cartographic Applications," pp. 2–4, 1975.

[14] S. Lakhera and T. Praveena, "Visual Analysis using Modified Ramer-Douglas – Peucker Algorithm on Time Series Data," *Int. Res. J. Eng. Technol.*, vol. 7, no. May, pp. 4225–4229, 2020.

[15] B. A. Priyaungga, D. B. Aji, M. Syahroni, N. T. S. Aji, and A. Saifudin, "Black Box Testing on Library Applications Using Equivalence Partitions Techniques," *J. Teknol. Sist. Inf. dan Apl.*, vol. 3, no. 3, p. 150, 2020, doi: 10.32493/jtsi.v3i3.5343.