IOP Conference Series: Earth and Environmental Science

Data Cubes Integration in Spatial OLAP for Agricultural Commodities

To cite this article: A I Putri and I S Sitanggang 2017 *IOP Conf. Ser.: Earth Environ. Sci.* **58** 012034

View the article online for updates and enhancements.

Data Cubes Integration in Spatial OLAP for Agricultural Commodities

A I Putri[1] and I S Sitanggang[1]

[1]Department of Computer Science, Bogor Agricultural University, Jalan Meranti Wing 20 level 5 Kampus IPB, Bogor, West Java, Indonesia

Email: ananda.putri152@gmail.com, imas.sitanggang@ipb.ac.id

Abstract. Ministry of Agriculture Indonesia collects data of agricultural commodities in Indonesia in the annual period. Agricultural commodities data include food crops, horticulture, plantations, and livestock. The data are available in the spreadsheet format. This study developed data cubes for food crops, plantations, and livestock using the galaxy schema of data warehouse and integrated the data cubes into the SOLAP Horticulture using SpagoBI. SOLAP is useful for data analysis and data visualization. The application displays agricultural commodities data in form of crosstab and chart. This study also developed the location intelligence module that visualizes agricultural commodities data on the map. The system was tested using the black box approach. The result showed that main functions including roll up, drill down, slice, dice, and pivot work properly. This application is expected to enable users to easily obtain data summaries of agricultural commodities.

1. Introduction

Agricultural commodities in Indonesia are very diverse which are divided into several sectors: food crops, horticulture, plantations, and livestock. Ministry of Agriculture Indonesia collects data of agricultural commodities in Indonesia in the annual period from the national to the district level. The data are available in the spreadsheet format so that it is not easy for users to get historical data summaries of agricultural commodities. Every year the data of agricultural commodities continue growing. The growth of data will generate datasets in very large quantities. Therefore, we need a tool that can automatically summarize information and transform data into a useful knowledge [1]. Data warehouse is a long-term storage space to store a collection of data from multiple sources in a scheme to help users in decision making [1]. A data warehouse can store data in a multidimensional space to make a data cube. A data warehouse uses the multidimensional data model and Online Analytical Processing (OLAP) to explore data in detail [2].

Spatial data are data that contain geographic information of a location on the earth's surface. A spatial data warehouse is a collection of spatial data which is subject oriented, integrated, time variant, and non-volatile to make decisions [1]. Development of a spatial data warehouse to integrate thematic and geographic data produce a spatial data cube for analyzing multidimensional spatial data [1]. Nowadays, there is a system which is called Spatial Online Analysis Processing (SOLAP) that can be used for data analysis and data visualization. A SOLAP is integrated with a spatial data warehouse to help users in obtaining summaries of historical data easily and quickly.

IOP Conf. Series: Earth and Environmental Science **58** (2017) 012034 doi:10.1088/1755-1315/58/1/012034

This study developed spatial data cubes for food crops, plantations, and livestock using the galaxy schema of data warehouse and integrated the data cubes into the SOLAP Horticulture using SpagoBI. SpagoBI was utilized to create data cubes. SpagoBI is a complete, open source, and flexible business intelligence tool. SpagoBI is complete because it covers all the needs of business intelligence with innovative solutions and provides a variety of analysis tools [3]. This study developed a SOLAP and a location intelligence module for agricultural commodities. The application provides data summaries represented in crosstab, chart or dynamic maps. We expect that the application can facilitate users to get data summaries of food crops, plantations or livestock appropriately.

2. Data and Method

2.1. Data

The data of agricultural commodities used in this study are food crops, plantations, and livestocks from 2000 to 2014 at the district level in Indonesia. The data were obtained from the website of Ministry of Agriculture Indonesia at http://aplikasi.pertanian.go.id/bdsp/newkom.asp in Microsoft Excel format (.xls) [4]. Attributes of the data include types of commodity, location, time, and status of the data. The measures in food crops data are harvested area (ha), production (ton), and productivity (kg/ha). The measures in plantations data are acreage (ha), production (ton), and productivity (kg/ha) and the measures in livestocks data are population (tail) and production (ton). Figure 1 shows the data of harvested area of food crop in Majalengka, West Java Province in the period of 2010-2019.

Hasil Pencarian Berdasarkan Komoditi

Sub Sektor	:	Tanaman Pangan
Indikator	:	Luas Panen
Level	:	Kabupaten/Kota
Provinsi	:	Jawa Barat
Kabupaten/Kota	:	Kab. Majalengka
Status Angka	:	Angka tetap
Tahun	:	2010-2019

Komoditi	Satuan	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
Jagung	Ha	18,577.00	16,062.00	19,738.00	17,149.00	15,943.00	0.00	0.00	0.00	0.00	0.00
Kacang Hijau	Ha	1,473.00	1,601.00	965.00	811.00	1,134.00	0.00	0.00	0.00	0.00	0.00
Kacang Tanah	Ha	1,112.00	962.00	696.00	679.00	553.00	0.00	0.00	0.00	0.00	0.00
Kedelai	Ha	2,347.00	1,514.00	1,428.00	507.00	1,344.00	0.00	0.00	0.00	0.00	0.00
Padi	Ha	103,392.00	98,289.00	101,279.00	108,008.00	105,558.00	0.00	0.00	0.00	0.00	0.00
Padi Ladang	Ha	2,284.00	1,534.00	1,818.00	1,887.00	2,758.00	0.00	0.00	0.00	0.00	0.00
Padi Sawah	Ha	101,108.00	96,755.00	99,461.00	106,121.00	102,800.00	0.00	0.00	0.00	0.00	0.00
Ubijalar	Ha	732.00	671.00	727.00	621.00	485.00	0.00	0.00	0.00	0.00	0.00
Ubikayu / Ketela Pohon	Ha	1,538.00	1,169.00	1,123.00	920.00	581.00	0.00	0.00	0.00	0.00	0.00

Sumber Data : Kementerian Pertanian

Figure 1. Food crop data.

This study was conducted in several steps including design of data warehouse, implementation of data warehouse, design of SOLAP, new data cubes integration, implementation of SOLAP for new data cubes, implementation of location intelligence for new data cubes, as well as SOLAP and location intelligence testing.

2.2. Design of data warehouse

A data warehouse is usually modeled by a multidimensional data structure called a data cube. The dimension is a collection of attributes in the scheme and the fact is value of several measurements [1]. The multidimensional data scheme used in this study is the galaxy scheme. In this scheme, there are some dimension tables and fact tables that can share dimension tables so the scheme contains some star schemas that are connected each other. In the galaxy scheme, one dimension table can be used simultaneously by multiple fact tables. The fact table contains some measures and some foreign keys that connect with the primary key in a dimension table [5]. A dimension table contains detailed data and descriptive text that describes a foreign key in the fact table.

2.3. Implementation of data warehouse

In implementation of the data warehouse, data are managed using the DBMS PostgreSQL. Multidimensional data consist of dimensions, measure, and facts. The data are stored in fact tables and dimension tables. Data cubes are implemented in SpagoBI Studio.

2.4. Design of SOLAP and new data cubes integration

SpagoBI Studio was used to design and to construct data cubes of food crops, plantations and live stocks. The scheme of data warehouse adopted is the galaxy scheme. The galaxy scheme can integrate new data cubes with dimension tables that have been built in the SOLAP for horticultural crops.

2.5. Implementation of SOLAP for new data cubes

SOLAP for agricultural commodities was built on the web application using SpagoBI as the OLAP server. The new data cubes stored in SpagoBI Studio are connected with SpagoBI Server to perform OLAP operations.

2.6. Implementation of location intelligence for new data cubes

Implementation of location intelligence was carried in SpagoBI Server. Maps for visualization of agriculture data are stored in GeoServer. SpagoBI has two types of dynamic map visualizations namely map zone and map point. A map zone is a polygon map. Each polygon has a different color representing different value of measures. The higher of measure value, the darker color of the polygon. A map point is a map showed by dots. Each point has a different size. The higher of measure value, the greater size of the dots on a map point.

2.7. SOLAP and location intelligence testing

Data verification was done before system testing to evaluate the data provided by the SOLAP. In addition, SOLAP testing was performed on OLAP operations including roll up, drill down, slice, dice, and pivot on each data cube. SOLAP testing was also conducted to evaluate data visualization in crosstab and chart for selected dimensions and measure. Some functions in the location intelligence were tested including displaying a map in a polygon or a point, displaying a map according to the selected indicators, and data filter on a map. System testing was done to find the functions that could not work properly. Moreover, testing results are important for further system development.

3. Result and Discussion

3.1. Preprocessing of data

Data of food crops, plantations, and livestock were obtained from the website of Ministry of Agriculture Indonesia. The data include agriculture commodities from 2000 to 2014 at the district level. Status number of the data used is fixed numbers. Measures of food crops, plantations and livestock have different units so data conversion on the measures is required. Number of records in food crops, plantations crops and livestock are 59,997, 80,140 and 199,610 respectively.

3.2. Design of data warehouse

The data warehouse scheme adopted in this study is the galaxy scheme. Dimension tables that are connected to the new fact tables are time, location, and status of data. This study created three new dimension tables and three fact tables. Three dimensions are food crop commodities, plantation commodities, and livestock commodities. Three fact tables are food crop fact, plantations fact, and livestock fact. Therefore the data warehouse of agriculture commodities has seven dimensions and four fact tables. The SOLAP for agricultural commodities has four data cubes including horticultural crops, food crops, plantations and livestock. Figure 2 shows the galaxy schema for data cubes of food crops, plantations and livestock.

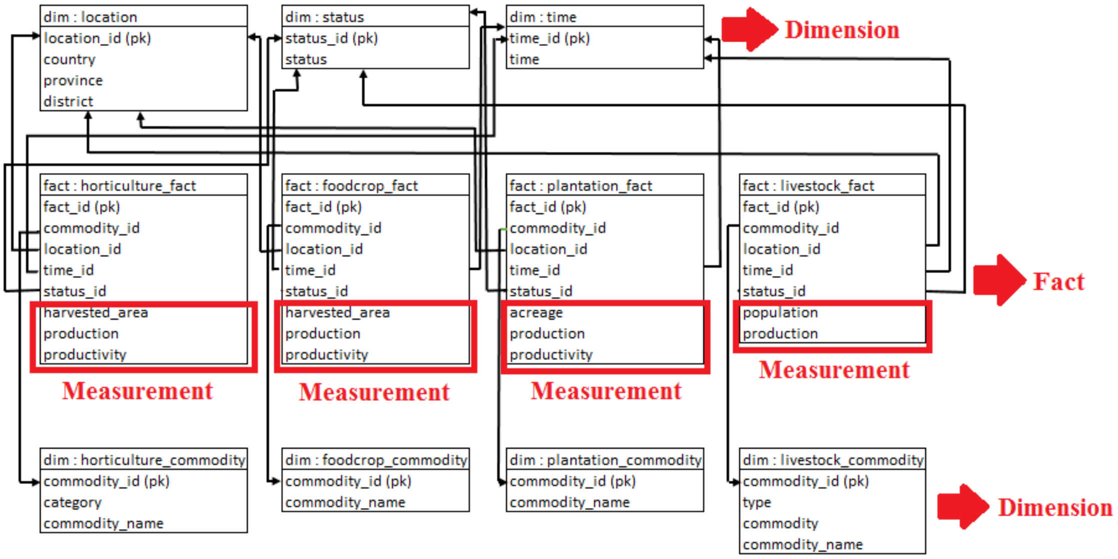

Figure 2. Galaxy scheme for data cubes of horticulture, food crop, plantation, and livestock.

3.3. Implementation of data warehouse

Fact tables and dimension tables in the data warehouse of agriculture commodities are stored on DBMS PostgreSQL. The data warehouse consists of seven dimension tables and four facts tables. Seven dimension tables are location, status, time, horticulture commodities (horticulture_commodity), food crop commodities (foodcrop_commodity), plantation commodities (plantation_commodity), and livestock commodities (livestock_commodity). Four fact tables are horticultures (horticulture_fact), food crops (foodcrop_fact), plantations (plantation_fact), and livestocks (livestock_fact). Dimension tables were implemented before the fact tables were created because each dimension table contains a primary key that is used as a foreign key in the fact table.

3.4. Design of SOLAP and new data cubes integration

This study designs SOLAP and data cubes using SpagoBI Studio. The dimension tables and fact tables stored in the DBMS PostgreSQL are connected to SpagoBI Studio using the PostgreSQL JDBC driver. Data cubes were created using SpagoBI Studio by assigning the type of dimension to dimension tables, fact tables to cubes, and measures in fact tables to measure.

3.5. Implementation of SOLAP for new data cubes

SOLAP was implemented by creating a server in SpagoBI Studio and then the server was connected to the data warehouse of agriculture commodities using SpagoBI Server. Furthermore, the OLAP template was made in SpagoBI Studio with the xml format for the multidimensional data model. The OLAP template is deployed into SpagoBI Server.

The result of SOLAP implementation is a spatial data warehouse of agriculture commodities which is available in the SpagoBI Server. OLAP operations such as roll up, drill down, slice, dice, and pivot are performed on the agriculture crops data and the results are displayed in crosstab and chart. The crosstab and chart can be downloaded into Microsoft Excel or PDF formats. Figure 3 shows the setting for OLAP operations.

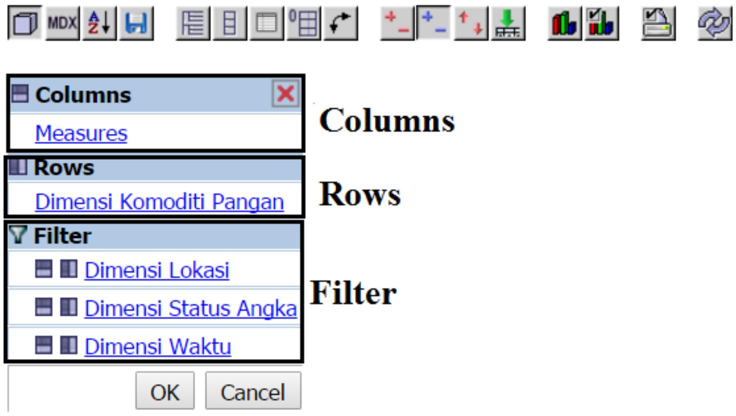

Figure 3. OLAP navigator.

Figure 3 shows the view of OLAP navigator. In the OLAP navigator, users can select measures and dimensions including the dimension of the location, commodity, status of data, and time. The dimensions and measure can be displayed as columns or rows in crosstab according to user's needs. The filter is used to select contents of the selected dimension. Figure 4 provides the result of SOLAP operation for food crops with two dimensions namely commodity and location from 2000 to 2014. Moreover, an example of SOLAP operation in harvested area of food crops in West Java from 2000 to 2014 is provided in figure 5.

		Measures		
Dimensi Komoditi Pangan	Dimensi Lokasi	Luas Panen (Ha)	Produktivitas (Ku/Ha)	Produksi (Ton)
−All Dimensi Komoditi Pangan	+All Dimensi Lokasi	452,369,204	3,250,682	2,142,393,875
[12] Jagung	+All Dimensi Lokasi	1,202,300	1,901	5,737,382
[12] Kacang Hijau	+All Dimensi Lokasi	90,016	421	96,465
[12] Kacang Tanah	+All Dimensi Lokasi	140,134	705	188,738
[12] Kedelai	+All Dimensi Lokasi	157,208	7,362	3,635,550
[12] Padi	+All Dimensi Lokasi	2,104,631	25,623	12,512,298
[12] Padi Ladang	+All Dimensi Lokasi	150,527	8,121	640,584
[12] Padi Sawah	+All Dimensi Lokasi	1,954,175	19,188	11,872,086
[12] Ubijalar	+All Dimensi Lokasi	13,823	7,337	315,477
[12] Ubikayu / Ketela Pohon	+All Dimensi Lokasi	159,117	9,491	3,677,400
Jagung	−All Dimensi Lokasi	47,413,416	210,446	173,757,133
	−jawa	24,353,778	96,863	98,326,616
	+banten	79,139	2,432	218,516
	+daerah istimewa yogyakarta	926,529	2,723	3,020,064
	+daerah khusus ibukota jakarta	1,634	361	529
	−jawa barat	1,672,351	54,583	7,013,854
	Kab. Bandung	132,940	702	491,688
	Kab. Bandung Barat	24,445	321	140,233
	Kab. Bekasi	510	321	737
	Kab. Bogor	2,847	212	11,426
	Kab. Ciamis	66,797	712	314,222
	Kab. Cianjur	88,098	569	286,874
	Kab. Cirebon	3,092	614	8,135
	Kab. Garut	705,315	785	3,327,015
	Kab. Indramayu	3,262	548	8,741
	Kab. Karawang	3,079	41,539	9,847
	Kab. Kuningan	60,278	601	194,605

Hierarchy

Figure 4. Food crops data summary in crosstab.

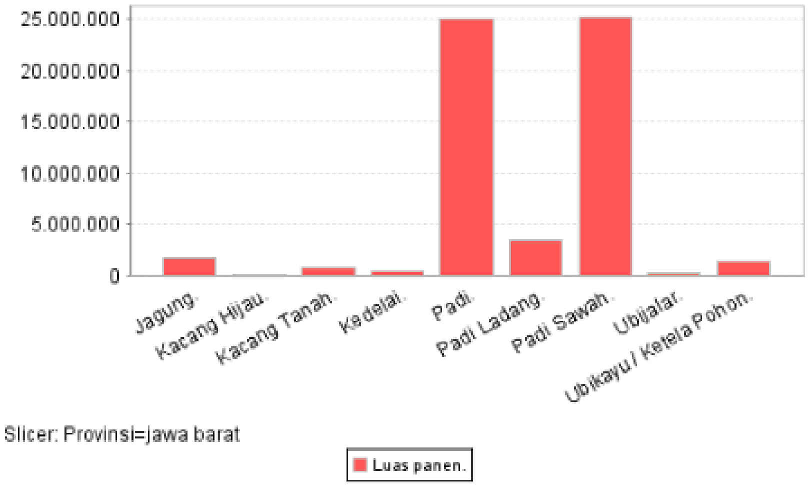

Figure 5. Food crops data summary in chart.

3.6. Implementation of location intelligence for new data cubes

The initial step in location intelligence implementation is preparing and storing the map of Indonesia in GeoServer in the shape file format (shp). The map of Indonesia was obtained from Geospatial Information Agency (BIG). Furthermore, the map is displayed in the SpagoBI server. The second step is creating datasets for each data cube in SpagoBI Server to be visualized on the map of Indonesia. Data retrieved from the data warehouse of agriculture commodities are commodity names, id locations, provinces, countries, times, harvested area, crop production and crop productivity. This study utilizes the dynamic map for data visualization. The dynamic map displays data in polygon maps (map zone) and point maps. The location intelligence requires a template to connect GeoServer to SpagoBI Server. The template is represented in the JSON format. In the template there are indicators to be presented on the map. The template is used to match the location id in GeoServer with the location id in the dataset. Figure 6 shows the result of the location intelligence module for harvested area of food crops that is represented in the map zone for the period of 2000-2014. The darker color of the polygon, the greater value of the harvested area in the region. In data visualization, there are some functions that can be used by users to explore the map. These functions include indicators, filters, and legends. The indicator has a function to show the measure. Filter is used to filter the data to be displayed on the map. Legend provides information related to different colors on the map.

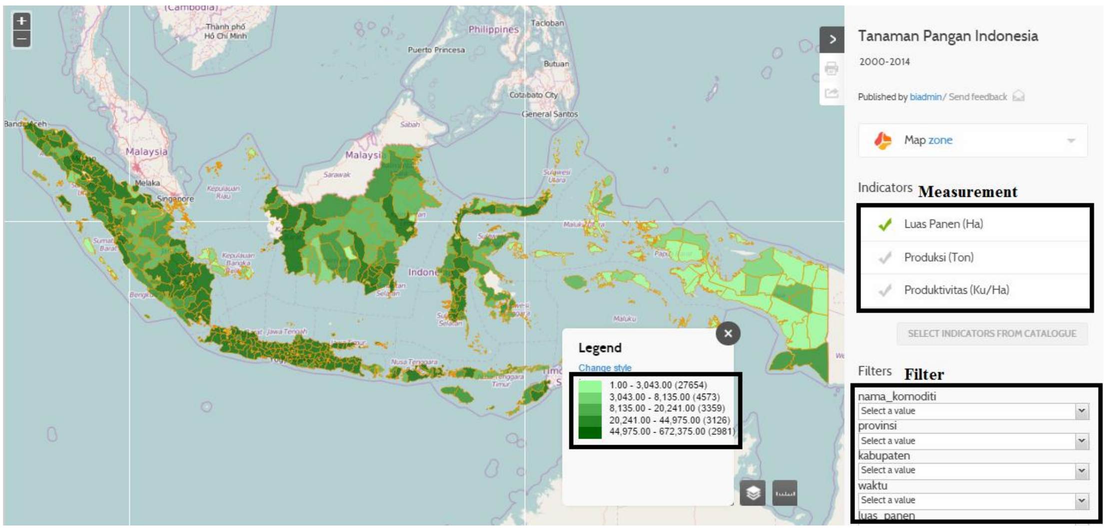

Figure 6. Representation of food crops data summary in a map zone.

Figure 7 shows the result of the location intelligence module for harvested area of food crops that is represented in the map point for the period of 2000-2014. In the figure 7, size of dots relates to the value of the harvested area in the region.

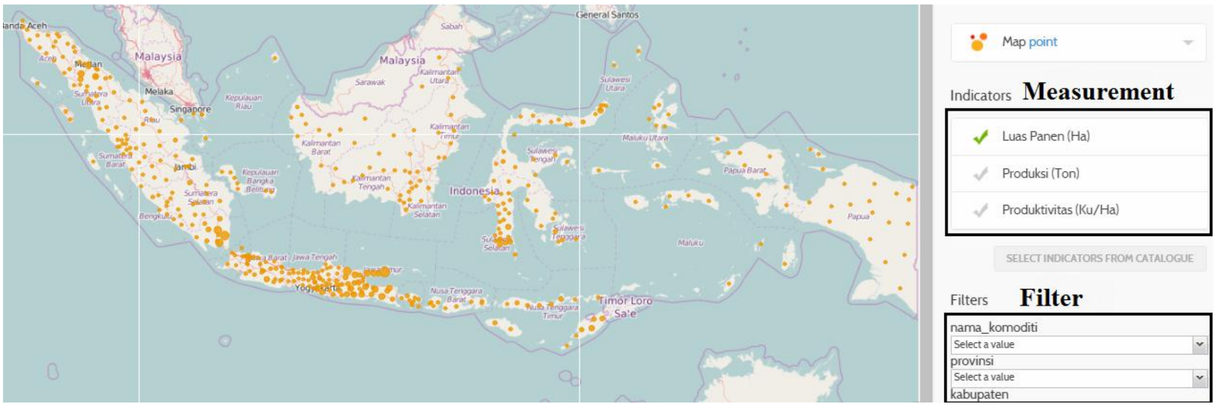

Figure 7. Food crops data in a map point.

3.7. SOLAP and location intelligence testing

Data verification was done before SOLAP and location intelligence testing. The result shows that data summaries provided by the SOLAP are valid and correspond to the original data that are available in website of Ministry of Agriculture. The testing of SOLAP for agricultural commodities and the location intelligence module was performed using the black box approach. Table 1 shows the test scenarios for SOLAP for agricultural commodities, whereas table 2 shows the test scenarios for the location intelligence module. The testing results on the SOLAP for agricultural commodities and the location intelligence module show that all functions on the system work properly.

Table 1. Testing results of the SOLAP for agricultural commodities.

No	Task	Scenario	Expected result	Result
1	Selecting data for certain dimension and measure	Examine each dimension of data cube	Data are displayed according to selected dimensions	The task works properly
2	Displaying data in crosstab	Selecting one or more dimension(s) in OLAP Navigator and click “ok” to display the crosstab	Data are represented in crosstab	The task works properly
3	Displaying data in chart	Selecting each dimension of data cube and choose “chart”	Data are represented in chart	The task works properly
4	Displaying data as the result of OLAP operations on each data cube	Selecting dimensions according to OLAP operations	Data are displayed according to selected OLAP operations	The task works properly

Table 2. Testing results of the location intelligence module.

No	Task	Scenario	Expected Result	Result
1	Displaying map zone for each data cube	Selecting map zone	Different color of polygons according to measure values for each district	The task works properly
2	Displaying map point for each data cube	Selecting map point	Different size of points according to measure values for each district	The task works properly
3	Filtering in map	Selecting one or more filter(s)	A map is displayed according to the filtering result	The task works properly
4	Displaying map according to indicator	Selecting one indicator	A map is displayed according to the indicator	The task works properly
5	Zoom in and zoom out on the map	Selecting zoom in or zoom out features on the map	A map shown to be larger or smaller than the previous map	The task works properly

4. Summary

Development and integration of new data cubes for horticulture crops, food crops, plantations and livestock in the SOLAP for agricultural commodities were successfully performed using SpagoBI. The SOLAP for agricultural commodities is integrated to the data warehouse developed based on the galaxy scheme. The SOLAP provides agriculture data summaries represented in crosstab and chart. The location intelligence module for each data cube was also successfully implemented to visualize agriculture data summaries in dynamic maps. The testing results of SOLAP and the location intelligence module for data cubes of horticultural crops, food crops, plantations and livestock indicate that all functions in the SOLAP and the location intelligence work properly.

References

[1] Han J, Kamber M and Pei J 2012 *Data Mining: Concept and Techniques* 3rd ed (San Francisco (US): Elsevier)

[2] Boulil K, Ber L F, Bimonte S, Grac C and Cernesson F 2014 Multidimensional modeling and analysis of large and complex watercourse data: an OLAP-based solution, *Ecological Informatics* (Elsevier) pp 30

[3] Cazzin G 2012 *Business Intelligence with SpagoBI* (Padua (IT): SpagoBI Competency Center)

[4] [Ministry of Agriculture] Ministry of Agriculture 2016 The database agricultural statistics retrieved from http://aplikasi.pertanian.go.id/bdsp/ newkom.asp.

[5] Kimball R and Ross M 2013 *The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling* (Indianapolis (IN-US): Wiley)