The importance of building data cube for MARDI dashboard development

(Kepentingan penyediaan kiub data bagi pembangunan papan pemuka MARDI)

Mohd Shafiq Azizan*, Aida Al-Quswa*, Zaireen Abdul Rahman*, Azrina Asmuni**, Elmaliana Albahari*, Muhammad Izzat Farid Musaddin* and Noor Marziah Mat Zain*

Keywords: analytical processing, data cube, dashboard, multi-dimensional, sequential

Abstract
Analyzing data is a very tedious work process if the data representation is in the traditional structure. This structure is usually represented in the report format in which the data is tabulated as a hardcopy and is not editable. The Strategic Planning and Innovation Management Centre (PI) of Malaysian Agricultural Research and Development Institute (MARDI) has encountered issues when analyzing the agriculture crop statistic reports for the dashboard development being published in this structure. Furthermore, going through enormous amount of data and stacks of reports is highly time consuming, prolonging the analysis process. Data cubes in the computer science context are known as a multi-dimensional array of values. Analytical processing can be done easily if the data structure is built in the data cube format. Therefore, this study aimed to show the role of crop statistic data cube in the analytical processing for MARDI's dashboard development. The approaches for developing the crop statistic data cube involved three (3) constructively built sequential steps: 1) requirements gathering, 2) database design, and 3) data preparation. Using this three-step approach, we were able to develop a dashboard that effectively solved issues encountered by PI, and fulfilled MARDI's stakeholder expectations and requirements.

Introduction

Data is wealth and consequently, every organization has their own data sets to leverage. It is important to understand the type of data sets in our possessions and ways to extract valuable information out of them. Additionally, this valuable information requires an adequate medium to project the outcome so that analyses can be further conducted. Hence, using dashboard is the best solution to interactively populate the information and visualize them in infographics charts, tables and maps so that the stakeholders can peruse them.

Dashboard is classified as one of the major business intelligence (BI) tools. BI is a process of applying tools and techniques to gather and analyze the data collected from multiple (both internal and external) sources by creating knowledge that helps in decision making (Sohollo 2011). Leveraging on this tool will help to answer all the business questions derived from the stakeholders using technologies based on interactive computer structures (Hall 2003).

On the other hand, analytical processing outcomes and valuable information are translated in dashboard.

*ICT Management Centre, MARDI Headquarters, Persiaran MARDI-UPM, 43400, Serdang, Selangor
**Strategic Planning and Innovation Management Centre, MARDI Headquarters, Persiaran MARDI-UPM, 43400 Serdang, Selangor
E-mail: shafiq@mardi.gov.my

They are developed based on the analysis performed from the transactional and historical data possessed by the organization. Before analytical processing can be achieved, data preparation process needs to be completed to standardize the data quality. The key features of data quality includes accurate and consistent data according to their type (Bhattacharjee et al. 2014). Missing information, misspelling or invalid data are among data quality problems that need to be dealt seriously.

The Strategic Planning and Innovation Management Centre (PI) of Malaysian Agricultural Research and Development Institute (MARDI) has been facing difficulties in analyzing the published agriculture crop statistic reports for the dashboard development. The data is represented in a traditional structure, with the report format data being tabulated as a hardcopy and not editable. Furthermore, going through an enormous amount of data and stacks of reports is highly time consuming, prolonging the analysis process.

This study was carried out to describe the role of data cube in order to overcome the issues encountered by the stakeholders. Based on a study by Putri and Sitanggang (2017), the data cubes integration for agricultural commodities are achievable in online analytical processing (OLAP).

The approaches of developing the crop statistic data cube involved three (3) constructively built sequential steps: 1) requirements gathering, 2) database design, and 3) data preparation. In order to develop the MARDI dashboard, the data cube needed to be prepared as the execution time for the data cube query was faster than the standard query as the data has been indexed and precomputed (von Davier et al. 2019).

Data Cube

Data cubes are multidimensional extension of 2-dimensional tables or 2-dimensional matrix (column and rows) (Kay 2004). The data cube is applied when it is necessary to extract important data or aggregate complex data. Online analytical processing (OLAP) is usually used to utilize the data cubes. It is a computing method that enables users to analyze the selective extract and query data from different perspectives (Rouse 2018). Data cubes are categorized into two categories as shown in *Table 1*.

The views are created based on these tables as they are represented as cuboid (Kay 2004). Comparing between both data cube types, ROLAP is better compared to MOLAP as it can handle larger data collections without making the cube sparser. Furthermore, the storage requirements of ROLAP cube are far more considerable

Table 1. Catagories of data cube

Multidimensional Online Analytical Processing (MOLAP)	Relational Online Analytical Processing (ROLAP)
Data is stored in multidimensional array	Data is stored in relation tables
Data is fetched from Multiple Dimensional Database	Data is fetched from Data Warehouse
Used for limited data volumes	Used for large data volumes
Dynamic multidimensional view of data is created	Static multidimensional view of data is created
Response time is less	Response time is high
When the number of dimension increases, the data cube will become sparser (several cells that represent specific attribute combinations will not contain any aggregated data)	Use of relational database model techniques such as indexes and joins, hence number of dimension does not increase

compared to MOLAP, which will increase according to time. Therefore, the conceptual framework for the crop statistic data cube development moves from the traditional data structure representation to the ROLAP data cube implementation.

Methodology

There were several actions that needed to be performed when building a data cube, including requirements gathering, database designing, and data preparation.

Requirements gathering

Before developing any data cube, requirements gathering was performed. In this phase, the stakeholders of an organization determine the information that the organization is required to project. In this study, we have identified the PI of MARDI as the stakeholder. A series of discussion was setup to gather the requirements from the stakeholder. This discussion included activities like question and answer (Q&A) session and brain storming session.

The stakeholder has decided that for this case study to focus on the Malaysian crop statistics data projection. The statistic crop reports referred as the primary source were the published report uploaded by the Department of Agriculture (DOA). The business requirement identified was that PI was able to analyze the crop statistic report while taking a minimal amount of time to go through the series of data represented in traditional structure. Hence, the goal for requirements gathering was to address all the concerns highlighted by the stakeholders.

Database design

Database can be defined as piles of data organized and stored in tables structurally in a computer (Dietrich 2017). Before data can be stored in a database, the database needs to be design. A structured and organized database will optimize the storage usage. Furthermore, retrieving records can be done easily when the data has been normalized because of absence of redundant records.

The database was built from tables where the data resided. The table consisted of field name, data types and length. Defining each data with its characteristic was important as resource utilization could be optimized to avoid resource wastage that could be a setback for an organization. Field name in the table was where the data was placed under a place holder. The naming convention for the data field name must be meaningful and unique as it was the main source to be referred to upon data retrieval.

Once the field name was clearly defined, it must hold a data type. A data type refers to a format of data storage that can hold for a range of values. Data types are categorized into 3 common categories:

a) Numeric
   - i) Integer - Numbers with signed or unsigned (up to 11 digits' width). Example: 273
   - ii) Double - A double precision floating-point number that cannot be unsigned.
     Example: 28.9991
   - iii) Decimal - An unpacked floating-point number that cannot be unsigned.
     Example:0.2221

b) Data and Time
   - i) Time - Stores the time in a HH:MM:SS format (Example: 12:11:33).
   - ii) Year - Stores a year in a 2-digit or a 4-digit format
     (Example: 2000).
   - iii) Timestamp - DATETIME format without the hyphens between numbers
     (Example: 3:30pm on December 30th, 1973 would be stored as 19731230153000).
   - iv) Date - A date in YYYY-MM-DD format, between 1000-01-01 and 9999-12-31

(Example: 10 July 2010 would be stored as 2010-07-10).

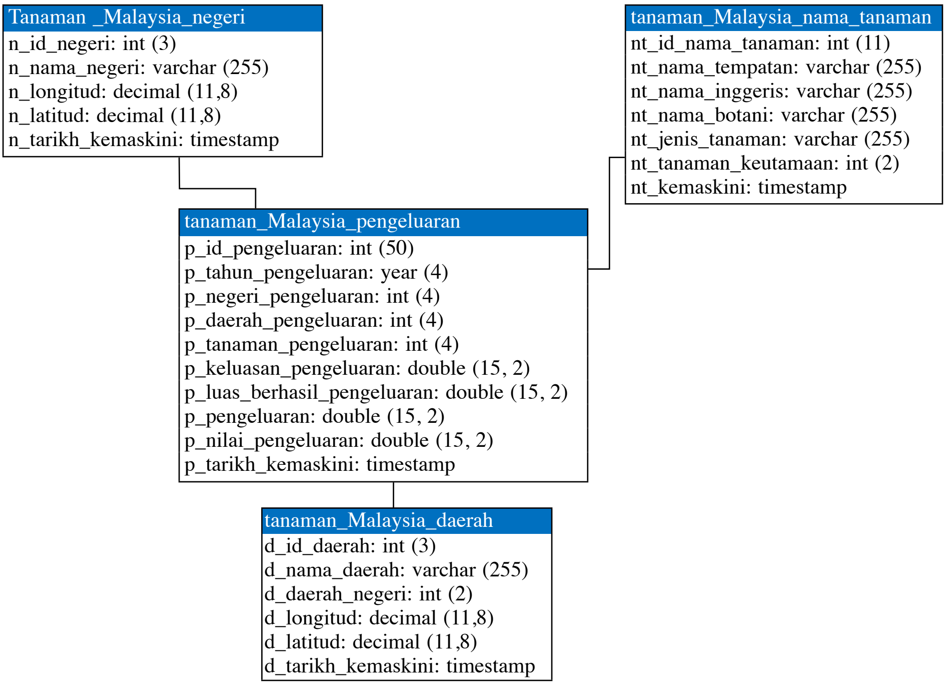
c) String
   i) Varchar - A variable-length string between 1 and 255 characters in length.
      Example: Hello.
   ii) Blob - BLOBs are “Binary Large Objects” and are used to store large amounts of binary data, such as images or other types of files.
      Example: Image file, Document file.
   iii) Enum - An enumeration, which is a fancy term for list.
      Example: If you wanted your field to contain ‘X’ or ‘Y’ or ‘Z’, you would define your ENUM as ENUM (‘X’, ‘Y’, ‘Z’).

Based on *Figure 1*, the crop statistics database design was divided into 4 tables as follows:

a) tanaman_malaysia_negeri
   Stores data such as Malaysia’s states name, longitude and latitude. Five field names have been identified for this table with its data type and length of values.
b) tanaman_malaysia_daerah
   Stores data for each Malaysia’s state district name, longitude and latitude. Six field names have been identified with its data type and length of values.
c) tanaman_malaysia_nama_tanaman;
   Stores data for each crop local name, English name, Botanic name, type of crops and its priority. Seven field names have been identified with its data type and length of values.
d) tanaman_malaysia_pengeluaran
   Stores data such as the crop hectarage, harvested area, production and value of production. Ten field names have been identified with its data type and length of values.

Data Preparation

In order to ensure the quality of data being projected, data cleansing process needs to take place. Eliminating data errors such as misspelling, missing information and invalid data are among the actions in data cleansing.



Tanaman _Malaysia_negeri
n_id_negeri: int (3)
n_nama_negeri: varchar (255)
n_longitud: decimal (11,8)
n_latitud: decimal (11,8)
n_tarikh_kemaskini: timestamp

tanaman_Malaysia_nama_tanaman
nt_id_nama_tanaman: int (11)
nt_nama_tempatan: varchar (255)
nt_nama_inggeris: varchar (255)
nt_nama_botani: varchar (255)
nt_jenis_tanaman: varchar (255)
nt_tanaman_keutamaan: int (2)
nt_kemaskini: timestamp

tanaman_Malaysia_pengeluaran
p_id_pengeluaran: int (50)
p_tahun_pengeluaran: year (4)
p_negeri_pengeluaran: int (4)
p_daerah_pengeluaran: int (4)
p_tanaman_pengeluaran: int (4)
p_keluasan_pengeluaran: double (15, 2)
p_luas_berhasil_pengeluaran: double (15, 2)
p_pengeluaran: double (15, 2)
p_nilai_pengeluaran: double (15, 2)
p_tarikh_kemaskini: timestamp

tanaman_Malaysia_daerah
d_id_daerah: int (3)
d_nama_daerah: varchar (255)
d_daerah_negeri: int (2)
d_longitud: decimal (11,8)
d_latitud: decimal (11,8)
d_tarikh_kemaskini: timestamp

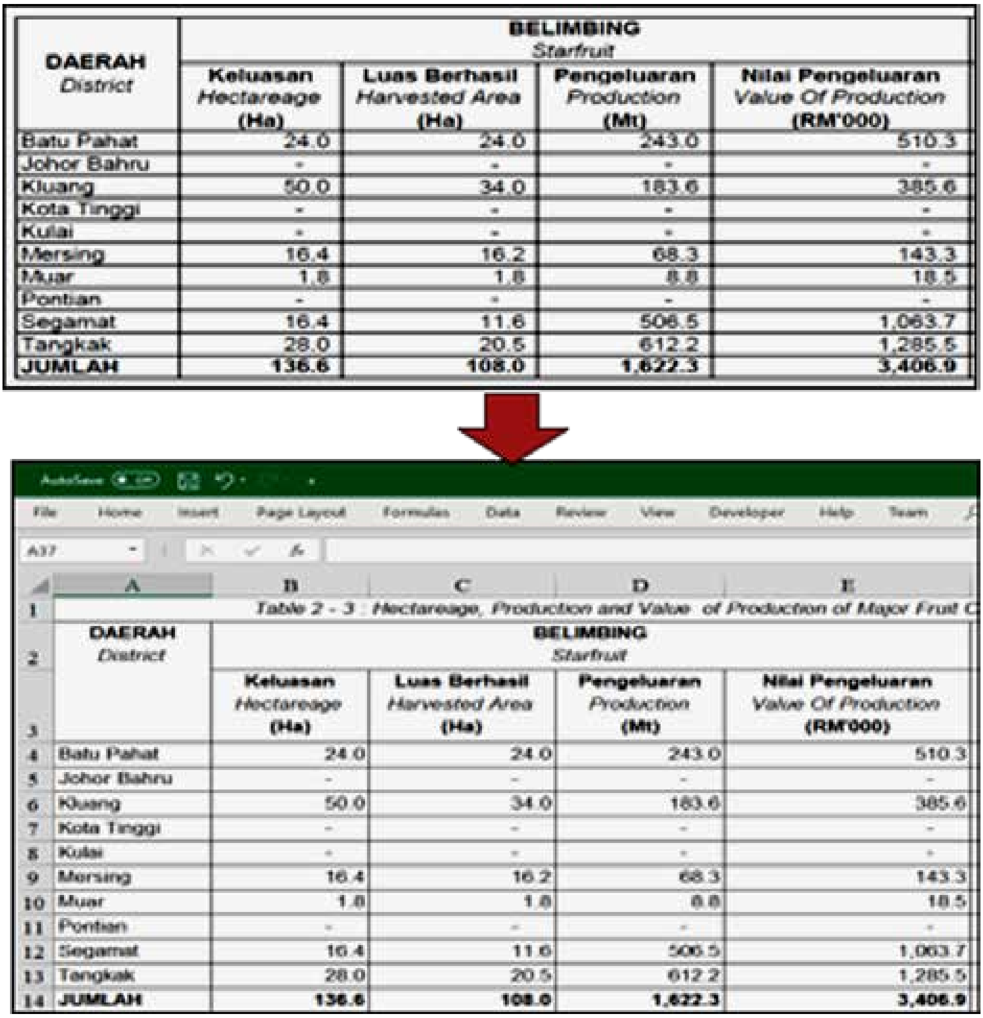
Figure 1. Crop Statistic Database Design

Extraction, Transformation and Loading (ETL) (Albrecht and Naumann 2008) is an integration process in data warehouse where the data from single or multiple source are extracted and then transformed into its desired format while ensuring the data quality. Firstly, the extraction process was done manually by obtaining the published report (Department of Agriculture Malaysia (DOA) 2020). Based on *Figure 2*, the PDF report format was converted into excel format.

Then, the excel report format was analyzed before moving to the next phase which was transformation. The populated records and figures shown in the report were analyzed based on what needed to be extracted, transformed and loaded. Storing unnecessary data in the database will increase resource wastage by the computer. Next, the extracted data was transformed according to the required format (*Figure 3*).

In figure 3, the extracted data was transformed as follows:

i) *Daerah* (District)
Highlighted in purple, the districts for Johor state was transformed into a number sequence. Referring to *Figure 3,* Batu Pahat district was converted to a number sequence, holding the number 1 value. The list continued for the next



DAERAH *District*	BELIMBING *Starfruit*			
	Keluasan *Hectareage* (Ha)	Luas Berhasil *Harvested Area* (Ha)	Pengeluaran *Production* (Mt)	Nilai Pengeluaran *Value Of Production* (RM'000)
Batu Pahat	24.0	24.0	243.0	510.3
Johor Bahru	-	-	-	-
Kluang	50.0	34.0	183.6	385.6
Kota Tinggi	-	-	-	-
Kulai	-	-	-	-
Mersing	16.4	16.2	68.3	143.3
Muar	1.8	1.8	8.8	18.5
Pontian	-	-	-	-
Segamat	16.4	11.6	506.5	1,063.7
Tangkak	28.0	20.5	612.2	1,285.5
JUMLAH	136.6	108.0	1,622.3	3,406.9

Table 2 - 3 : Hectareage, Production and Value of Production of Major Fruit C

	A	B	C	D	E
1		*Table 2 - 3 : Hectareage, Production and Value of Production of Major Fruit C*			
2	DAERAH *District*	BELIMBING *Starfruit*			
3		Keluasan *Hectareage* (Ha)	Luas Berhasil *Harvested Area* (Ha)	Pengeluaran *Production* (Mt)	Nilai Pengeluaran *Value Of Production* (RM'000)
4	Batu Pahat	24.0	24.0	243.0	510.3
5	Johor Bahru	-	-	-	-
6	Kluang	50.0	34.0	183.6	385.6
7	Kota Tinggi	-	-	-	-
8	Kulai	-	-	-	-
9	Mersing	16.4	16.2	68.3	143.3
10	Muar	1.8	1.8	8.8	18.5
11	Pontian	-	-	-	-
12	Segamat	16.4	11.6	506.5	1,063.7
13	Tangkak	28.0	20.5	612.2	1,285.5
14	JUMLAH	136.6	108.0	1,622.3	3,406.9

Figure 2. Extraction of data from PDF report format to Excel format

district, which was Johor Bahru assigned with the number 2 value, and ended with Tangkak district which had number 10 value. This conversion could eliminate misspelling the district names and retain the data quality.

ii) *Tanaman* (Crops)

Starfruit crops was selected for this example in *Figure 3*. Highlighted in green, the fruit crops for starfruit was transformed into a number sequence, holding the number 1 value. It was observed that storing a single digit number value used less storage to store the data compared to its full crop name.

iii) *Keluasan, Luas Berhasil, Pengeluaran dan Nilai Pengeluaran* (hectarage, harvested area, production and value of production)

The remaining figures fields, such as the hectarage, harvested area, production and value of production are highlighted in red in *Figure 3*. Realigning all the decimal figures shown to two decimal places formally standardized the data projection. Records holding a '-' value were transformed to '0.00' value. This conversion process allowed consistent data quality stored in the database.

The final step in the ETL process was data loading process. Data transformed according to the required format was loaded to the database.

Issue and challenges

Obtaining the published reports was done manually by downloading from the DOA's website (Department of Agriculture Malaysia (DOA) 2020). Furthermore, the reports were uploaded on a yearly basis. Hence, retrieving the past 5 years' crop statistic reports was a challenging task to complete. The next challenge encountered was the report representation in the PDF format which was non-editable.

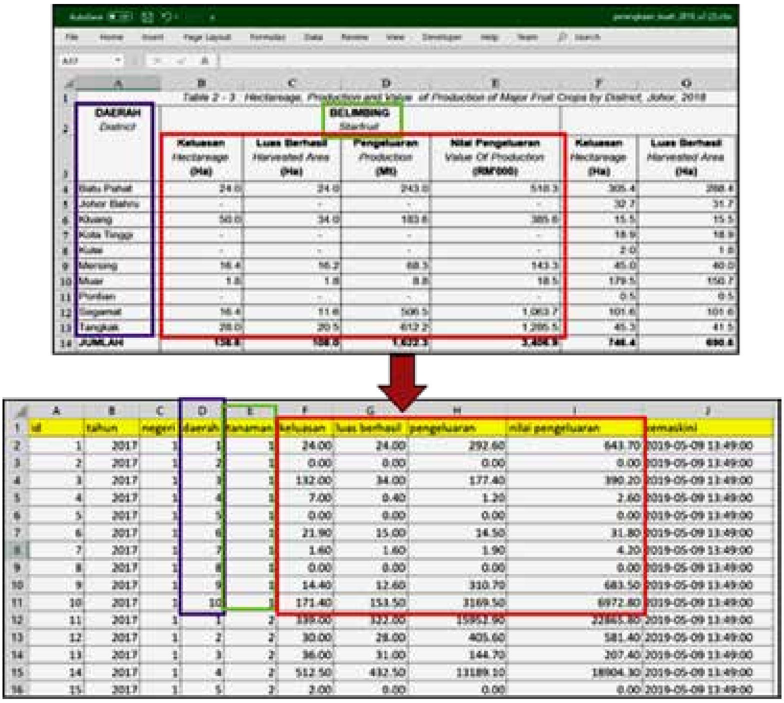


Table 2 - 3 Hectarage, Production and Value of Production of Major Fruit Crops by District, Johor 2018

DAERAH *District*	BELIMBING *Starfruit*					
	Keluasan *Hectarage* (Ha)	Luas Berhasil *Harvested Area* (Ha)	Pengeluaran *Production* (Mt)	Nilai Pengeluaran *Value Of Production* (RM'000)	Keluasan *Hectarage* (Ha)	Luas Berhasil *Harvested Area* (Ha)
Batu Pahat	24.0	24.0	243.0	510.3	305.4	288.4
Johor Bahru	-	-	-	-	32.7	31.7
Kluang	50.0	34.0	183.6	385.6	15.5	15.5
Kota Tinggi	-	-	-	-	18.9	18.9
Kulai	-	-	-	-	2.0	1.6
Mersing	16.4	16.2	68.3	143.3	45.0	40.0
Muar	1.6	1.6	8.8	18.5	179.5	150.7
Pontian	-	-	-	-	0.5	0.5
Segamat	16.4	11.6	506.5	1,063.7	101.6	101.6
Tangkak	28.0	20.5	612.2	1,285.5	45.3	41.5
JUMLAH	136.4	108.9	1,622.3	3,406.9	746.4	690.4

id	tahun	negeri	daerah	tanaman	keluasan	luas berhasil	pengeluaran	nilai pengeluaran	dimasukkini
1	2017	1	1	1	24.00	24.00	292.60	643.70	2019-05-09 13:49:00
2	2017	1	2	1	0.00	0.00	0.00	0.00	2019-05-09 13:49:00
3	2017	1	3	1	132.00	34.00	177.40	390.20	2019-05-09 13:49:00
4	2017	1	4	1	7.00	0.40	1.20	2.60	2019-05-09 13:49:00
5	2017	1	5	1	0.00	0.00	0.00	0.00	2019-05-09 13:49:00
6	2017	1	6	1	21.90	15.00	14.50	31.80	2019-05-09 13:49:00
7	2017	1	7	1	1.60	1.60	1.90	4.20	2019-05-09 13:49:00
8	2017	1	8	1	0.00	0.00	0.00	0.00	2019-05-09 13:49:00
9	2017	1	9	1	14.40	12.60	310.70	683.50	2019-05-09 13:49:00
10	2017	1	10	1	171.40	153.50	3169.50	6972.80	2019-05-09 13:49:00
11	2017	1	1	2	139.00	122.00	15952.90	22865.80	2019-05-09 13:49:00
12	2017	1	2	2	30.00	28.00	405.60	581.40	2019-05-09 13:49:00
13	2017	1	3	2	36.00	31.00	144.70	207.40	2019-05-09 13:49:00
14	2017	1	4	2	512.50	432.50	13189.10	18904.30	2019-05-09 13:49:00
15	2017	1	5	2	2.00	0.00	0.00	0.00	2019-05-09 13:49:00

Figure 3. Data transformation process

In order to accomplish data transformation, report format conversion process was performed. Therefore, through a third-party web application file converter, the PDF format report was converted to an editable Excel format.

Result and discussion

In the data warehouse, several multidimensional schemas were available to be implemented: Star Schema, Snowflake Schema and Galaxy Schema. Out of these three schemas, Star schema was the simplest type compared to the other two. It consisted of one fact table and a couple of dimension tables that resembled a star (Krishna 2020). Snowflake schema on the other hand, was very complex design whereby one fact table was surrounded by dimension tables which were in turn surrounded by more dimension table. Two fact tables sharing the same dimension tables between them was the characteristic of Galaxy Schema.

Galaxy schema has been adopted in the data cube integration in Spatial OLAP for agricultural commodities (Putri and Sitanggang 2017) as the schema suits best in the case study. However, the adoption of star schema was done in this study as it had one fact table and connected to three dimension tables (Kimball and Ross 2013). Based on *Figure 4*, the three dimension tables were tanaman_malaysia_negeri, tanaman_malaysia_daerah and tanaman_malaysia_nama_tanaman linked to the fact table, tanaman_malaysia_pengeluaran.

Once the data was successfully transformed according to the desired format, the data could be loaded to any chosen database technology. For this case study, Oracle MySQL Community database was selected. The transformed data was placed according to the design database which was defined with its field names, data types and length values.

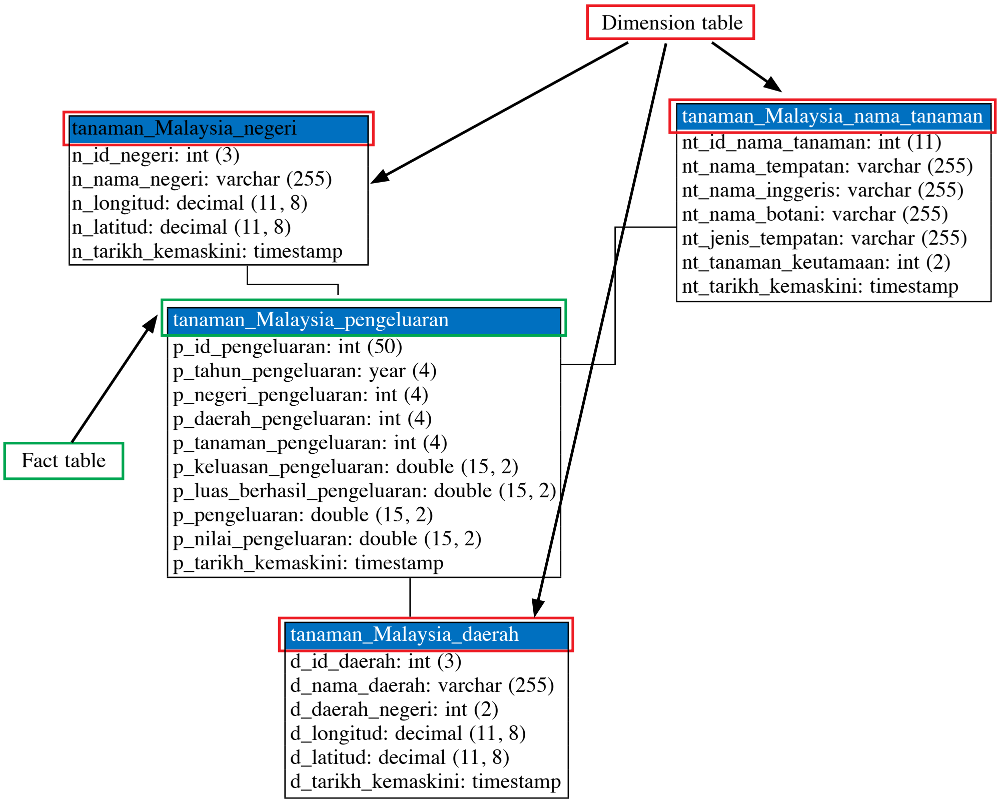
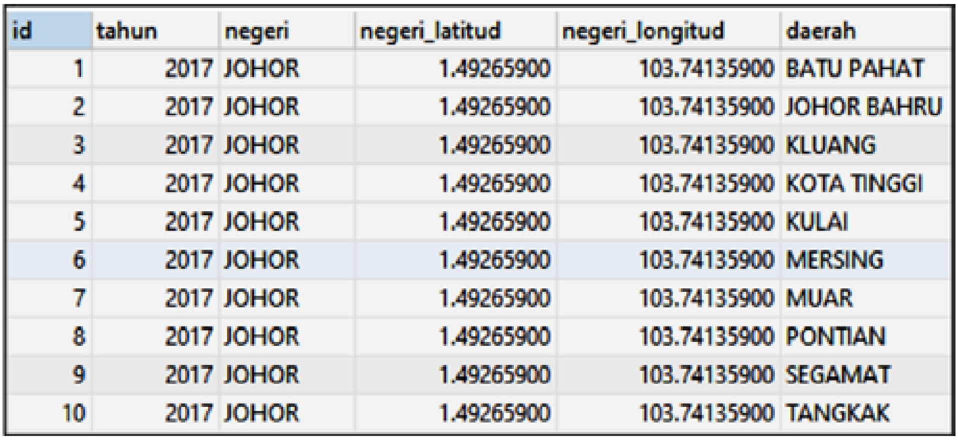

Dimension table
tanaman_Malaysia_negeri
n_id_negeri: int (3)
n_nama_negeri: varchar (255)
n_longitud: decimal (11, 8)
n_latitud: decimal (11, 8)
n_tarikh_kemaskini: timestamp
tanaman_Malaysia_nama_tanaman
nt_id_nama_tanaman: int (11)
nt_nama_tempatan: varchar (255)
nt_nama_inggeris: varchar (255)
nt_nama_botani: varchar (255)
nt_jenis_tempatan: varchar (255)
nt_tanaman_keutamaan: int (2)
nt_tarikh_kemaskini: timestamp
tanaman_Malaysia_pengeluaran
p_id_pengeluaran: int (50)
p_tahun_pengeluaran: year (4)
p_negeri_pengeluaran: int (4)
p_daerah_pengeluaran: int (4)
p_tanaman_pengeluaran: int (4)
p_keluasan_pengeluaran: double (15, 2)
p_luas_berhasil_pengeluaran: double (15, 2)
p_pengeluaran: double (15, 2)
p_nilai_pengeluaran: double (15, 2)
p_tarikh_kemaskini: timestamp
Fact table
tanaman_Malaysia_daerah
d_id_daerah: int (3)
d_nama_daerah: varchar (255)
d_daerah_negeri: int (2)
d_longitud: decimal (11, 8)
d_latitud: decimal (11, 8)
d_tarikh_kemaskini: timestamp

Figure 4. *Star Schema for data cube of Crop Statistic*

Retrieving a single record from the database required the execution of an SQL query to combine multiple tables to retrieve them. This representation of data led to the data being grouped into different dimensions, indexing data and precomputing queries (von Davier et al. 2019) known as a multidimensional data cube.

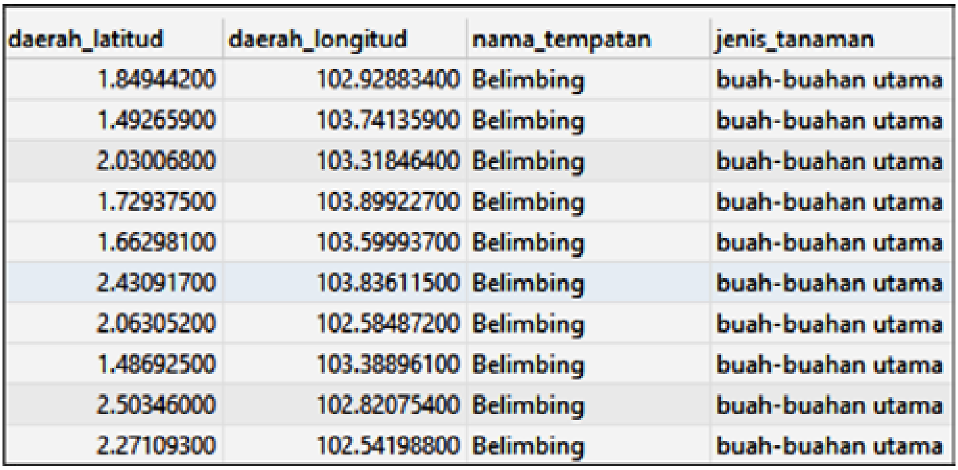
Based on *Figure 5(i), (ii)* and *(iii),* the data was retrieved from a single-view query from the data cube crop statistics shown in *Figure 4*. A database view is literally a temporary table created in a database and is accessible by a single select statement ('Learn SQL: SQL Views' n.d.). In *Figure 5(i),* negeri and daerah column projected the full name of the 'Johor' state and its district names.

On the other hand, in *Figure 5(ii)* the consistency of Johor's district geolocation was populated (longitude and latitude) with 8 decimal places. Belimbing fruit crop was shown for the crop type and referred in the nama_tempatan column. Furthermore, the



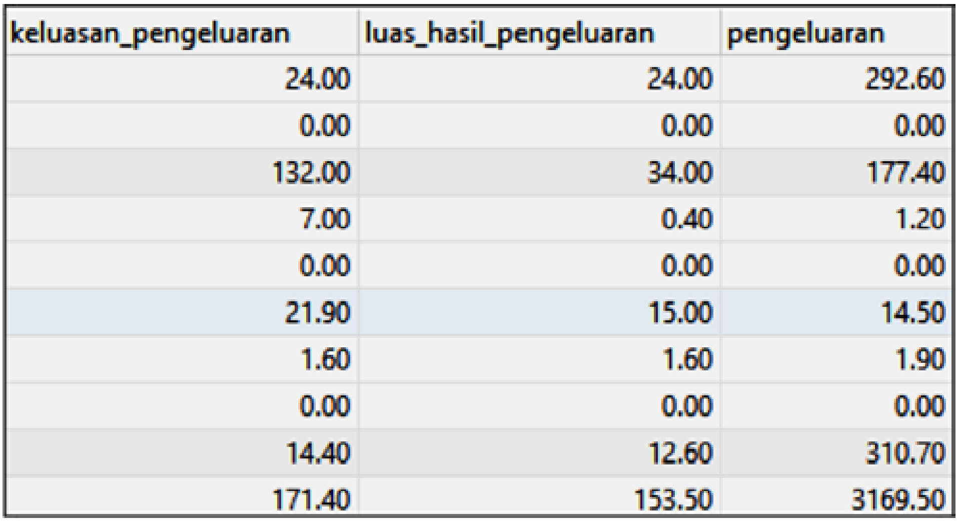
id	tahun	negeri	negeri_latitud	negeri_longitud	daerah
1	2017	JOHOR	1.49265900	103.74135900	BATU PAHAT
2	2017	JOHOR	1.49265900	103.74135900	JOHOR BAHRU
3	2017	JOHOR	1.49265900	103.74135900	KLUANG
4	2017	JOHOR	1.49265900	103.74135900	KOTA TINGGI
5	2017	JOHOR	1.49265900	103.74135900	KULAI
6	2017	JOHOR	1.49265900	103.74135900	MERSING
7	2017	JOHOR	1.49265900	103.74135900	MUAR
8	2017	JOHOR	1.49265900	103.74135900	PONTIAN
9	2017	JOHOR	1.49265900	103.74135900	SEGAMAT
10	2017	JOHOR	1.49265900	103.74135900	TANGKAK

Figure 5(i). Transformed data loaded to the database



daerah_latitud	daerah_longitud	nama_tempatan	jenis_tanaman
1.84944200	102.92883400	Belimbing	buah-buahan utama
1.49265900	103.74135900	Belimbing	buah-buahan utama
2.03006800	103.31846400	Belimbing	buah-buahan utama
1.72937500	103.89922700	Belimbing	buah-buahan utama
1.66298100	103.59993700	Belimbing	buah-buahan utama
2.43091700	103.83611500	Belimbing	buah-buahan utama
2.06305200	102.58487200	Belimbing	buah-buahan utama
1.48692500	103.38896100	Belimbing	buah-buahan utama
2.50346000	102.82075400	Belimbing	buah-buahan utama
2.27109300	102.54198800	Belimbing	buah-buahan utama

Figure 5(ii). Cont. transformed data loaded to the database



keluasan_pengeluaran	luas_hasil_pengeluaran	pengeluaran
24.00	24.00	292.60
0.00	0.00	0.00
132.00	34.00	177.40
7.00	0.40	1.20
0.00	0.00	0.00
21.90	15.00	14.50
1.60	1.60	1.90
0.00	0.00	0.00
14.40	12.60	310.70
171.40	153.50	3169.50

Figure 5(iii). Cont. transformed data loaded to the database

values for the fruit crop hectarage, harvested area, production and value of production can be referred in *Figure 5*(*iii*) in a standardized form of two decimal places.

Implementing the star schema in the crop statistic data cube yielded a positive result as the Strategic Planning and Innovation Management Centre (PI) of MARDI was able to analyze the crop statistic report without a hassle. Furthermore, the analysis process could be done to the granular level when the data cube functionality such as slicing, dicing, rolling up and rolling down was leveraged.

The analysis including the type of crop that had the highest yields for a specific district in a state could be done easily by referring to the fully developed crop statistic data cube. By referring to *Figure 5(i), (ii)* and *(iii),* in the Johor state, the Tangkak district (daerah) had the highest crop yield production at 3169.50 tonnes, followed by Segamat at 310.70 tonnes for the Belimbing crop. This detailed analysis can be further extended to other states in Malaysia or their districts by utilizing these slice, dice and rolling data cube functions.

Conclusion

Building a data cube was crucial before developing a dashboard. Preparing the data in a cuboid manner yielded greater benefits to the stakeholders as analytical processing could be done in a short period of time. Thus, moving from the traditional data structure representation towards the data cube framework has helped the PI and stakeholders in an effective decision making.

References

Albrecht, Alexander and Felix Naumann. (2008). Managing ETL Processes. ACM VLDB ‘08, 24 – 30 Aug. 2008, Auckland, New Zealand

Bhattacharjee, Arup Kumar, Partha Chatterjee, Prasad Shaw and Manomoy Chakraborty. (2014). ETL Based Cleaning on Database. International Journal of Computer Applications. Vol. 105

Davier, Alina A. von, Pak Chung Wong, Steve Polyak, and Michael Yudelson. (2019). The Argument for a ‘Data Cube’ for Large-Scale Psychometric Data. Frontiers in Education 4 (July): 1 – 14. https://doi.org/10.3389/feduc.2019.00071.

Department of Agriculture Malaysia (DOA). (2020). Crop Statistic Report. Retrieved on 15 Jan 2020 from http://www.doa.gov.my/index.php/pages/view/622?mid=239.

Dietrich, Erik. (2017). A Look at the History of RDBMS - Monitis Blog. Retrieved on 29 June 2020 from http://www.monitis.com/blog/a-look-at-the-history-of-rdbms/.

Hall, O.P. (2003). "Using Dashboard Based Business Intelligence Systems - An Approach to Improving Business Performance." Graziadio Business Review 6 (4): 1–12. Retrieved on 13 July 2020 from https://gbr.pepperdine.edu/2010/08/using-dashboard-based-business-intelligence-systems/.

Kay, Russell. (2004). Data Cubes. Retrieved on 13 July 2020 from https://www.computerworld.com/article/2564238/data-cubes.html.

Kimball, Ralph, and Margy Ross. (2013). The Data Warehouse Toolkit : The Definitive Guide to Dimensional Modeling.

Krishna. (2020). Star and SnowFlake Schema in Data Warehouse. Retrieved 22 Sept. 2020 date from https://www.guru99.com/star-snowflake-data-warehousing.html.

"Learn SQL: SQL Views." (n.d). Retrieved on 18 July 2020 from https://www.sqlshack.com/learn-sql-sql-views/.

Putri, A. I., and I. S. Sitanggang. (2017). "Data Cubes Integration in Spatial OLAP for Agricultural Commodities." IOP Conference Series: Earth and Environmental Science 58 (1). https://doi.org/10.1088/1755-1315/58/1/012034.

Rouse, Margaret. (2018). "What Is OLAP (Online Analytical Processing). Retrieved on 16 July 2020 from https://searchdatamanagement.techtarget.com/definition/OLAP.

Abstrak

Penganalisaan data merupakan suatu proses kerja yang rumit jika penyampaian datanya dalam bentuk struktur tradisional. Struktur ini kebiasaannya adalah dalam format buku laporan yang mana datanya dipersembahkan ke dalam jadual dan tidak boleh dikemaskini. Pusat Perancangan Strategik dan Pengurusan Inovasi (PI), mengalami kesukaran dalam menganalisis laporan statistik data tanaman yang diterbitkan dalam struktur tradisional bagi pembangunan papan pemuka. Selain itu, proses analisis data mengambil masa yang lama akan kerana perlu meneliti serta menyemak jumlah data yang banyak pada beberapa halaman laporan. Kiub data dalam konteks sains komputer dikenali sebagai nilai susunan multidimensi. Proses analisis mudah dilaksanakan jika struktur data dibangunkan secara format kiub data. Oleh itu, kajian ini dijalankan bagi menjelaskan peranan kiub statistik tanaman dalam proses menganalisis bagi tujuan pembangunan papan pemuka. Pendekatan yang telah diambil adalah dengan membangunkan kiub data statistik tanaman melalui tiga turutan langkah iaitu: 1) pengumpulan kehendak pengguna 2) reka bentuk pangkalan data dan 3) penyediaan data. Penggunaan pendekatan metadologi tiga langkah ini dapat menghasilkan papan pemuka yang efektif bagi penyelesaian masalah yang dihadapi pihak PI dan memenuhi keseluruhan harapan dan kehendak pemegang taruh MARDI.