International Journal of Advances in Computer Science and its Applications – IJCSIA
Volume 4 : Issue 1 [ISSN 2250 – 3765]
Publication Date : 09 January 2014

The Development of Data Warehouse System for Rice Cultivation

Saisunee Jabjone, Krittika Puannguluam and Chatchai Jiamrum

***Abstract*—This research aims to develop the data warehouse system for applying to a decision support system based on rice cultivation. The study area is Amphoe Phimai, Nakhon Ratchasima Province, Thailand. The data warehouse system for rice cultivation (DWR) aimed to enhance the analysis of rice cultivation datasets compare to currently used statistical methods. The factor of rice growth has been collected which is proving the rice related information to the agricultural expert in an integrated manner or generating a quality advice. DWR would suggest efficient information for rice growth and monitoring the rice situation by using OLAP. It allows for quick analysis of all possible interesting aggregated data by employing drag-drop and mouse-click and is used into forecast the yield of rice. The assessment of user satisfaction was used to evaluate the system. Then this system was published to the government agencies and the public via the online system. The result of the study showed that the users satisfied the system. The mean and SD scores were 4.33 and 0.54 respectively**.

***Keywords*—rice cultivation, data warehouse, decision support system, snowflake schema, OLAP**

I. Introduction

Thailand is one of the world's major agricultural exports in particularly rice product. Thailand is ranked sixth with approximately 22 million tons of paddy rice annually. Thai farmers have around 3.72 million households (16 million people) or 65 percent of total agricultural households (5.76 million households) are in this industry [1]. In domestic, the rice demand has increased every year due to population growth. The domestic consumption is approximately 65% of rice production annually. The annual excess supply, 35% of the total production, goes to the world market [2].

Saisunee Jabjone
Informatics Department
Nakhon Rachasima Rajabhat University
Nakhon Rachasima, Thailand
e-mail : a1102923@hotmail.com

Krittika Puannguluam
Informatics Department
Nakhon Rachasima Rajabhat University
Nakhon Rachasima, Thailand
e-mail : krittika_p@hotmail.com

Chatchai Jiamrum
Information Department
Suranaree University of Technology
Nakhon Rachasima, Thailand
e-mail : cjiamrum@gmail.com

The most produced strain of rice in Thailand is jasmine rice, which is a higher quality type of rice. There are two main growing seasons for rice, the wet season and the dry season. The first crop (or wet season crop) is cultivated from June to August, and harvested during October to January. Hundreds of rice varieties are grown in Thailand. The old varieties are being replaced by new varieties and/or hybrid varieties. Generally, traditional varieties are cultivated in the wet season and new varieties high–yielding usually non–photo sensitive can be grown in the wet as well as the second (dry) season. (February to April and harvested during April to June). The production in the wet season is approximately 18 million tons and 4 million tons from the dry season [3].

In Thailand, the rice cultivation information has collected through several organizations, therefore, it lacks of data integration in particularly in district level. Furthermore, the local governor and farmers do not have enough data for decision-making such as rice yield, rice cultivation area, irrigation management, disaster of flood and drought, total demand and supply. Traditionally, agriculture decision making in Thailand is not data driven, but usually based on expert judgment. Consequently, the researchers developed the data warehouse system for rice cultivation (DWR) for support the decision-making by using data from Amphoe Phimai, Nakhon Ratchasima Province, Thailand. Phimai locates in northeast of Thailand. It is about 300 kilometer from Bangkok. This district is subdivided into 12 subdistricts (tambon), which are further subdivided into 208 villages (mooban). The number of population is 129,849 which majority of population are farmers. There are three seasons (summer, rains and winter).

DWR is a data warehouse that is designed for query and analysis rather than for transaction processing. It contains historical data derived from transaction data. A data warehouse environment includes an extraction, transportation, transformation, and loading (ETL) solution, an online analytical processing (OLAP) engine, client analysis tools, and other applications that manage the process of gathering data and delivering it to business users [4]. It is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process [5]. It is the concept of data extracted from operational systems and stores integrated sets of historical data from multiple operating systems, provides end-user access, enables the knowledge worker to make better decisions or conduct accurate analysis [6]. Thus, the data warehouse is the foundation of Decision Support System (DSS) processing. As such, the access, use, technology, and performance requirements are completely different from those in a transaction-oriented operational environment. The volume of data in data warehousing can be very high, particularly when considering the requirements [7].

II. Materials and Methods

The data warehouse system for rice cultivation (DWR) implements the process to access data sources, clean, filter, and transform the data, and store the data in a structure that is easy to access, understand, and use. The data is then used for query, reporting, and data analysis. This data warehouse provides online functionalities of spatial data analysis. The analysis using OLAP is able to provide information in understandable format to present status and predicts future trends in the concerned domain of rice cultivation. It includes generation of reports and graphs analysis includes generation of reports and graphs as per the specific requirements. Figure 1 shows the process for implementing the DWR.

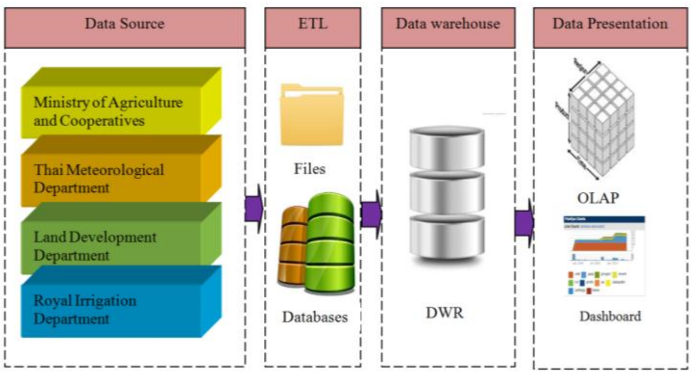

Figure 1: DWR Process

A. Data Source

The rice cultivation data is collected through several organizations in Thailand such as Ministry of Agriculture and Cooperatives, Thai Meteorological Department, Land Development Department and Royal Irrigation Department. In addition, the volunteers from agriculture of Phimai district interviewed the local farmer during their field visits about rice yield, type of rice cultivation, effect of disaster from flood and rainless. The data was collected for 5 year from 2008 to 2013. These databases were implemented on SQL server 2008. Each of these data marts has information related to statistics of rice yield, breed, soil, rainfall, irrigation, climate and area. The validation checks were implemented wherever possible to provide the quality data for the end users. These validation checks are of different forms such as detections of outliers, consistency checks of the data flow and its aggregations from lower level to upper level in a hierarchy, aggregation checks within the same level of hierarchy, etc.

B. ETL

ETL is used to migrate from one database to another, to form data marts and data warehouse also to convert databases from one format or type to another. The rice data extracted from diverse sources will have to be checked for integrity and will have to be cleaned and then loaded into the warehouse for meaningful analysis.

C. Data warehouse schema Architecture

Data Warehouse environment usually transforms the relational data model into some special architectures. There are many schema models designed for data warehousing but the most commonly used are start schema and snowflake schema [8]. For DWR, the preferred structure is the snowflake schema. The snowflake schema is more complex than a star schema. It represents the delivered hierarchy with one table for each hierarchy level. Each table has a foreign key to the level above. The primary key of each table is the member ID of the corresponding hierarchy level [9]. This schema is a normalized representation of the dimension. Dimensions contain the description of data that give a meaning to the numbers contained in the table Facts. The link between the tables "Dimensions" and "Facts" is realized by link-codes between the corresponding tables from the snowflake schema. The DWR schema shows in fig. 2.

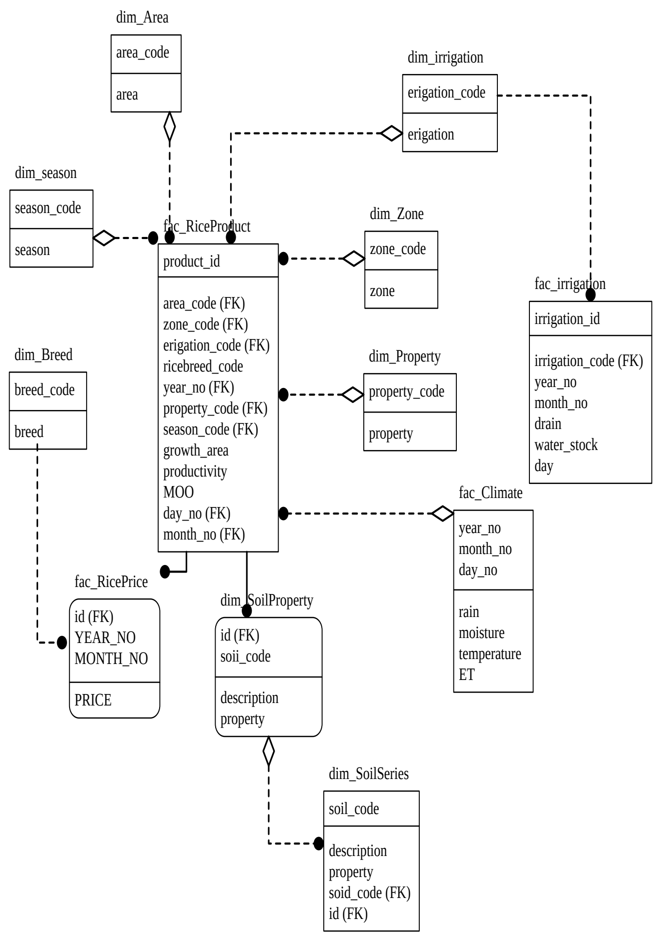

Figure 2: DWR Snowflake Schema

D. Data analysis using OLAP

DWR used the dimensional attributes contained in the table. The data can be summed, counted, and analyzed by any of attributes. The attributes in district level planning and decision support processes require access to data for many different resources such as rice yield, land used, soil, rainfall, temperature, humidity, irrigation and water through. Information on production (demand and supply), price level, and population is also expected. Using OLAP functionality, it supports the user to analyze the data including: trend analysis over sequential time periods, slicing subsets for on-screen viewing, drill-down to deeper levels of consolidation, reach-through to underlying detail data, rotation to new dimensional comparisons in the viewing area (fig 3).

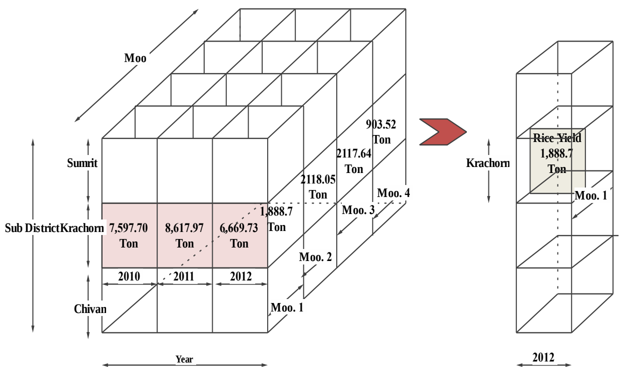

Figure 3: The rice yield per farmer difference season and subdistrict

III. Results

The DWR was developed for three user types, (1) district governor, (2) farmer and (3) officer of agriculture and irrigation. The DWR is developed for store the information and will be sued for analysis and reporting the rice situation in Phimai. To facilitate complex analyses and visualization, the data in a warehouse is typically modeled multidimensionally. In DWR, the user could operation by rollup (increasing the level of aggregation) and drill-down (decreasing the level of aggregation or increasing detail) along one or more dimension hierarchies, slice and dice, and pivot. From a decision maker's perspective, the visualization tools such as chart was used to provide a useful way to view data and information. Outcomes displayed include graphical trend rice cultivation analysis, capacity of rice product in each subdistrict, geographical maps, percentage of rice yield per cultivation area and trend of demand and supply of rice in Phimai. Each chart type user interface design allows presentation of complex relationships and performance metrics in a format that is easily understandable and digestible by time pressured managers. More specifically, such interface designs significantly shorten the learning curve and thus increase the likelihood of effective utilization. Figure 4 shows the average and total rice yield by using geographical map.

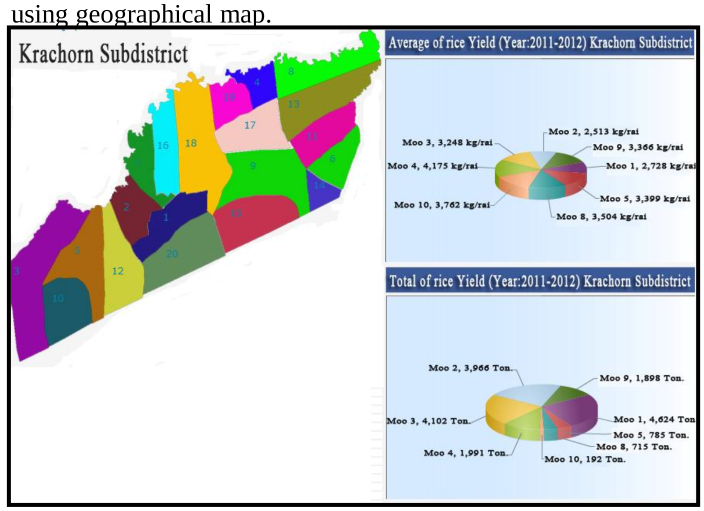

Figure 4: Geographical maps shows the average and total rice yield in Krachon subdistrict

The DWR provides information about the present status and predicts the future trends in the concerned sector of rice cultivation. These can be analyzed using flexible and user friendly features of the cubes such as drill-up, frill-down, slice and dice, etc. The DWR provide user-friendly environment for interactive data analysis. It has functionalities such as creating different graphs, simple statistical calculations from the table rows and columns, swapping of rows and columns with difference view (fig 5).

Year	Cultivated Season	Breed				
Subdistrict	in season				off season	
Moo		Jasmin	Rice No. 105	Sticky Rice		Total Sum of Ric
Moo. 10	770.8	475.66	295.14		265.94	1 036.74
Moo. 11	201.36	136.04	65.32			201.36
Moo. 12	34.24	34.24				34.24
Moo. 13	6.04	6.04				6.04
Moo. 16	18.54	18.54				18.54
Moo. 18	5.08	5.08				5.08
Moo. 19	170.68	170.68				170.68
Moo. 2	1 390.65	1 351.95	7.65	31.05	0.63	1 391.28
Moo. 20	21.42	9.98	11.44			21.42
Moo. 22	9.04	9.04				9.04
Moo. 23	4.44	4.44				4.44
Moo. 25	2.38	2.38			0.96	3.34
Moo. 27	28.38	28.38			8.88	37.26
Moo. 3	2 823.92	755.59	2 068.33			2 823.92
Moo. 4	1 988.38	535.96	1 452.42		12.01	2 000.39

Figure 5: Drill-down chart shows the capacity of rice product in each subdistrict per year and season

The data can be sorted in the table element by selecting a particular row or column and selecting ascending/descending orders. DWR also has option for identifying major contribution rows or columns in the marginal totals, user can highlight exceptional values of the table through automatically as well as user defined criterion. The standard bars and charts can be displayed on the screen by simply clicking the respective options. In addition, this system has facility to store history of the analysis performed by a particular user. It presented by using simple bars, pie diagram, pie diagram, multi-line graphs, 3-D bars, dashboard, etc. The graphs and analysis performed through OLAP cubes can be directly exported for report preparation. Tabular data displayed on the screen can be exported to popular formats for further analysis. The example of graph shows in fig 6.

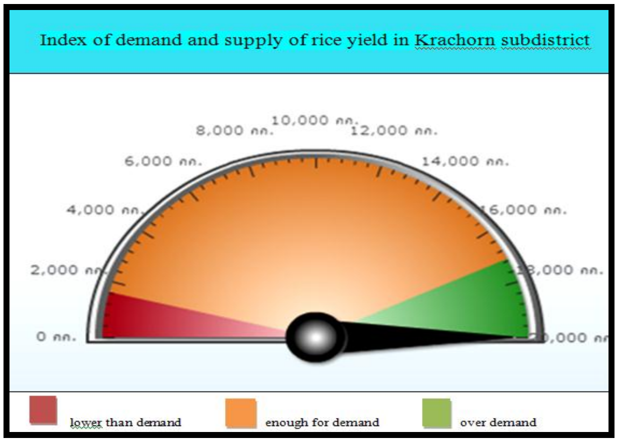

Figure 6: Graph shows the index of demand and supply of rice yield in each subdistrict

The training workshop was held for imparting hands-on training of DWR tools to the end user. The forty participants from Phimai district participated with this workshop. The participants were given the questionnaire for evaluating the satisfaction of the system. The measure for DWR satisfaction was information accuracy, format and preciseness, GUI, access time, and fulfillment of end-user needs. As the DRW is needed to maintain huge volumes of data for quick access; thus, the efficiency and response time are crucial, particularly while retrieving data from the warehouse. From these participants, 3.80% are very satisfied with the system while 69.20% are satisfied. The rest (26.90%) are not satisfied. The participants were then divided into the two categories of clients; farmers and governor. In the governors' category, 13% are very satisfied and more than half are satisfied while the rest are not satisfied. In the farmers' category, 65% are satisfied and the rest are not satisfied. The evaluation of all participant satisfaction found that majority of them satisfied the system. The mean and SD scores were 4.33 and 0.54 respectively. From these results, we observe that the initial implementation of DWR is successful. However, there is much scope for improvement of the system as seen from the feedback and results of the user study.

IV. Discussion and Conclusion

Increasing food demands due to high rates of Thai population growth and major changes in Thai economic and Thai social systems that have created an urgent need to develop new and revise many existing agricultural systems and practices. Regional Productivity analysis involves evaluating spatial soil and weather variability, identifying optimum rice management practices, and predicting productivity of the region under different climatic and management scenarios. Consequently, the DWR is developed for store the information and will be sued for analysis and reporting the rice situation by using a case study in Phimai district. Data was kept in the warehouse which farmer, district governor and officer who involve the data may use and get a benefit from this system. A major component of a DWR is the ability to accurately assess the condition of rice growth and rice yield to provide appropriate strategies or countermeasures. It can help local governor to rice cultivation monitoring, early warning of rice situation such as over supply and more demand. This analysis can help regional planners and policy makers in delineating acreage and distribution of areas with high productivity and developing management recommendations for rice cultivation. In addition, it could give the great information for managing the irrigation system. However, this system has still some limitations such as limited statistical data about the rice yield in last 10 year. Most of monitoring data are to be interpreted manually, therefore, it might need the expert to interpret the result. In finally, the data in the system was only validated in Phimai. It should expand to the other area for more advantage in the future.

Acknowledgment

This research was made possible by the funding of Nakhon Ratchasima Rajabhat University and Research and Development Institute. I would like to thank Mrs. Parichart who provided raw data and contacted to district governor. Thank you to officer of Ministry of Agriculture and Cooperatives, Thai Meteorological Department, Land Development Department and Royal Irrigation Department. who supported the data, gave the advice and tested the system.

References

[1] Bank of Thailand, Agricultural Report 2009, Bangkok, 2010.

[2] Thawatchai Dechachete, Office of Agricultural Economics, Government of Thailand, Bangkok, 2011.

[3] Sayamwara, A. and Na Ranong, W., Thailand Development Research Institute. Compilation of Rice Knowledge, Bangkok, 2010.

[4] Inmon, B., Building the Data Warehouse. Fourth Edition. John Wiley & Sons, Inc., New York, 2005.

[5] Kelly, S., Data warehousing in action. Wiley, Indianapolis, 2007.

[6] Kimball, R., The Data Warehouse Toolkit : The Complate Gride to Dimensional Modeling. John Wiley & Sons, Inc., New York, 2002.

[7] Silvers, F., Building and maintaining a data warehouse. CRC Press, Boca Raton, 2008.

[8] Sherman, Rick., Beyond ETL and Data Warehousing. InfoManagement Direct, 2009.

[9] Sen, A;, Sinha, A.P., A comparison of data warehousing methodologies, Commun, ACM 48(3), pp. 1-11, 2005

About Author (s):

Saisunee Jabjone is an Assistant Professor at the department of Informatics, Nakhon Ratchasima Rajabhat University.
Her research interest is in the area of data mining, data warehouse, ontology, data management, database, bioinformatics, and education technology.

Krittika Puannguluam is a lecturer at the department of Informatics, Nakhon Ratchasima Rajabhat University. She is interested in computer networks, adhoc networks, data warehouse, data mining and other fields of information technology

Chatchai Jiamram works as a system analyst at Suranaree University of Technology. His research interest is in the area of data warehouse, data mining, data management, simulation, modeling and Artificial intelligence.