Design and Development of Data Warehouse Framework of Highland Vegetable Crops for Benguet

A R Quitaleg and M G Ortiz
University of the Cordilleras, Baguio City, Philippines

oliangelanne@yahoo.com, mariegraceortiz@gmail.com

Abstract. The data warehousing technology has been widely used in organizations providing management, data integration and helps do the decision-making. Data warehouse offers an effective way for analysis and statistics to the big data in agriculture. Agriculture is one of the major issues in the Philippines and needs technological advancement. There is an increasing need for farmers and decisions makers in processing agricultural data. Agricultural datasets are diversified, non-standard and very large. It captures information about crops, farmers, buyers and other factors and requirements pertaining to the agricultural production. Designing a data warehouse for agricultural data is a great challenge for the researchers. 

1. Introduction

The world population grows rapidly but the resources for crop production continuously diminish. There is a big shortfall between the amount of food we produce today and the amount needed to feed everyone in 2050 [1]. There will be nearly 10 billion people on Earth by 2050.

In recent years the growth rates of world agricultural production and crop yields have slowed. This has raised fears that the world may not be able to grow enough food and other commodities to ensure that future populations are adequately fed. In the Philippines, according to Philippine Statistics Authority (PSA), agricultural production grew at only 0.56% in 2018, greatly missing the target of 4% for the year. This was a sharp drop from 2017’s record of 6.69% [2].

The core question is whether today’s agriculture and food systems are capable of meeting the needs of a global population. The international community has recognized these challenges. In particular, the 2030 Agenda for Sustainable Development, provides a compelling, but challenging, vision on how multiple objectives can be combined to define new sustainable development pathways. The second Sustainable Development Goal explicitly aims at ending hunger, achieving food security and improved nutrition, and promoting sustainable agriculture, simultaneously by 2030 [3].

In addition, the World Health Organization (WHO) reported fruit and vegetables are important components of a healthy diet, and their sufficient daily consumption could help prevent major diseases. Approximately 16.0 million (1.0%) disability adjusted life years and 1.7 million (2.8%) of deaths
worldwide are attributable to low fruit and vegetable consumption. Thus, WHO recommends 400 grams per day of fruit and vegetables per day for the prevention of chronic diseases and alleviation of several micronutrient deficiencies, especially in less developed countries.

High Value Crops Development Program (HVCDP) has identified top priority vegetables and these are the lowland crops, highland crops, and export crops. The study will focus on the highland vegetable crops. Benguet is popularly known as the "Salad Bowl" of the Philippines because virtually every type of vegetable used in salads is produced here.

To satisfy the increasing demand for food particularly highland vegetables, the use of new farming techniques must significantly increase crop yield, thus increasing agricultural income. Considering that vegetable farming provide employment for the province and its neighboring provinces.

With technological advances in the area of information and communication technology (ICT), the agricultural data processing is an increasing need for farmers and other decision-makers. Agricultural data is varied, non-standard and there are islands of information. Decision-makers including farmers, cooperatives and other agencies need to collect, consolidate and summarize data from various sources to fulfill their needs and meet the increasing demand of vegetables. Due to these circumstances, data warehouse (DW) is an essential need for agriculture. Data from each cooperatives, organizations, farmers, buyers, and agencies must be stored in a common database and used for comprehensive data analysis. This data is used to inform important business decisions. In order to make good decisions, relevant data should be taken into consideration and the best source for that data is a well-designed data warehouse.

This study focuses on the design and development of data warehouse framework of highland vegetable crops for Benguet. Building data storage will be used in the agricultural data analysis process that can be used to facilitate decision-making on farming and marketing. It aims to increase the volume of commodities, better farm management, and better marketing system.

2. Literature Review

Big data analytics can help stakeholders design and implement predictive models based on crop type and algorithm for farm optimization. Big Data is defined [4] as high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision-making, and process automation.

The paper will focus on the predictive analytics, which aim to something about “what might happen next?” It involves extrapolating trends and patterns in the future. By considering historical data, making smarter decision about the future in present is called predictive analysis model. According to the study conducted by [5], this model encompasses different patterns from historical facts to analyze the future predictions. This is a type of data mining technique that derives results from previous year data and predicting the behavior using history of data. This includes analyzing what event had occurred in the past, when it is occurred, and monitoring current terminologies and facts about data, then making smarter decision using machine-learning methods. In prediction, previous year data are referred as training set and the data's are classified based on training set [6].

One of the most significant bodies of work for big data is the project named SatSure founded by the 33-year-old Abhishek Raju works on using Big Data and its allied technologies like data science and IoT to better the lives of farmers. Abhishek shares the fact that he was deeply moved by the rate of farmer suicides and the lack of application of science and newer technologies in the oldest Indian industry – agriculture. His solution to this is SatSure [7].

Another notable contribution is The Yield, as stated by [8] in her article entitled “Transforming the agriculture industry using IoT and predictive analytics”. The initiative uses data, sensing technologies, and analytics to solve problems for Tasmania's agriculture ecosystem.

The above-mentioned projects mentioned did not consider building a data warehouse for the organization to support their decision-making.

A paper entitled “An Efficient Data Warehouse for Crop Yield Production” [9] was proposed for precision agriculture. It presented system architecture and a database schema for design and
implementing a DW. The authors used a constellation schema or galaxy schema because it focuses on various areas in farming such as treatment, operation, trading, and yield. However, in the paper, it did not address a specific issue of a particular organization.

Also in the paper [10] “Indian Agriculture Data Warehouse Design employed the star schema data model, however, it concerns only on time and location dimensions. The DW system provides spatial analysis of data with the help of Geographic Information System (GIS). However, the DW concerns livestock farming, not for agricultural crop production.

The need for integrated information is the biggest driver for investing in data warehouse systems, because it represents both operational and strategic business needs to support day-to-day business operations. With these, the researchers would like to propose a DW design to increase the agricultural production of highland vegetable crops in Benguet.

3. Design and Development of Data Warehouse Framework

The country’s biggest agricultural trading center is in La Trinidad, Benguet. Benguet Agri-Pinoy Trading Center (BAPTC) offers research and extension and instruction expertise to further develop the vegetable and root crop industry of Benguet. The increase in the trading operations of the BAPTC has driven more farmers, buyers, market facilitators, packers and porters from different wholesale market sites to transact at the facility. To maximize its potential as catalyst of improving the lives of highland vegetable farmers in the Cordilleras, the center needs good management to efficiently handle both the prospects and challenges of the project. With these, BAPTC, and its accredited farmer groups and individuals need to document all transactions. However, they have not yet utilized technology as a solution to these problems but they are willing to adopt a new technological solution.

One of the most important assets of the center is its information. The goals of this study are to 1) maintain historical information, 2) facilitate reporting and analysis, 3) be the foundation for better decisions making, and 4) be an adaptive source of information. This study is also intended to offer solutions to help BAPTC achieve its mission and vision, “Modernizing farmers' access to global markets” and “Building avenues for farmers and consumers to have better market access through participatory, capability building and provision of innovative technologies and quality facilities.”

Data used by BAPTC comes from various sources such as accredited farmers group and individuals, other trading posts which has many forms and formats, structured and unstructured, and they wanted to use this data and convert it into useful information by integrating and storing it in a single location. Data warehouse design is the process of building a solution to integrate data from multiple sources that support analytical reporting and data analysis [11]. The researchers would like to propose a data warehouse framework for the BAPTC so that all information about the center’s transactions and activities are consolidated in one place.

3.1. Data Warehouse Architecture

The researchers considered the system architecture of the various technical components and data architecture of the information stored in the DW. Data architecture refers to how data is organized within the warehouse.

The three-tier architecture [7] was used in this study composed of bottom tier, middle tier, and top tier. In the bottom tier, data from multiple sources are cleansed, extracted, transformed, and loaded into the database of the data warehouse serves. The middle tier is an application tier that presents an abstracted view of the database using OLAP. Top tier is the front-end client layer that uses query, reporting, analysis, and data mining tools to extract data from data warehouse. Figure 1 shows the data warehouse architecture for BAPTC.

The end-users of the data warehouse are the BAPTC employees and cooperative’s staff. In the data presentation layer, users can interact and access data for business insights for ad-hoc analysis, reports and dashboards. The top-down approach is where the data warehouse is designed first and then data marts are built on top of data warehouse.
3.2. Business Process
Understanding the business requirements with an understanding of the available source of data and identifying what business process to model is the first step in the design. In BAPTC, the center wants to better understand management monitor the volume of crops produced by a farmer at a particular field and its selling/buying price to wet market buyers and institutional buyers.

These information is critical to the center to meets its objectives namely 1) A multi-stakeholder approach to sustainable linkage of farmers to market, and 2) to increase the farmers' incomes and protect them from the effects of market price fluctuations.

BAPTC identifies farmers and buyers who will compose the market. It provides processing and storage facility to farmers and area accessible to traders. The center determines the commodity and volume requirement to match the projected of the market with the production capacity of farmers. The logistical cost of moving the commodities from farm to the center and from the center to the market outlets and consumers are also monitored. BAPTC gathers historical and actual production data. The center also identifies potential crops.

3.3. Grain
After the business process has been identified which include monitoring volume and prices of commodities, monitoring buyers and sellers, supervising on crops produced by farmers, and other business processes in the BAPTC, the researchers face a serious decision about granularity. Declaring the grain is central and important step in a dimensional design. The grain establishes exactly what a single fact table row represents. The grain must be declared before choosing dimensions. Atomic data is highly dimensional [12]. In BAPTC, the most granular data is the daily production of commodities of each farmer in each crop on a particular season.

3.4. Dimensions
After the grain of the fact table has been chosen, the choice of dimensions is straightforward. Dimension tables are sometimes called the “soul” of the data warehouse because they contain the entry points and descriptive labels that enable the DW system to be leveraged for business analysis. Dimensions can be attributed to the production of commodities measurements, such as the farmer who produced the commodity, the type of crop produced, the facility used by the farmer, the, the name of the buyer/seller who purchased the commodities, the season when the crop is produced, the BAPTC staff who assisted in the transaction, the coop or group of the farmer, and the place where the crop is harvested. The following descriptive dimensions apply to BAPTC and its descriptive attributes of each dimension are described in Table 1.

3.5. Facts
The fourth and final step in the design is to make a careful determination of which fact will appear in the fact table. The term fact represents a business measure [12]. Facts are determined by answering the question, “What is the process measuring?” The grain declaration helps in the process.
In this case, the daily production of commodities of each crop will be the fact that will include the volume of the commodity produced and purchased, the unit price of the commodity, and the total amount of the commodity produced and purchased. The fact table in this case is presented in Table 2.

### Table 1. Dimension Tables

| No | Dimension Table | Descriptive Attributes                                      |
|----|-----------------|-------------------------------------------------------------|
| 1  | Crop            | cropID, crop type, crop name                                |
| 2  | Farmer          | farmerID, lastName, firstName                               |
| 3  | Buyer/Seller    | buyerID, name, address, contactNo                          |
| 4  | Facility        | facilityID, facility name, facility type, facility rental fee|
| 5  | Season          | seasonID, season name, season month start, season month end |
| 6  | Staff           | staffID, staff lastName, staff firstName                    |
| 7  | Date            | dateID, date, quarter, weekly indicator                    |
| 8  | Location        | locationID, location barangay, location municipality        |
| 9  | Field           | fieldID, field name, locationID                            |
| 10 | Truck           | truckID, truck type, plate no.                             |
| 11 | Coop            | coopID, coop name, address, contact no.                    |

### Table 2. Production Fact table

| Production Fact | dateID (FK) | cropID (FK) | farmerID (FK) | fieldID (FK) | buyerID (FK) | staffID (FK) | facilityID (FK) | volume or weight of commodity | price per unit | total amount |
|-----------------|-------------|-------------|---------------|--------------|--------------|--------------|-----------------|-----------------------------|----------------|--------------|

#### 3.6. Snowflake Schema

After defining requirements, the next step is to determine how data structures will be available, combined, processed and stored in the data warehouse. This process is known as data modeling [12]. DW design and data sources as defined in this step.

The proposed DW design employs snowflake schema. A snowflake schema is a variation on the star schema, in which very large dimension tables are normalized into multiple tables. Figure 2 shows the data warehouse snowflake schema and how data from the data warehouse will be used for BAPTC. Product Fact is connected to multiple dimensions represents the snowflake schema. Dimensions such as Crop, Field, Buyer, and Farmer are normalized into multiple related tables. In the figure, field dimension is elaborated into location dimension to understand on what Barangay and Municipality a crop is produced. The crop dimension is elaborated into season dimension to determine on what season a crop is best produced. The farmer dimension is extended into two (2) relationships to understand what truck farmer in the trading areas and on what cooperative used or organization the farmer belongs to.

A data warehouse contains data that is used to support business decisions which can be used in various ways such as with data visualization tool to provide reports and dashboards; with advanced data analytics such as forecasting, data mining, sentiment analysis; and can be used as a source of data to other systems such as supply chain management and marketing.
4. Conclusion and Recommendations

A data warehouse design was presented as a solution to agricultural data issues. For BAPTC to operate efficiently and achieve their goals, from decisions-makes, farmers to individual contributors need efficient access to data and analytics that provide insights into how the center is operating, areas of concern, and opportunities for competitive advantage. By using snowflake approach in the DW design will save space. Data from DW can be utilized for data mining, forecasting, predictive analysis, accurate reports, and informative dashboard. This will help management resolve various complex queries as per their requirement. Further, this research could be extended to develop predictive models to increase the yield of highland vegetable crops in Benguet.

References

[1] Ranganathan J, Waite R, Searchinger T and Hanson C 2018 How to sustainably feed 10 billion people by 2050 in 21 charts (Washington DC: World Resources Institute)
[2] Rivas R 2019 Agriculture grows a meager 0.56% in 2018, misses target Retrieved from: https://www.rappler.com/business/221717-agriculture-growth-philippines-2018
[3] United Nations Sustainable Development Goals
[4] Bayer M and Laney D 2012 The Importance of ’Big Data’: A Definition (Gartner, Inc)
[5] Kumar P, Sarangi A, Singh DK, Panhar S S and Sahoo R N 2015 Simulation of salt dynamics in the root zone and yield of wheat crop under irrigated saline regimes using SWAP model Agricultural Water Management vol 148 C pp 72-83 DOI: 10.1016/j.agwat.2014.09.014
[6] Gupta S 2018 Data Science in the Indian Agriculture Industry May 28th 2018
[7] Guru 2019 Data Warehouse Architecture, Concepts and Components Retrieved from https://www.guru99.com/data-warehouse-architecture.html
[8] Bindi T 2017 Harnessing IoT in the Enterprise: Transforming the agriculture industry using IoT and predictive analytics February 1st 2017 Retrieved from https://www.zdnet.com/article/transforming-the-agriculture-industry-using-iot-and-predictive-analytics/
[9] Ngo V M, Le-Khac N and Kehchadi M 2018 An Efficient Data Warehouse for Crop Yield Prediction Proc. of the 14th Int. Conf. on Precision Agriculture

[10] Rai A, Nilakanta S and Scheibe K 2007 Indian Agriculture Data Warehouse Design Managing Worldwide Operations and Communications with Information technology pp 535-8

[11] Bui A 2017 Data Warehouse Design: The Bad, The Good, The Ugly Retrieved from https://blog.panoply.io/data-warehouse-design-the-good-the-bad-the-ugly

[12] Kimball R and Ross M 2013 The Data Warehouse Toolkit: The Definitive Guide to Dimensional Model 3rd edition (New Jersey: John Wiley & Sons, Inc)