A DATA WAREHOUSE FOR UNCONTROLLED DECISION MAKING AND DECISION MAKING IN GREENHOUSES

Daniele Aparecida Antunes, E-mail: daniele.alves.batista@hotmail.com
Aline da Silva Souza, E-mail: alinedassouza@gmail.com
Luan Ricardo Gonçalves França, E-mail: luanlrgf7@gmail.com
Emili Everz Golombiéski, E-mail: emilieverz043@gmail.com
Maria Salete Marcon Gomes Vaz, E-mail: salete@uepg.br
Universidade Estadual de Ponta Grossa (UEPG)
Programa de Pós-Graduação em Computação Aplicada

Abstract. *Agriculture has been studied by scientists using empirical methods to analyze the low profitability of production and which factors predominate for this purpose. To improve agricultural production, agricultural greenhouses were created, which are closed environments and allow for the control of conditions necessary for the development of plants, such as temperature, humidity, and light. A greenhouse allows the construction of models and systems that help to improve production and crop quality. Temperature, humidity, and light sensors are examples of components that receive information from the environment and enable greater control of these variables. The readings taken by sensors in greenhouses bring a wealth of information that could be stored and used to assist in decision support and long-term greenhouse improvements, not just be verified in real-time. One way to manage large volumes of data is through the use of Data warehouses (DW), which store summarized, aggregated, and consolidated data for long periods. Based on this reality, this article presents an analysis of a DW that enables strategic support in the planning and control of greenhouses, storing information on temperature, humidity, greenhouse luminosity, as well as external factors such as weather conditions, investments, costs, and productivity.*

Keywords: *Greenhouses, Data Warehouse, Decision Making.*

1. INTRODUCTION

Greenhouses are closed environments that enable the control of conditions necessary for the development of plants, such as temperature, humidity, and light. Temperature, humidity, and light sensors are examples of components that receive information from the environment and enable greater control of these variables. Actuators, in turn, act on the environment, changing natural conditions, such as heaters and drippers, helping with humidity and temperature rates. The control of these conditions may present difficulties in its stabilization, in Dos Santos (2017) the development of a prototype of temperature control in an oven can be verified, comparing PID and Split Range control techniques for the adjustment of temperature in ovens, in addition to displaying this data in mobile device software.

The readings taken by sensors in greenhouses bring a wealth of information that could be stored and used to assist in decision support and long-term greenhouse improvements, not just be verified in real-time. One way to manage large volumes of data is through the use of Data warehouses (DW), which store summarized, aggregated, and consolidated data for long periods.

DW is a computer system used to store data related to an organization's activities, in the form of historical series, enabling analysis of past events and helping to support decision-making in future events (MACHADO, 2000). In NGO et al. (2019) a DW is developed combining Hive, MongoDB, Cassandra, and a constellation schema using real agricultural datasets. DW was developed using characteristics necessary for use in agricultural big data, such as high storage capacity, performance, and speed, storage structure flexibility, data ingestion, pre-processing, governance, monitoring, security, and veracity. The actual data recorded took into account information about crops, locations, fields, farmers, fertilizers, field facts, and operating time. Complex queries performed in DW showed 3.19 times faster speed when compared to queries in Mysql.

With the volume of data generated by sensors in greenhouses, a DW could be used to store this information, helping in decision making and improvements. The data would be stored for long periods and used for analysis, such as checking for interference from external environmental conditions inside the greenhouse, changing the functioning of sensors and actuators, affecting costs such as electricity consumption, or the need to exchange components. Therefore, this research proposes the development of a DW that enables strategic support in the planning and control of greenhouses, storing information on temperature, humidity, greenhouse luminosity, as well as external factors such as weather conditions, investments, costs, and productivity.

2. BACKGROUND REVIEW

In this section, the concepts relevant to this article are described, approaching aspects related to the characteristics of agricultural greenhouses and Data Warehouse concepts as a support in decision-making control.

2.1 AGRICULTURAL GREENHOUSE

For Reis (2022), greenhouses in agriculture are an instrument of environmental protection for the production of plants, such as vegetables and flowers. They are structures built with different materials, such as wood, concrete, iron, aluminum, etc., covered with transparent materials that allow the passage of sunlight for the growth and development of plants. The use of these structures can be partial or full, depending on the characteristics explored. A typical example of partial use is the use of covering the structure to obtain the umbrella effect, very common in tropical regions. On the other hand, it is possible to explore the full potential of this type of structure, building a complete greenhouse, with all the controls for the coverage and protection of plants about adverse meteorological parameters, such as rainfall, and with side curtains for generation and trapping of heat. In the latter case, the greenhouse effect of this structure is used, which is why the greenhouses are better known as greenhouses.

In the 19th century in Leiden, Holland, the first greenhouse was built by the botanist Charles Lucien Bonaporte. The first greenhouses had the job of preventing the orange trees from freezing, so the first greenhouses were called “orange trees” at the time. Greenhouse structures were improved through studies starting in the 1960s with larger and stronger sheets of polyethylene film becoming more durable (NNADI and IDACHABA, 2018).

Greenhouses can also be installed in places where it is not possible to plant some types of food due to climatic conditions that do not favor the development of plants. From this, it is assumed that the intermediate period between one harvest and another, subsequent of a product, does not happen, guaranteeing the necessary growth throughout the year. The raw material used to build a greenhouse is responsible for protecting the plants and the environment as a cover, and the material can vary, this material can be from the simplest to the most elaborate, storing the heating energy of the plants. The most common material is polyethylene, which is used in tropical areas, as it has the advantage of sustaining mechanical and thermal stresses caused by sunlight and other adverse weather conditions (NNADI and IDACHABA, 2018).

Many elements can compose a greenhouse, thus constituting an integrated system, where energy flows and information are generated efficiently with different significances for decision making. The cultivar present inside the greenhouse is a determining component, as it is the element that is affected by climatic factors, physical structure, and environmental handling. All conditions combined can interact with each other, increasing the level of difficulty of the various systems, significantly changing the plant's development cycle (RODRÍGUEZ et al., 2015).

Greenhouses are closed environments that offer a sustainable and agile development of plants over time, enabling the control of conditions necessary for the development of such as temperature, humidity, and light. Temperature, humidity, and light sensors are examples of components that receive information from the environment and enable greater control of these variables. Researchers have worked with systems that obtain control of factors that can affect plant development, for example, the water sprinkler and irrigation system (DANITA et al., 2018).

The correct control and monitoring inside a greenhouse define the life and production of the cultivars. For the control and monitoring to be efficient, it is necessary to implement a system capable of generating the functions as a support in decision-making in the greenhouses. So, for the greenhouse cultivars to be efficiently monitored, the environment in which it is found must be automated, capable of generating accurate and real-time information for decision-making. With the volume of data generated by sensors in greenhouses, a DW for example could be used to store this information, helping in decision making and improvements. Information from a DW can minimize errors caused by human interventions.

2.2 DATA WAREHOUSE CONCEPTS

The Data Warehouse emerged in the 1990s with the proposal to solve the access to managerial information within an organization. As at that time there were already a large number of business management systems, consequently the volume of data also grew exponentially. Databases evolved to meet technological demand, but at the time it was necessary to work in a context of distinct data and unite them externally. Then, through an academic project of 1980, it started in 1990 with the first tests and analyses with the Data Warehouse (CALAZANS, 2003).

The name of Data Warehousing is given to the systematic process of building a DW system, consisting of a set of technologies, algorithms, techniques, and elaborate construction, contributing to the storage and management of large volumes of data, with the proposal of making it easier for executives and analysts to understand the business (CHAUDHURI; DAYAL, 1997; GARDNER, 1998 apud CALAZANS, 2003).

The DW can be called a data organization technique aimed at managing the techniques of institutions, not having the purpose of controlling the business. DWs have summarized, aggregated, and consolidated data that are stored for long periods and are a way to manage very large volumes of data that are generally spread across different systems of an organization. The DW of corporations are built from different managerial databases of an organization (CALAZANS, 2003).

The concept of DW corresponds to structuring corporate data in an integrated manner, providing the institution with the differentiation of propensities, assisting in decision-making through the information generated, and consequently, increasing its profits and results.

DW provides a general data model for different areas of interest, regardless of the data source, making the analysis of information more efficient. Errors from different data sources are stipulated before the information is loaded into the DW, thus making the process of building reports and analysis simpler and more objective.

One of the purposes of DW is to generate information about an institution that is easily accessible. For Kimball (2002) the information generated must be accurate to end-users, with understandable, didactic, and reliable data. According to Machado (2000), one of the characteristics of DW is the flexibility and adaptability associated with changes and the level of security that projects data. Other advantages of features that the author cites are:

- Availability of information for the management of corporations;
- Generation of behavior curves;
- Agility in the use of decision support tools;
- Security of information for decision;
- Greater scope in the view of indicators.

Over the lifetime of a DW, there can be a high cost of storage and it can quickly become outdated, which can bring great losses to the organization. In addition to this disadvantage, during the DW process, data must be extracted, transformed, and loaded into the warehouse, presenting a latency period (Boateng et al. 2013).

3. Methodology

Taking into account the characteristics necessary for precision agriculture in greenhouses, a star model was developed using the Lucidchart diagram building site.

The sensor table represents the possible sensors of a greenhouse, storing information such as type, measurement, and identifier. With the growing diversity of sensors, it is important to separate the types of sensors, that's why the sensor table was developed, defining for each sensor a type and the measurement that is measured. In the same patterns, we have the actuator table that stores the actuator identification information, type, and form of action in the middle.

As we need to have the data cataloged by id and date, the date table is represented, so that the user can filter by time, day, week, month, quarter, or/and year, the date table being primordial for subsequent database queries and the possible creation of models that take into account the seasonality of cultivars.

As this is a crop-oriented DW, the crop table is of paramount importance, which includes the crop identifier, the foreign key of the crop table, which will be explained below, the crop planting date, the date of harvest. The location of the greenhouse is taken into account if a farmer has more than one greenhouse, or wishes to identify different areas within the same greenhouse, so it is possible to catalog not only the crop but also the place where the cultivar was planted, for further evaluation of productivity versus location in the greenhouse where the cultivar was. Investment is another variable considered in the harvest table, as well as expected return and actual return, these three variables are for analysis of monetary values, as the data is collected year by year, the farmer can have a broad view of points of improvement, through sensors and actuators.

For each crop, there is a cultivar, this being said, the need for the crop table, which is related to the crop, is plausible, this table contains the unit price paid for seeds, seedlings etc., the type of unit such as bags of seeds, trays of seeds seedlings, among others., the date of acquisition, as the value may vary depending on the date, the type of crop, consisting of oilseeds, fruits, vegetables, and other types. Finally, the table has the name field, where the name of the culture is described.

The facts table was named as measures, bringing as foreign keys the identifying values of the tables: Sensor, Actuator, Crop and Data. The attribute measure value aims to store the values verified by the sensors or actuators, while the variable electrical energy consumption records the energy consumption data of sensors and actuators when activated in an instance. The tables and attributes of the developed model can be seen in Figure 1.

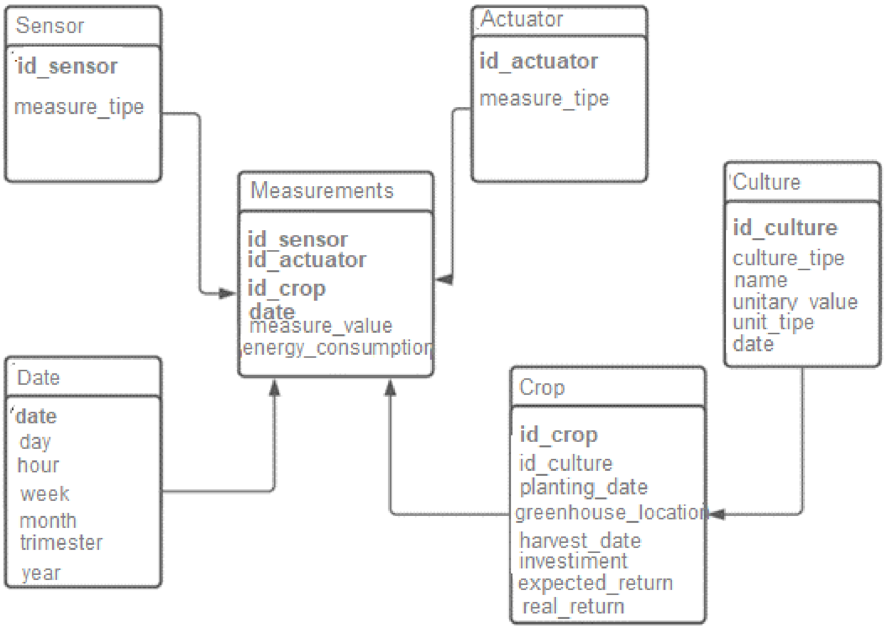

Figure 1: Attributes table (Source: Authors)

4. RESULTS

The presented model aimed to provide a structure for recording data that could help a producer in the management of greenhouses on a property, encompassing issues such as investments, sensor and actuator control, monitoring of environmental conditions and crops.

From queries in the proposed Data Warehouse, the producer can check the conditions presented by the greenhouse. Information about the investment and return values can help the owner in managing his resources, identifying crops in which there was a greater financial return, and the efficiency in the use of electrical energy by sensors and actuators.

The recording of measurements read by the sensors can help in their maintenance, as the change in the way it behaves under some condition could be verified, changing the values read, and enabling an estimate of the component's useful life over the seasons. The information about the action of the actuators is also relevant, taking into account the time an actuator would need to remain on to fulfill a goal, such as irrigation or heating, its wear could be noticed if there was an increase in the period spent for the execution of the assignment.

The saved historical periods could be used to verify how external environmental conditions interfered with the sensor reading and actuator action. A period of colder weather may require the use of a heater for longer, increasing energy expenditure, while adverse environmental conditions may affect the reading of information by a sensor.

An operation that can be stored through the Data Warehouse is the use of energy in the greenhouse with high technology, capable of proposing energy savings based on variable speed ventilation, where a controlled induction motor system (DTC) is used. Torque Control) as a robust and efficient control and a photovoltaic generator (PVG from English Photovoltaic Generator) feeding the induction motor.

The power produced by the photovoltaic generator during the day reaches 980 watts (W), which is enough to power the variable-speed ventilation system that requires only 705 (W). Therefore, the system contributes to significantly reducing the use of the electrical network, as well as achieving a photovoltaic energy gain with an average of 300 (W/day) that could be stored or used for other purposes. Simulation results of this abatement system can achieve a significant amount of energy saved by drastically reducing the costs of greenhouse farming agriculture (MAHER et al., 2016).

5. CONCLUSION

The proposed model presented a data recording structure that can assist the producer in managing a greenhouse, providing access to information about sensors, actuators, crops, and historical records of facts. There are advances in research on the use of information technologies and automatic control in greenhouses, aiming at the economy, quality, and efficiency in supporting decisions in agricultural production.

Future works can use this model as a basis for the development and application of a data manipulation system, in addition to carrying out tests in the storage of information, checking possible redundancies, and optimizing the model. Information about the ways of cultivation and treatment can be added to the model.

ACKNOWLEDGEMENTS

To CAPES for the financial support.

REFERENCES

D0S SANTOS, O. L. (2017) **Sistema de Controle de Temperatura para uma Estufa com Monitoramento via Aplicativo**. RCT - Revista de Ciência e Tecnologia, [S.l.], v. 3, n. 4, july 2017.

MACHADO, F. (2000) **Projeto de data warehouse: uma visão multidimensional**. São Paulo: Érica, 2000.

NGO, V. M.; LE-KHAC, N. And KECHADI, M. (2019) **Designing and Implementing Data Warehouse for Agricultural Big Data**. ArXiv abs/1905.12411 (2019).

REIS, N. V. B. **Construção de estufas para produção de hortaliças nas Regiões Norte, Nordeste e Centro-Oeste**. (2022) Available in <https://www.embrapa.br/documents/1355126/9124396/Constru%C3%A7%C3%A3o+de+estufas.pdf/8bec74eb-2206-44ff-9aad-538141520c4a>. Accessed on May 21, 2022.

NNADI, S. N.; IDACHABA, F. E. **Design and Implementation of a Sustainable IOT Enabled Greenhouse Prototype**. (2018) IEEE 5G World Forum (5GWF), [s. l.], 2018. DOI 10.1109/5GWF.2018.8517006.

DANITA, M.; BLESSY, M.; SHEREEN, N.; NAMRATA, S.; PAUL, J. J. **IoT Based Automated Greenhouse Monitoring System**. (2018) Second International Conference on Intelligent Computing and Control Systems (ICICCS), [s. l.], 2018. DOI 10.1109/ICCONS.2018.8662911.

RODRIGUEZ, F. et al. **Modeling and Control of Greenhouse Crop Growth**. 1. ed. Glasgow: Springer International Publishing, 2015. 250 p.

CALAZANS, A. T. S. **Medição de Tamanho para Sistemas de Data Mart**. (2003) 138 f. Dissertação (Mestrado em Gestão do Conhecimento e Tecnologia da Informação) – Universidade Católica de Brasília. Brasília: UCB, 2003.

KIMBALL, R. and ROSS, M. **The Data warehouse toolkit: guia completo para modelagem dimensional**. 4. ed. Rio de Janeiro: Campus, 2002.

BOATENG, O.; GREESHMA, J. S.; GREESHMA, P. S. **Data warehousing**. Business Intelligence Journal, Renton, WA, v. 5, n. 2, p. 224-234, Jul 2012.

LUCIDCART. **Site para construção de diagramas**, (2022) Disponível em: <https://www.lucidchart.com/> . Accessed on Jun 3, 2022.

MAHER, A., KAMEL, E., ENRICO, F. et al. **Um sistema inteligente para o controle do clima e economia de energia em estufas agrícolas**. (2016) Eficiência energética 9, 1241–1255 (2016). https://doi.org/10.1007/s12053-015-9421-8