Energetic Assessment of Dairy Activities using OLAP systems

Sandro Bimonte, Jean-Pierre Chanet, Irstea, TSCF, 9 avenue Blaise Pascal CS 20085 63178 Aubière, France, sandro.bimonte@irstea.fr, jean-pierre.chanet@irstea.fr
Jacques Capdeville, Yves Lefrileux, Institut De L'elevage Jacques, 31321 CASTANET TO-LOSAN, France, Capdeville@idele.fr, Yves.Lefrileux@idele.fr

Abstract

Agricultural energy consumption is an important economic, environmental and social issue. The French national project EDEN aims at the analysis of energies consummations for two main activities: agricultural productions and dairy activities using OLAP systems. In this paper, we focus on energetic assessment of dairy activities. We present a multidimensional model allowing for monitoring of energies used for dairy activities. The developed model is complex but in counterpart it allows for analysis of inputs, outputs and their relations of the dairy farm system. In this way decision-makers can understand and correlate energetic trends overall the global production process.

Keywords: OLAP, Data warehouse, agricultural energies

1. Introduction

Agricultural energy consumption is an important environmental and social issue. Several diagnoses have been proposed to define indicators for analyzing energy consumption at large scale of agricultural farm activities (year, farm, family of production, etc.) (Bimonte et al., 2014). However, to define ad-hoc environmental energetic policies to better monitor and control energy consumption, new indicators at a most detailed scale are needed. Moreover, by defining detailed scale indicators, large quantities of data need to be collected to feed these energetic diagnoses. This huge volume of data represents another important limitation of systems that implement these diagnoses because they are usually based on classical data storage systems (such as spreadsheet tools and Database Management Systems) (Bimonte et al., 2014). These systems do not allow for interactive analysis at different granularities/scales of huge volumes of data. By contrast, OLAP and data warehouse (DW) systems allow for the analysis of huge volumes of data by providing aggregated numerical values visualized by means of interactive tabular, graphical and cartographic displays (Kimball, 1996).

In this paper, in the context of the French national project EDEN, we focus on energetic assessment of dairy activities. We present a multidimensional model allowing for monitoring of energies used for dairy-farming activities. In this paper, we focus on energetic assessment of dairy activities. We present a multidimensional model allowing for monitoring of energies used for dairy activities. The developed model is com-
plex but in counterpart it allows for analysis of inputs, outputs and their relations of the dairy farm system. In this way decision-makers can understand and correlate energetic trends overall the global production process.

The paper is organized in the following way: Section 2 described main concepts of OLAP and related work about using OLAP systems for animal production; Section 3 presents our multidimensional model for dairy analysis and the architectural implementation is described in Section 4; Section 5 discuses main issues tackled in this project, and Section 6 concludes the paper.

2 Main concepts and related work

In this section we present main concepts of data warehouse and OLAP (Sec. 2.1) and existing work about dairy-farming and OLAP (Sec 2.2)

2.1 Data Warehouse and OLAP

In this section, we present an example from Trujillo et al. (1998). The facts of a data warehouse are the values of the indicator to analyze (Malinowski and Zimanyi 2008). In the example, we consider the facts of the data warehouse to be the product sales of a company in dollars. Each of the company's stores provides these data. In a data warehouse, an analysis results from the use of an aggregation operation (e.g., sum or average) on the facts. In the example, a possible analysis is the sum of sales calculated by category of product, by store and by month. The result of this analysis can be represented in a cube (Trujillo et al. 2001) - see Figure 1.a. Each dimension of the cube corresponds to a criterion of analysis: type of products, store and month. The cells of the cube are called measures. They store the sums of sales for each tuple <type of products, store, month>. For instance, in Figure 1.a, the sum of sales for the tuple <Water, Store 1, December> is 4. In data warehouses, the criteria of analysis are structured in hierarchies called dimensions. Figure 1.b shows the three dimensions presented by Trujillo et al. (2001). A data warehouse can produce many analyses by combining different levels of dimensions. For example, other cubes could be calculated:
- sums of sales by city,
- sums of sales by brand, by city, by year,
- sums of sales by type, by state, by season, etc.

Note that data warehouses generally support n-dimensional cubes. Data can be combined to provide previously unknown causal links. To do so, users can visualize cubes from the data warehouse using tools like OLAP (On-line Analytical Processing) (Malinowski and Zimanyi 2008).

Using data warehouses is therefore important within a decision-making context. For example, a data warehouse containing economic, urban and environmental information will help decision-makers find the best place to establish a new infrastructure. The concept of the data warehouse has great potential for assessing the impact of actions, practices, scenarios and programs from both the socio-economic and the environmental point of view.
OLAP systems have been successfully used in several application domains such as marketing, health, environment (Bimonte et al., 2010), etc. In the context of animal production, to best of our knowledge only few work have been proposed. Rai, Dubeya, Chaturvedia and Malhotraa (2008) propose a system for the SOLAP analysis of define a system for the multidimensional analysis of data related to animal resources in relation with other agricultural factors such as vegetation, meteorology etc. They propose several multidimensional models and they detail a classical Relational OLAP architecture where data coming from external data sources are automatically integrated via ETL tools into a unique data warehouse. However, this work does not provide details about measures and dimensions used for the multidimensional analysis, and about data sources, but authors focus more on the global architecture with reports composed of tabular and cartographic displays. 

Along the same lines, Schulze, Spilke and Lehner (2007) analyze animals. In particular, they present a database model for storing information about cows and a multidimensional model with two measures: production of milk and treatments for cows, which are analyzed according dimensions representing time and animals. They provide some rules for mapping the conceptual model into a relational model. They extend the conceptual model for DW with textual information about measure: the type of aggregation, the meaning, the derivation formula.

Nilakanta, Scheibe and Rai (2008) present an OLAP system for the analysis of animals over different points of view. They define multidimensional models for the information about the population, the livestock census, the infrastructures available, and the production. All these DW are correlated and integrated in a unique constellation schema allowing to compare measures of different DWs along common dimensions. Moreover, the paper focuses on issues related to the temporal and spatial dimensions that must be defined at different granularities for those DWs. Finally, authors show also user-centered evaluation of the decisional system defining a set of questions.

Figure 1. Example of a data warehouse
To conclude, to best of our knowledge there are not work that detail indicators and analysis dimensions for dairy farming and a issues about integration of data coming from classical external sources, and sensor network.

3 Multidimensional model for Dairy-Farming Activities

In this section we present the multidimensional model we define for the analysis of dairy farm activities. It is shown on Figure 2 using the ICSOLAP UML profile (Bimonte et al., 2013). Note that since the model has been defined in the context of a French national project, we use French terms of dimensions and measures.

The purpose of UML profiles is to allow customizing UML for particular domains or platforms by extending its meta-classes (class, property, etc.). A profile is defined using three key concepts: stereotypes, tagged values and constraints. A stereotype is an extension of a UML meta-class. In the ICSOLAP UML profile, the multidimensional model is represented as a package stereotyped <<Hypercube>>. It contains dimensions, also represented as packages. Each dimension is composed of hierarchies, also represented as packages, which organize levels. A level (<<AggLevel>> stereotype) is a class composed of a set of descriptive attributes (<<DescriptiveAttribute>> stereotype) and an identifying attribute. <<TemporalAggLevel>> stereotype designates a level, in a temporal dimension. A fact is represented using the stereotype <<Fact>>, which is a class with attributes that are measures (<<NumericalMeasure>> stereotype for example).

The spatio-multidimensional model presents 8 dimensions:

- The production dimension (Production): the outputs of the farm (e.g. milk, cheese, etc.)
- The products-animals dimension (ProductD): the inputs of the farm (animals or other resources) used (e.g. cow, fodder, electricity, etc.)
- The equipment dimension (Equipment): equipment used in the technical operation (cooling tank, etc.)
- The technical operation (Operation_technique): dairy activity (e.g. dairy barn)
- The time dimension (Temps): classical temporal dimension with days, months and years
- The period dimension (Period): zootechnological period (e.g. summer grazing)
- The campaign dimension (Campagne): annual campaign (e.g. 2012)
- The location dimension (Localisation): it regroups rooms into farm and region.

It is important to note the time, campaign and period dimensions are three temporal dimensions that cannot be modeled as a multiple hierarchy since period and campaign depend on rooms and animals.
Several measures are defined. The first measure is the total of animals involved in a technical operation \((\text{sum}\_\text{intrans}\_\text{quantité})\). In this way, decision-makers can analyze for example how many cows are kept for milking per room and day (Figure 3a). In order to count the number of animals and evaluate the performance of the milking technical operation in the different farms, an indicator representing the total number of animals present in each group is needed (Figure 3b) \((\text{nbAnimal})\). This indicator is not associated to any technical operation. Then a dummy technical operation \((\text{NO}\_\text{OPERATION})\) is introduced as shown on Figure 3b. Since OLAP tool are addressed to final users, we have provided a derived measure \(\text{intrans}/\text{nbAnimal}\) (i.e. inputs/number of animals) that allow the ratio between the animals present and the animals involved in some technical operation (Figure 3c).
Animals are not the only resources used in the dairy activities. Energies (electricity, fuel, etc.) or animal food (fodder, etc.) are also used. In order to evaluate the usage of these inputs a measure \( \text{sum}_\text{intrant}_\text{quantité} \) (i.e. sum of quantities of inputs) is defined. Table of Figure 4a allows for example to analyze the quantity of fodder used for alimentation per room.

In order to understand relationship between animals and other farm inputs, for example how much kg of fodder is used per cow, we have defined a new derived measure \( \text{intrant}/\text{NbAnimaux} \) (i.e. inputs/number of animals) as shown on Figure 4b.

Farm activities use input to produce output like milk for example. In order to monitor outputs we have added a measure \( \text{extrant}_\text{quantité}_w \) (i.e. quantity of outputs). Figure 5 shows for example the milk produce by the milking operation per room and day.
In the same way we have compared different inputs, decision-makers need also to compare inputs and outputs for example providing a ratio between the milk produced and the number of cows involved in the milking operation (Figure 6) in order to answer to this kind of questions a new derived measure extrant/intrant (i.e. output/input) is defined.

Once dimensions and measures have been described we present how multidimensional analysis is provided using OLAP operators. In particular using the temporal dimensions is possible to understand the temporal evolution of indicators. For example the sum of milk produced can be shown by year, campaign and period. Using the drill-down operator is possible to provide an analysis at finest level for example milk produce by day. It is important that for the animal present in the farm, the sum cannot be applied on all dimensions. This measure is aggregated using the average on the temporal dimensions to avoid a double counting. Figure 7 shows the total of milk produced per campaign. In the same way it is possible to aggregate data by location as shown on Figure 8 where produced milk is summed per farm (Figure 8).
4 EDEN Architecture

In this section we present the architecture of our decision support system. As described in the previous section, the EDEN project aims at monitoring energies of agricultural production, irrigation data and dairy energies. Then, the global architecture is shown on Figure 9, where data are collected and integrated in a unique database, implemented using PostgreSQL. Then, data is extracted from this database and loaded in the OLAP system dedicated to the particular energy analysis. We need three different OLAP implementations since the multidimensional models are quite different (Bimonte et al., 2014).

![Figure 9. EDEN architecture](image)

Let us now focus on data of our case study. As shown on Figure 10 data is collected manually by farmers for example data about produced milk.

![Figure 10. Dairy architecture](image)

Contrary, some other information is automatically collected using sensors, for example data about electricity consummation. That information are transformed before to be integrated in the EDEN database, since those transformations concern the semantic of dairy data and are not generic routines valid also for other energy data sources. In particular at this stage, missing value is estimated and manually collected data are sometimes manually corrected by dairy farmers. Then, data are extracted from the EDEN database and loaded in the OLAP system using the ETL tool Talend.
We have used a classical Relational OLAP architecture for dairy data. It is composed of the Data warehouse tier that stores data is implemented using Postgresql. The OLAP server tier, which provides OLAP functionalities, is implemented using Mondrian. Finally, the OLAP client is JRubik.

5 Discussion

OLAP systems allow on-line and multidimensional analysis of huge volumes of data. Although, they are successfully used in several application domains (e.g. marketing, etc.) few work study integration of dairy data into those kind of systems. In the context of French project EDEN, we have addressed this issue. The main two tackled difficulties have been: the complexity of analysis needs of final decision-makers, and the data collection process. The dairy farming application domain is very complex. Thus in order to achieve the analysis of inputs and outputs of dairy activities with a unique and coherent data warehouse, some very complex multidimensional design solutions have been defined (e.g. multi-aggregation functions, complex hierarchies, etc.). It is important underline that the defined model is valid a generic model and it is valid for other animals such as goats.

Moreover, the collection of data is achieved using wireless sensor networks, with nodes that measure electricity, water, fuel consumption for each step of the milk production. The sample rate of data is 10 minutes, so the raw data are filtered (i.e. removal of errors) and aggregated (integration of data at a daily level) before being involved in the ETL process. This means that ETL routines have been developed taking into account these particular issues of wireless sensor networks.

6 Conclusion and future work

Energetic assessment of dairy activities is necessary for monitoring the agricultural farm and establishes references values for diagnosis tools. In the context of the EDEN project in this paper, we have presented a multidimensional model for the analysis of inputs and outputs of dairy activities. We have also proposed a Relational OLAP architecture based on manual and sensors data sources. Our ongoing work is the collection of data and the analysis with our OLAP system.

7 Acknowledgements

The authors would like to thanks the French Minister of Agriculture for its support of this project.

8 References


