Shendryk, Vira; Parfenenko, Yuliia; Boiko, Olha et al.

Article

Aggregation of multidimensional data for the decision support process for the management of microgrids with renewable energy sources

Technology audit and production reserves

Provided in Cooperation with:
ZBW OAS

Reference: Shendryk, Vira/Parfenenko, Yuliia et. al. (2022). Aggregation of multidimensional data for the decision support process for the management of microgrids with renewable energy sources. In: Technology audit and production reserves 2 (2/64), S. 16 - 20.
http://journals.uran.ua/tarp/article/download/255957/253185/589452.
doi:10.15587/2706-5448.2022.255957.

ISSN 2664-9969

UDC 004.65
DOI: 10.15587/2706-5448.2022.255957
Article type «Reports on Research Projects»

Vira Shendryk,
Yuliia Parfenenko,
Olha Boiko,
Sergii Shendryk,
Yaroslava Bielka

AGGREGATION OF MULTIDIMENSIONAL DATA FOR THE DECISION SUPPORT PROCESS FOR THE MANAGEMENT OF MICROGRIDS WITH RENEWABLE ENERGY SOURCES

The object of research is the process of processing and storing data when making decisions on managing the life cycle of electricity generation and consumption in microgrids with renewable sources. The prospects of the study are due to the fact that in order to provide a full-fledged decision support process in the management of microgrids with renewable energy sources, it is necessary to consolidate and manipulate multidimensional data in multithreading and online information processing. To solve the problem, the theoretical methods of analysis, abstraction, induction and deduction were used. To ensure multidimensionality and multithreading of data processing, it is proposed to develop a data warehouse based on the snowflake data model. Efficiency of information processing in real time is provided by an operational database built on the principle of OLTP. The organization of the joint work of the data warehouse with the operational database, the consolidation and manipulation of data is provided by triggers.

The result of the work is a data warehouse that will be used in the decision support system for managing energy microgrids, which will improve the efficiency of data processing and storage. This is achieved by combining the work of a centralized data warehouse with an operational database, as well as the use of a separate data mart for each user of the system. The practical significance of the work lies in the fact that the data warehouse will become part of the decision support system for processing information about the life cycle of energy in the management of energy infrastructure. Compared to using a single database for a decision support system, this approach ensures the speed of working with data and allows differentiating between the use of a data warehouse for analytics and data manipulation operations.

The data warehouse was deployed in a cloud environment on the Amazon Web Services (AWS) platform and the Amazon Relational Database Service (Amazon RDS) web service. Secure access to client data is implemented using data marts.

Keywords: *data processing and storage, data warehouse, database, data mart, triggers.*

Received date: 14.03.2022
Accepted date: 18.04.2022
Published date: 30.04.2022

How to cite

Shendryk, V., Parfenenko, Y., Boiko, O., Shendryk, S., Bielka, Y. (2022). Aggregation of multidimensional data for the decision support process for the management of microgrids with renewable energy sources. Technology Audit and Production Reserves, 2 (2 (64)), 16–20. doi: http://doi.org/10.15587/2706-5448.2022.255957

1. Introduction

The transition to the use of green energy requires the installation of special equipment, solar panels, wind or hydroelectric power plants. Their work is influenced by various important factors, such as wind speed for windmills and the length of the world day, cloud cover, other weather conditions and geographic location for solar panels. Integration of renewable sources into microgrids, effective management of energy generation and energy consumption requires the use of an energy management system, consisting of a complex of software and hardware. Determining the need for electricity, the possibility of its production depending on the existing conditions, the processes of forecasting and optimizing energy costs are accompanied by the processing of large data arrays. The larger the data sample, the more accurate the forecasting model, therefore, the more correct the management decisions made. To store and process information, a special data organization is required – a data warehouse. A data warehouse is preferred over a database because effective analytics requires the data warehouse it, which is developed as a layer on top of another database or databases. The data warehouse takes all this data and creates a layer optimized for quick and easy analytics. From the definition of a database, it follows that it is a set of data organized for storage, access, and retrieval, and a data warehouse is a type of database that combines copies of transaction data from various source systems and makes them available for analytical use. The advantage of the data warehouse is the use of OLAP (online analytical processing) technology for data processing, which consists in preparing aggregated information based on large data arrays structured according to the principle of multidimensionality. The basis for the aggregation of information can be databases built on the principle of OLTP (Online Transaction Processing), in which the data warehouse system processes small-sized transactions performed in a large stream with a minimum response time to a request. Thus, the advantages of both technologies can be used to achieve the goal of responsive and flexible management of renewable energy microgrids.

Comparative features of databases and data warehouses are presented in Table 1 [1].

Table 1

Comparison of databases and data warehouses

Parameter	Database	Data warehouse
Types	There are various types of databases, but the term is usually applied to an OLTP database	A data warehouse is an OLAP database, OLTP floor or other databases to perform analytics
Similarity	Both OLTP and OLAP systems store and manage data in the form of tables, columns, indexes, keys, views, and data types and use SQL to query the data	
Usage	Usually limited to one program: one program is one database. OLTP allows to quickly process transactions in real time	Stores data for any number of applications and databases. OLAP allows to organize one source of information for the entire organization used to manage analysis and decision making

Therefore, the data warehouse serves users in the process of collecting, storing, analyzing data and making decisions. Such systems can organize and present data in different formats according to the needs of different users [2].

A data warehouse is a domain-specific, integrated, nonvolatile data set that changes over time and supports the ability to manage decision-making [3]. It includes five main components: data sources, data extraction processes, central repository, data marts, and data access and analysis tools [4]. The data stored in the data warehouse has the characteristics of subject-oriented, integrated, and time-variable [5].

Therefore, *the object of research* is the process of processing and storing data when making decisions on managing the life cycle of electricity generation and consumption in microgrids with renewable energy sources.

The aim of research is to develop cloud storage for an information system supporting energy infrastructure management. Using a data warehouse will provide secure multi-threaded access to data for different types of users.

2. Research methodology

The renewable energy microgrid management process consists of separate sub-processes, each of which must be accompanied by appropriate data sets. The initial sub-process is the structuring of the collected current data on the state of power plants, weather data, power consumption level, information storage. The next sub-process is to use the data to predict the level of electricity consumption for the next period, as well as to predict the possible level of electricity generation depending on the forecast weather conditions. The final sub-process is decision support for managing the operation of the microgrid. Therefore, all data manipulation tasks are provided thanks to the data warehouse.

In this work, theoretical methods of analysis, abstraction, induction and deduction were used to develop a data warehouse model for a decision support system for managing energy microgrids.

Two types of data models in storages are considered: a star and a snowflake. In [4], devoted to the development of data warehouse architecture to support the energy management of smart electrical systems, a star model was used. This model has a denormalized data structure, so queries run faster. But at the same time, the same information can be repeated in different tables. In [6] on the development of a system for monitoring and analyzing energy optimization, a star scheme was also used and the importance of using an operational data warehouse was emphasized.

The star schema has one central fact table surrounding the dimension tables. The fact table is joined to the dimension tables by foreign keys. Foreign keys in a fact table can be used as a composite primary key because, as a rule, together foreign keys uniquely identify each row in the fact table. In relation to dimensions, the fact table is on the side of «many» [7].

The snowflake schema contains three types of tables: fact table, dimension tables and dimension tables, which are made by normalizing dimension tables, so the star schema can be distinguished as a snowflake schema in normalized form. Due to this, the data is not redundant, without repetition, which means less memory is required, which means less storage space required. There is one drawback due to the increase in the number of tables; queries for data from these tables will be more difficult [8].

A comparison of the two data warehouse models led to the conclusion that the star schema includes redundant data. This causes the data processing to cause a delay in time. The size of the table used in the star schema is also larger because the tables are not partitioned. Therefore, processing a larger table is somewhat time consuming. In addition, the snowflake schema is normalized tables that keep normalized data without redundancy, which makes queries faster [9]. Therefore, in this work, a snowflake model is used to build a data warehouse. An approach was also used, which consists in combining a data warehouse with an operational database that stores real-time data, for example, from sensors installed in the power system, for a limited period of time, after which the data becomes archived.

To deploy the data warehouse, hosting in a cloud environment was chosen. This provides high availability, fault tolerance, virtually unlimited resources, scaling and elasticity of resources, reliability and protection, which repeatedly compensate for the possible performance degradation [10]. In addition, this approach provides an opportunity for each client of the service to create a set of services that meet its needs for information for making management decisions.

3. Research results and discussion

A data warehouse has been developed, which is a source of unified consistent information for managing energy microgrids with renewable energy sources. End users have direct access to data as needed. Users of the system can be both individuals and legal entities – customers on which an electric microgrid is registered, while each user has access to a data mart, which, accordingly, provides access to information about their own energy microgrid or microgrid cluster.

ISSN 2664-9969

Data is received from data providers (sensors, meters, convectors, devices and weather APIs) and stored in an operational database. Information retrieval processes transfer important accumulated data to the warehouse and from there are delivered through data marts to end users. The architecture of the data warehouse is shown in Fig. 1.

The data in the operational database, which stores information about the current state of the microgrid, comes from various sources of supplying data on the state of the power grid. Weather forecast data is recorded at three-hour intervals to provide forecasts of renewable energy generation for the day, three days, or week ahead. Every hour data is received from microgrid devices. The information necessary for the functioning of the decision support system fills the data warehouse on a weekly basis. The storage filling procedure should take place at midnight GMT in order to minimize delays in the operation of the entire system.

In order to ensure speed and save disk space, data from the storage is annually archived and written to .csv files. The correct file name will allow quickly finding the desired file and calculating it as needed. To serve the activities of a specific user, data is supplied to individual data marts.

The data flow diagram in DFD notation, which shows the relationships between the operational database, data warehouse and data marts, as well as the processes of processing and using data, is shown in Fig. 2.

In conceptual modeling, database entities were identified, the main ones are described below. Entity – Customer, the person to whom the electric microgrid is registered. It has the right to edit the initial information about the power system, its configuration. An unlimited number of users can be added to the customer – Users entities that can use the system with access to the customer's microgrid data. Depending on the completeness or restrictions of rights, the system provides user roles – UserRoles. The City entity stores the coordinates of the location of the microgrid (longitude and latitude) for use in the weather API lookup request, as well as information about the time zone. The Location entity represents a microgrid owned by a specific customer. The ActualWeather entity stores current weather data at a microgrid location. The DeviceType entity will store three types of supported devices: solar panels, wind turbines, and batteries. The Device entity is designed to store a data array of devices used in customer microgrids. The DeviceParameters entity stores device parameter values. The LocationDevices entity stores location records for solar panels and windmills. The LocationData entity stores data received from network switches, on or off, voltage and power sensors. The ForecastData and ActualData entities are designed to store predicted and actual data on electricity generation for each device at the location. The ForecastConsumption entity stores the predicted electricity consumption for a location.

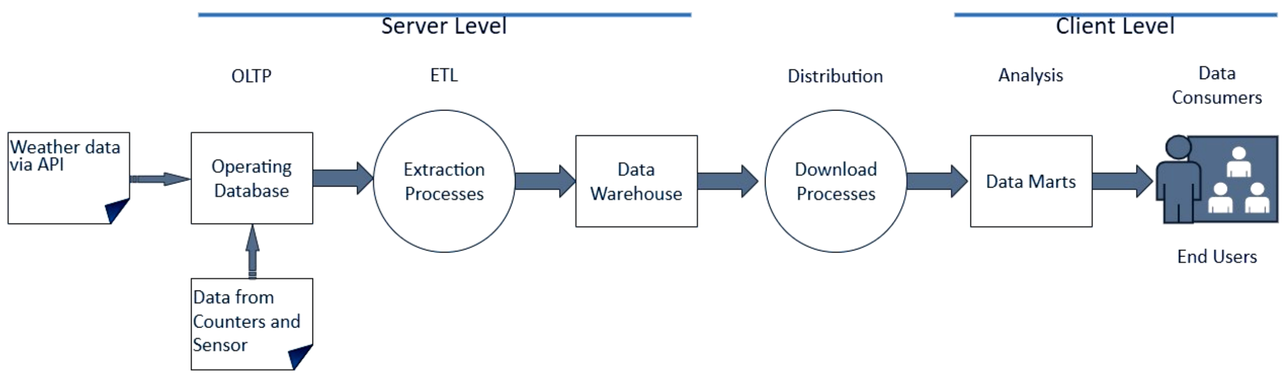

Fig. 1. Data warehouse architecture

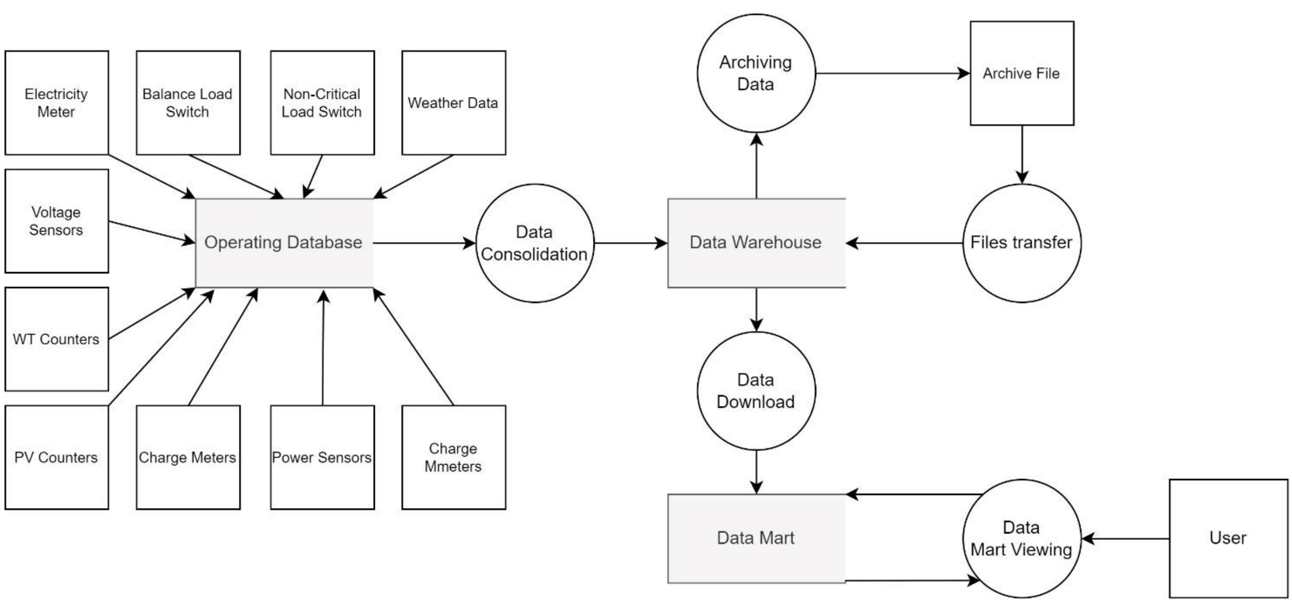

Fig. 2. Data flow diagram

The data of all entities, except for Users, UserRoles, DeviceParameters, Parameters, is transferred to the data warehouses as a result of the data consolidation operation once a week. Although the warehouse is populated on a weekly basis, at the same time there is data that must go into the warehouse after being added to the database, which ensures uniformity of identifiers in the database and storage for easier data manipulation. Triggers are used to ensure this process. A trigger is a task that is executed in response to a specific predefined event that occurs in the database, such as after a new row has been added to a specific table. In particular, this event involves inserting, modifying, or deleting table data, and the trigger may fire before or immediately after any such event.

The SQL code of triggers for adding data from the Device and DeviceType database tables to the storage tables DimDevices and DimDeviceType is shown in Fig. 3.

A trigger is created with the keywords «Create Trigger» followed by the name of the trigger. The next parameter is the trigger execution time (before or after), the event (insert – insert, update – update or delete – delete) and the table in which the event occurred. The last one is a description of the action to be performed.

The data entered into the repository periodically include:
- type 1 – record of actual data from location sensors;
- type 2.1 – record of the actual generation of energy from each source of electricity at the location;
- type 2.2 – predicted energy generation for each energy source at the location, depending on weather data;
- type 3 – record of the predicted electricity consumption at the location.

The procedure for filling the data warehouse consists of three functions: FunctionAddActualLocationData – adds records of type 1, FunctionAddActualDeviceData – records of type 2.1, FunctionAddForecastData – works with records of types 2.2 and 3, since the forecast data is combined with forecast weather records. Two parameters start_ and end_ are passed to the procedure and functions, which define the period of time for which records should be added to the storage.

Data marts have been created to present data to a specific user. Archiving of data from the data warehouse is carried out in a .csv file.

For deployment in the cloud, Amazon Web Services (AWS) cloud platforms and Amazon Relational Database Service (Amazon RDS) web service for working with relational databases in the cloud were selected.

The limitation of deploying a data warehouse in a cloud environment is the amount of memory allocated for it, so before deploying a data warehouse for a particular client, it is necessary to determine the amount of data that needs to be manipulated, taking into account the possible addition of microgrid objects, or the destruction of a microgrid cluster.

Further research consists of integrating a data monitoring subsystem, a data warehouse-based data warehouse subsystem with a subsystem for predicting the possible states of a microgrid in operational and long-term modes, as well as developing a knowledge base and rules for decision support in the management of microgrids with renewable energy sources.

4. Conclusions

The article proposes the use of an aggregated method for processing multidimensional data for the decision support process in the management of microgrids with renewable energy sources. To ensure a full-fledged decision support process, it is necessary to operate with sufficiently large data arrays. Some of this data is processed online in real time; some is used in long-term planning processes and processed over a longer period. In addition, the distribution of roles and boundaries of access to data for a different range of users is performed. To ensure multidimensionality and multithreading of data processing, it is proposed to use a data warehouse based on the snowflake data model. For operational processing of information in real time, an operational database is used as a sublayer of the data warehouse. To consolidate and manipulate data, triggers have been developed that allow organizing the joint work of the data warehouse with the operational database. Each user gets secure access to the right data through their respective data marts. To ensure uninterrupted and secure use, the data warehouse is deployed in the cloud on the Amazon Web Services (AWS) cloud platform and Amazon Relational Database Service (Amazon RDS) web services.

References

1. Cardon, D. (2018). *Database vs. Data Warehouse: A Comparative Review.* Available at: https://www.healthcatalyst.com/insights/database-vs-data-warehouse-a-comparative-review/
2. Han, J., Kamber, M., Pei, J. (2012). *Data Mining Consepts and Techniques.* Elsevier, 740. doi: http://doi.org/10.1016/c2009-0-61819-5

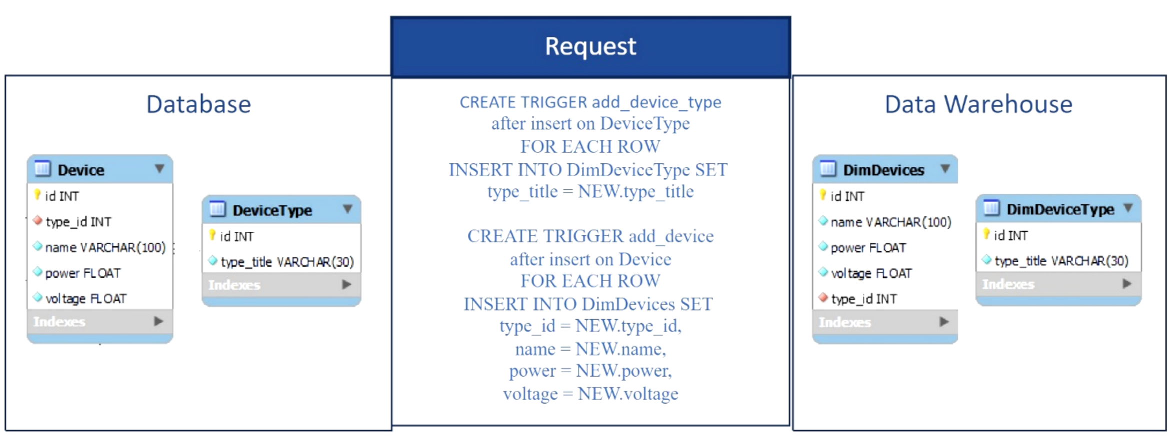

Fig. 3. Triggers for fixture tables and fixture types

3. Yang, Q., Ge, M., Helfert, M. (2019). Analysis of Data Warehouse Architectures: Modeling and Classification. *Proceedings of 21st International Conference on Enterprise Information Systems*, 604–611. doi: http://doi.org/10.5220/0007728006040611
4. Li, Y., Wang, L., Ji, L., Liao, C. (2013). A Data Warehouse Architecture supporting Energy Management of Intelligent Electricity System. *Proceedings of the 2nd International Conference on Computer Science and Electronics Engineering*, 696–699. doi: http://doi.org/10.2991/iccsee.2013.177
5. Samsinar, R., Endro Suseno, J., Edi Widodo, C. (2018). Power Distribution Analysis For Electrical Usage In Province Area Using Olap (Online Analytical Processing). *Proceedings of the 2nd International Conference on Energy, Environmental and Information System*, 1–5. doi: http://doi.org/10.1051/e3sconf/20183111010
6. Gökçe, H. U., Gökçe, K. U. (2014). Multi-dimensional energy monitoring, analysis and optimization system for energy efficient building operations. *Sustainable Cities and Society, 10,* 161–173. doi: http://doi.org/10.1016/j.scs.2013.08.004
7. Sarka, D., Lah, M., Jerkic, G. (2014). *Implementing a Data Warehouse with Microsoft SQL Server 2012*. Microsoft Official Academic Course, 816.
8. Levene, M., Loizou, G. (2003). Why is the snowflake schema a good data warehouse design? *Information Systems, 28 (3),* 225–240. doi: http://doi.org/10.1016/s0306-4379(02)00021-2
9. Iqbal, M. Z., Mustafa, G., Sarwar, N., Wajid, S. H., Nasir, J., Siddque, S. (2020). A review of Star schema and Snowflakes Schema. *Intelligent Technologies and Applications,* 129–140. doi: http://doi.org/10.1007/978-981-15-5232-8_12
10. Freedman, A. (2018). *Dedicated vs. Cloud: Comparing dedicated and cloud infrastructure for high availability (HA) and non-high availability applications.* Available at: https://silo.tips/download/dedicated-vs-cloud-comparing-dedicated-and-cloud-infrastructure-for-high-availab

Vira Shendryk, *PhD, Department of Information Technology, Sumy State University, Sumy, Ukraine, ORCID: https://orcid.org/0000-0001-8325-3115*

✉***Yuliia Parfenenko,*** *PhD, Department of Information Technology, Sumy State University, Sumy, Ukraine, ORCID: http://orcid.org/0000-0003-4377-5132, e-mail: yuliya_p@cs.sumdu.edu.ua*

Olha Boiko, *PhD, Department of Information Technology, Sumy State University, Sumy, Ukraine, ORCID: https://orcid.org/0000-0001-8557-2267*

Sergii Shendryk, *PhD, Department of Cybernetic and Informatics, Sumy National Agrarian University, Sumy, Ukraine, ORCID: https://orcid.org/0000-0002-2034-7920*

Yaroslava Bielka, *Independent Researcher, Sumy, Ukraine, ORCID: https://orcid.org/0000-0002-5952-658X*

✉*Corresponding author*