Modern concepts for logistics information management

Nazila Rahimova, Taleh Asgarov, Samad Humbatov, Elmaddin Huseynov

Abstract

The article describes the ETL and storage subsystems for logistics information management and the processes performed there. For this purpose, the interactions of the components included in the structure for data storage are taken into account. Data Warehouse, OLAP and Data Mining technologies are used for storage, visualization and processing of logistic information.

Keywords: logistics, ETL process, data storage, data processing

Management in logistics systems is associated with the collection, storage, processing and transmission of large volumes of data. The requirements for management functions require the development of an information system that provides continuous access to reliable, complete and accurate information. In modern times, enterprises dealing with logistics issues widely use the potential of information and communication technologies for the implementation of their functions. Depending on the type of logistics information resources, their collection, processing and meaningful delivery to users in chronological order is carried out by stepwise processes. For this, it is appropriate to study the structure of the logistics system and to know the rules by which the components that make up it work. In such a case, efficient data storage, generalization, establishment of inter-level information relations, and efficient processing of information in logistics systems are carried out with the help of the ETL (Extract, Transform, Load) process [1, 2].

It is known that in modern information systems processing of logistic data is carried out by levels. Electronic documents are presented to decision-makers at the final stage after the processing of logistic data by levels. Submitted documents are prepared periodically, at the end of the month, quarter and year, as well as according to requests related to the work of the organization. In general, the organization of logistics structures is designed based on the concept of a data warehouse (VA, Data Warehouse-DW). Logistic systems implemented according to this concept are based on modern information technologies and increase management efficiency in the enterprise. It is with this concept that electronic documents received from separate logistics enterprises are harmonized, brought to a uniform format and elimination of duplications. As a result, the user collects detailed and reliable information according to the requested request. In most cases, aggregated data is used to answer queries. A number of applications (ETL, Oracle Data Integrator, SAS Data Integration Server, Informatica powercenter, etc.) Are available to enable the aggregation of such data in an enterprise manner. In general, systems using VA are built on the basis of "client-server" and Intranet technologies [3, 4].

In the article, the structure of the logistics system is built on a three-level scheme, and the following issues are solved for the processing of logistics data at separate levels included in the scheme:

Data extraction, transformation and loading (ETL);

Data storage (data storage);

Data analysis (data analysis).

The organization and modeling of the ETL process in the data warehouse is detailed in [2]. Lately, the ETL process is applied when the enterprises dealing with the log stock are working with large volumes of data. This frees the enterprise's main corporate network from many issues such as replication, backup copies, and data recovery. The ETL process performs the following key functions:

Extraction of information from external information sources (Extract);

Data conversion (transformation) and cleaning (Transform);

Data preparation and loading (Load).

In organizing such a process, a data flow scheme is formed and that scheme is equipped with a data source (initial data), intermediate memory and data receiver.

Figure 1 shows a generalized scheme of the ETL subsystem for logistics information management. These

subsystems include "ETL" and "data storage" (hereinafter we will call the data storage subsystem - data warehouse). There are certain mutual information relations between them. The picture clearly shows the mutual information relations and processes between these subsystems. As can be seen from the diagram, the procedure that takes place in the VA (Staging Area and Hard Area) with the help of the ETL process is carried out in the following order:

A) "Data Extraction" process ensures the extraction of different types of information from separate information sources;

B) The extracted data set is added to the relevant tables of the "Extracted data source" component (Data Source Area);

C) With the help of the "Data transformation" process, the data in the Data Source Area block is called, the intermediate results of the obtained transformation are transferred to the "Transformed Data" component (Transformation Area);

D) After appropriate conversions, the obtained data are uploaded to the database called "Operational Data Store";

E) Reading of data from the "Long-term data storage" area (Hard Area) is carried out by the "Data loading" process;

F) During the "Data Upload" process, the completeness of the data is checked and it is ensured that they are uploaded to the database called "Data Set";

G) Performs data set reading for the "Data Aggregation" process.

H) The results of the "Data Aggregation" process are written to the "Aggregate" (Summary Area) and "Data Marts" databases.

Now let's get acquainted with the information processes carried out in the ETL subsystem.

Here, data, information sources - automated information systems, data files, electronic archives, etc. Provides removal. The data collected from those sources are written to the corresponding tables of the "Extracted data source" component (Data Source Area) without undergoing transformation.

Performs the procedure of preparing transformation tables based on the data extracted in this process. This process performs the following functions:

Separation of the selected articles into subsets, determination of dimensions based on those subsets;

Formation of temporary keys (surrogate key) corresponding to the records stored in the tables of the "Transformation Area" database;

Remembering the transformation of the records produced in the tables of the Transformation Area or Data Source Area bases, compiling the attributes corresponding to those records and organizing their structures in the Operational Data Store component;

Formation of records related to information sources in tables and ensuring their permanent and long-term storage in the Hard Area field with the help of appropriate algorithms;

Formation of records in coded tables;

Generation of tables by time.

In this process, data is transferred to temporary (Staging Area) and long-term storage (Hard Area) areas.

At this time, the data is loaded into factological tables and aggregated data is realized.

The download process is performed according to the data flow scheme and this process is divided into the following classes:

According to the nature of loading: a) initial loading process (Initial load); b) updated (changed) loading process (Refreshing load).

According to the type of data source: a) SCF (Structuring Classification Files) – structured text files; b) UCF (User Classification Files) – standard classifier of the operating system, user classifier; c) MFT (Multiple Facts Tables) – factual tables of the operating system.

After the ETL process, the data is stored in the Data Storage subsystem. The structure of data in this

subsystem is described by metadata [5]. The Data Storage subsystem has a certain mutual information relationship with the ETL subsystem. The Data Storage subsystem intended for the further development of VA consists of temporary (Staging Area) and long-term storage (Hard Area) areas of data.
The Staging Area includes the following components:
Extracted data source (Data Source Area);
Transformed data (Transformation Area);
Operational data storage component (Operational Data Store).
The Hard Area is made up of the following parts.
Loaded data set (Data Set);
Aggregated data (Summary Area);
Data showcase (or kiosks) (Data Marts).
Due to the fact that data comes from different sources and in different formats, they are collected in a structured way in the Data Source Area component.The Transformation Area component provides a unified framework by classifying records by type (date, number, text type) to eliminate contradictions between information sources (databases, electronic archives, etc.). This component is important in the representation of facts in VA with various schemes (star, snowflake schemes).
The task of the Operational Data Store component is to realize a logical and physical view of corporate data. Here the data is stored in a structured way. Data is not modified after uploading, but new entries are allowed to be added. Hardware and software tools can be used to protect information at a reliable level.
A Data Set stores a set of downloaded records.
Summary Area implements aggregated data processing. Here the data is summarized according to different hierarchy levels. If the data refers to a small period of time (day, ten days, month), it is possible to summarize them according to a larger period of time (month, quarter, year). Aggregation rate can be set for each aggregate.
The Data Marts database can be viewed as a specialized thematic database that provides solutions to analytical issues that characterize the organization's activities. As is known, multidimensional (MOLAP), relational (ROLAP) and hybrid (HOLAP) models are used as VA data representation models [6]. For large VA, the ROLAP model is considered more convenient. To efficiently perform data analysis in VA, it is advisable to implement the ROLAP model with "star" (if the number of dimension tables is not large) or "snowflake" (if the number of dimension tables is large) schemes.
Logistic activity is multifaceted and determined by the nature of the problem being solved. In order to solve analytical problems in decision-making, values of logistic information are stored for specific time periods. Here, the relevant decision-making process is carried out in a "top-down" direction. Primary information is integrated and summarized when it is transferred from a lower level to a higher level [7]. Decisions become more specific as they are directed from the top down. Special automated workstations are used to provide users with applications to the system in making decisions. Users access VA through client applications (or through a browser). Currently, information visualization is of great importance. In solving the problem under consideration, visualization objects can be divided into four groups: standard reports, unregulated queries, multidimensional analysis (OLAP) and data mining (Data Mining). OLAP and Data Mining technologies are implemented with the help of appropriate tools (for example, Oracle reports, Oracle Discoverer, Discoverer for OLAP, OLAP Spreadsheet, Oracle data miner, etc.).

Conclusion

Recently, the management of logistics information has actualized the application of corporate data warehouses and information systems. The main condition is the optimal storage and management of documents collected in the information systems operated in the computer networks of large logistics enterprises, as well as provision of uninterrupted access to them. For this, the architectural foundations of the ETL system, which supports decision-making based on Data Warehouse, OLAP and Data Mining

technologies in logistics organizations, are provided. It is shown that by applying ETL and storage subsystems for logistics information management, data summarization or vice versa, detailing, creation of backup copies, separation of information arrays into categories, etc. Such issues are easily resolved.

References

[1] Karimov S.Q. & Askerov T.K. The structure of the corporate information-analytical system and the organization of the ETL process in the system. News of ANAS, 2012, No. 2, p. 49-54.

[2] Karimov S.Q. & Karimova G.K. Organization and modeling of the ETL process in the data warehouse. News of ANAS, 2011, Volume XXXI, No. 6, p. 20-26.

[3] Askerov T.K. Organization of the work process in a system that supports decision-making based on a database with different architectures. News of Baku University, Series of Physical and Mathematical Sciences, 2014, No. 1 p. 111-117.

[4] Askerov T.K. & Abdulayev V.H. About the context search method applied in information-search systems. Institute of Physics of ANAS, Problems of Energy, 2013, No. 4, p. 93-95.

[5] Torben B. & Christian Y. Technology of multidimensional data bases. Open systems. 2002, No. 2, p. 71-74.

[6] The procedure for developing ETL processes http://citforum.ru/gazeta/5/

[7] Architecture of corporate storage.http://www.prj-exp.ru/dwh/dwh_architecture.php