

PAPER • OPEN ACCESS

Development of datamining software for the city water supply company

To cite this article: O G Orlinskaya and E V Boiko 2018 *J. Phys.: Conf. Ser.* **1015** 042047

View the [article online](#) for updates and enhancements.

You may also like

- [Implementing Online Analytical Processing in Hotel Customer Relationship Management](#)
M Taufik, F Renaldi and F R Umbara
- [Improve HR Decision-Making Based On Data Mart and OLAP](#)
Alaa Khalaf Hamoud, Maysaa Abd Ulkareem, Hisham Noori Hussain et al.
- [Algorithms for synthesizing management solutions based on OLAP-technologies](#)
A M Pishchukhin and G F Akhmedyanova



ECS
The
Electrochemical
Society
Advancing solid state &
electrochemical science & technology

DISCOVER
how sustainability
intersects with
electrochemistry & solid
state science research

Development of datamining software for the city water supply company

O G Orlinskaya¹, E V Boiko²

¹ North-Caucasus Federal University, 2 Kulakov Ave., Stavropol, 355000, Russia

² Junior accounting automation specialist, "Business IT ltd." 3a Promishlenaya St., Stavropol, 355035, Russia

E-mail: O.Orlinskaya@rambler.ru

Abstract. The article considers issues of datamining software development for city water supply enterprises. Main stages of OLAP and datamining systems development are proposed. The system will allow water supply companies analyse accumulated data. Accordingly, improving the quality of data analysis would improve the manageability of the company and help to make the right managerial decisions by executives of various levels.

1. Introduction

In today's information society, the development of technology has led to the fact that the amount of data accumulating in the databases of large companies is rapidly increasing. However, raw information is not useful to analysts and executives, as they simply cannot handle all this data and base their business decisions on it. Therefore, to determine the "hidden knowledge" and various patterns in large arrays of "raw" data, it is necessary to use datamining.

Datamining has become an integral part of business planning, modelling and forecasting. In the current environment of global competition, finding patterns can be a source of additional competitive advantages. For example, this can be done through careful analysis of customer needs, financial forecasting, and analysis of the company's fixed assets.

2. Materials and methods

Today, 1C-branded information systems are the most popular on the Russian market. According to analysts, about 83% of jobs in Russia are automated using the "1C:Enterprise" platform [1].

In order to analyse data from "1C:Enterprise 8" databases and derive various statistics for reports, one can use the data analysis and forecasting tool included in the platform.

The data analysis and forecasting tool is one of the mechanisms for forming economic and analytical reports. It provides users (economists, analysts, etc.) with the ability to search for non-obvious patterns in data accumulated in the database [3]. The tool allows one to:

- search for regularities in the source data;
- control parameters of the performed analysis, either programmatically or interactively;
- programmatically access the result of the analysis;
- automatically output the result of the analysis onto a spreadsheet;
- create forecast models that allow predicting further events or values.



The data analysis tool is a set of interrelated objects in the platform's embedded programming language. The analyst may use its components in any combination in any applied solution. The analysis parameters are changed through the interactive settings, and the results are displayed in a spreadsheet.

The data source for the tool does not have to be an information database; it can also be data from external sources in spreadsheet form.

This tool is implemented in the "1C:Enterprise" platform via the following objects [3]:

- data analysis, which performs the calculations itself. It is necessary to define the data source and the required parameters for analysis;
- result of analysis, which stems from data analysis;
- forecast model, which is generated based on the analysis result.

The object is the final link in the 1C analysis mechanism. It generates a table with predicted values.

A result is obtained by applying one of the analysis types to the original data. The result of analysis is a model of data behavior. The result of analysis is displayed in the final document and can be saved for later use.

Further use of the analysis result lies in creating a model that allows predicting the behavior of new data in accordance with the existing model.

The data analysis and forecasting tool can perform several types of data analysis:

- general statistics;
- sequence search;
- cluster analysis;
- decision tree;
- forecast model.

The current version of the 1C platform has some OLAP elements: for example, the operative accounting register keeps data on turnover sums and intermediate points in balances, as well as the data layout scheme (DLS), which allows one to build various reports, presenting data in multidimensional tables or charts.

However, the imperfection of these mechanisms, high computational complexity and slow speed of the DLS, especially when dealing with information over a fairly long period, significantly slows down the system and prevents managers and other stakeholders from obtaining information promptly.

Due to the aforementioned shortcomings of the "1C:Enterprise" platform, an external OLAP system must be used to analyse data and produce required reports, as this will reduce the load on the 1C database and allow to analyse in real time.

At the time of writing, the water supply company in the city of Stavropol serves more than 500 000 clients monthly, getting a mix of information about consumers and the water supply system such as data on water consumption, the number of applicants for connection, payment data and other indicators. The organization uses an automated system "Vodokanal: billing" which is based on the "1C:Enterprise 8.2" platform.

This software serves the following purposes:

- automating the process of identifying debtors among clients and analysing the main parameters which can determine a debtor;
- automating the process of determining the type of property for clients and analyzing the main parameters which can determine the type of property;
- automating the process of calculating future cash inflow and volume of consumption for future periods;
- reducing labor and time costs for generating reports for operational purposes;
- simplifying the process of creating accounting reports without the assistance of a specialist programmer.

In view of these objectives, final requirements were stated for this software before development started.

We suggest using Microsoft SQL Server to develop Business Intelligence solutions. This DBMS works well with the "1C:Enterprise" platform. Microsoft SQL Server comes with Business Intelligence Development Studio, which allows creating business intelligence projects. The environment includes projects from Microsoft SQL Server Analysis Services (SSAS) and SQL Server Integration Services (SSIS). This environment allows creating a variety of projects including object templates and tools for solving problems of business intelligence.

Data mining with SQL Server has the following advantages:

- the ability to use a variety of spreadsheet data as a source for data mining. It is possible mine OLAP cubes created in Analysis Services;
- built-in data cleansing, data management and reporting: one can create ETL processes for cleansing data in preparation for modeling;
- custom algorithms in addition to clustering, neural networks and decision trees;
- model validation infrastructure: cross-validation, classification matrices, lift charts, and point charts;
- queries and detailing: SQL Server provides the DMX language for integrating predictive queries into the application;
- client tools: datamining add-ons for MS Excel.
- support for scripting languages and managed APIs: every datamining object is fully programmable;
- security and deployment: provides security based on roles though Analysis Services.

This datamining system with OLAP technologies is meant for the Stavropol city water supply company "Vodokanal ltd.", specifically, the consumer department, the analytics department, and management.

The main purpose of the datamining system for the city water supply company is to generate real-time reports.

The development of the OLAP datamining system consists of three stages [2]:

- collecting information about data in the 1C database and developing a database for transferring relevant data;
- creating an OLAP system with the database;
- creating a mining model using SSAS.

During the first stage of development, let us explore the datasets in the system "Vodokanal: billing" and determine the type and composition of the information contained in the database. Based on the obtained data, let us build a database of appropriate structure.

To build a database it is necessary to include the following subsystems from "Vodokanal: billing": "Clients", "Debtors", "Legal Info", "Cash Flow". In addition, this stage involves the transition from a fileserver of the 1C database to a client-server version for Microsoft SQL Server. A client-server version is appropriate for the larger scale of workgroups or enterprise-wide use. We use a three-tier client-server architecture. In this embodiment, the database is stored with one of the supported database management systems, and the interaction between the client application and the DBMS is done by the "1C:Enterprise 8" server cluster [3].

During the second stage of development, let us design and build an OLAP system based on a client-server version of "1C:Enterprise". This stage is the most time consuming, because the original 1C database may have a different structure, and the information stored in them may contain errors, duplicates, and have a different format. At this stage, the data is retrieved from external sources, transformed, and cleansed, then loaded into the newly created OLAP database.

The third step is constructing a model of multidimensional analysis. The datamining model in Business Intelligence Development Studio can be built using a variety of datamining algorithms. Analysis Services include the following types of algorithms [2]:

- classification algorithms;
- regression algorithms;

- segmentation algorithms;
- association algorithms;
- algorithms for sequence analysis.

It is important to note that before the creation of the multidimensional data model, a schema for the database must be defined. The authors selected the "snowflake" schema.

The snowflake schema got its name because of its shape, which shows a logical schema of tables in a multidimensional database. Same as the star schema, the snowflake schema is represented by a centralized fact table connected to dimension tables. The difference is that snowflake dimension tables are normalized with a number of other associated measurement tables, whereas in the star schema, dimension tables are fully denormalized and each dimension is represented in a single table with no links to related tables.

After defining the structure of the database, let us create an OLAP cube: identify data sources for analysis, aggregation and cleansing. If necessary, this OLAP cube can be used as an external data source object to connect to the "1C:Enterprise" database and to use it, for example, as a data source for a query.

The last step is creating and programming reports using PivotTables in MS Excel and creating various datamining models. PivotTable Services is a library used by Excel and other applications when creating multidimensional databases.

A Pivot Table is an interactive table for statistical analysis or summary of large amounts of raw data that can be described on a range of Excel cells or as a result of a database query. The basis for interactive tables is data from several columns of a source table. If a Pivot Table is created based on an OLAP cube, the fields are already pre-defined according to the type of OLAP data they belong to. Provided the same report structure, generating a table with DLS takes 17.7 seconds (according to data obtained by the "Measurement of performance" tool on the "1C: Enterprise" platform). At the same time, a Pivot Table with the required results is generated in 5 seconds – obviously a significant reduction in processing time, which reduces waiting time by the end user and increases productivity.

3. Results and discussion

The city water supply company uses several models for its datamining needs, such as models based on decision trees, Bayes algorithm, and time series. The datamining model based on decision trees has a prediction accuracy of 85%, the one based on Bayesian algorithm has an average accuracy of 68%.

The city water supply company needs reports on water usage by consumers not only from the whole city but also from specific areas, or in the context of a specific timeframe. Reports can also be generated by consumers and sorted by various parameters.

4. Conclusion

The proposed system will allow city water supply companies conduct analysis of accumulated knowledge. The resulting information will not only help predict cash flows and plan future expenses and revenues, but also to correctly configure the water supply system in the city, especially in developing areas. Assessment of the volume of consumption will allow for inspection and timely reconstruction of engineering networks, construction of new water supply systems and sanitation, which will further lead to improving the quality of customer service.

Accordingly, improving the quality of data analysis would improve the manageability of the company and help to make the right managerial decisions by executives of various levels.

References

- [1] Anton Malkov 2017 *Enterprise management system (Russian market)*. Retrieved from: <http://tadviser.ru/a/56070>

- [2] Boiko E V, Orlinskaya O G 2016 Using OLAP systems for applications on 1C:Enterprise V *Student science for development of information society vol 1* (Stavropol: North-Caucasus Federal University) pp 183-185
- [3] Boiko E V, Orlinskaya O G 2015 City water supply datamining system *III Student science for development of information society vol 1* (Stavropol: North-Caucasus Federal University) pp 112-114
- [4] Goncharenko M V, Babenko R Yu 2017 *Vodokanal:billing*. Retrieved from: http://teplovodokanal.biz-it.ru/vodokanal_citizens/