SPECIFICS OF THE TABULAR MODELS FOR THE OPERATIONAL DATA ANALYSIS OF RELATIONAL SOURCES

Bekarevich Yuri
St. Petersburg State University, 7/9 Universitetskaya nab., St. Petersburg, 199034 Russia

Abstract

The solution of analytical tasks on visual models, which are adequate for the structure of relational data sources, in the interactive mode with a convenient data view is of current interest for small and mid-size businesses with limited resources.

When using MS Excel и Power Pivot tools, the relational data source model is designed to have the presence of table relationships under surrogate keys. The surrogate keys are supposed to be input only into dimension tables with composite keys. For this, it is necessary to determine the indexes for the real keys in dimension tables. These actions are sufficient for the automatic construction of tabular model during import tables and their relationships into a tabular model, while keeping the advantages of database integrity support.

To reduce the analytical processing time, it is offered to input prepared tables with aggregated data into a tabular model. Such tables should include the most frequently used aggregates and could be created with SQL queries. These queries could also employ relational projection operation to select only those fields, which are used in the analysis.

The model characteristics could be improved by preliminary joining fields of interrelated dimension tables into one table of a lower level, which allows creating hierarchies, the use of which makes data analysis in aggregated reports more convenient.

Keywords: operational data analysis, tabular model, aggregation, relational data sources

JEL code: C81, L86

Introduction

To make a timely management decision, business professionals should have detailed up-to-date information on business performance and should have a possibility for its immediate operational analysis.

The purpose of an operational analysis is to get quick responses to spontaneous real-time analytical requests, to quickly make reports on aggregated parameters of business processes, which depend on various indicators, as well as to present these results in a human-friendly view, convenient for decision making.

The technology of On-Line Analytical Processing – OLAP covers the whole specter of data processing, necessary for a decision-making process. This information is presented by the facts under analysis, that depend on a large number of dimensions (Bekarevich and Pushkina 2013). In the field of business analytics Multidimensional Data Analysis takes almost an exclusive position. This tool provides flexibility and productivity in the solution of analytical issues for large companies. However, it is too bulky and unreasonably complicated for business analysis in small companies, since it requires the use of complex analytical server systems (Lidberg 2013; Hassana et al. 2014). For small and mid-sized businesses showing the rising interest towards business analytics methods one of the available tools are MS Excel pivot tables and diagrams and its Power Pivot add-in. The new versions of these tools allow interactive preparation of tabular models, that provide an adequate presentation of the interrelations of analyzed data (Ferrari and Russo 2014; Tim Hill 2015). These models also employ relational

model constructions, such as tables and relationships as well as a memory analytics module for data storage and calculation (Jelen and Collie 2014).

Company's data warehouses or transactional relation bases are usually used as the main source of operational analysis. The use of the data warehouse as a source provides quick and convenient construction of data models, as the storage structure is well focused on the analysis problems solutions (Kimball and Ross 2013). However, for many enterprises the creation of a data warehouse is a too expensive procedure, that requires additional hardware and software resources as well as a highly qualified personnel (Bekarevich and Pushkina 2014).

The use of a transactional database as a source is more affordable for small and mid-sized businesses. Excel and Power Pivot allow an operational analysis at MS Access local relational databases (Bekarevich and Pushkina 2014) and MS SQL Server, Oracle, IBM DB2, Informix, Teradata, Sybase server databases (Sarka et al., 2014), as well as at other data sources available through OLE DB and ODBC interfaces.

It is necessary to say that for the purposes of data analysis, Excel pivot tables and diagrams can employ tabular models or multi-dimensional model OLAP cubes, prepared in Analysis Services SQL Server (Lidberg 2013). The server analytical system, such as Analysis Services SQL Server, create tabular models or cubes OLAP based on the data source view (Russo et al. 2012).

With limited resources, it is necessary to find a solution which decreases expenses on the preparation of a tabular model of operational analysis data. Most part of business operations is registered, monitored and analyzed in the form of data stored in relational bases. The data scheme of such bases is similar to a tabular model of the analysis (Bekarevich and Pushkina 2014). However, to simplify the creation of a tabular model of the analysis in Excel and Power Pivot, a relational database structure must meet a number of requirements. Once those requirements are met, the tabular model can be received automatically. In this case, in compliance to the tasks of the analysis within the subject area, according to the analytical tasks within the subject, it is enough to set the analyzed indicators (facts) and parameters affecting them (dimensions), and to consider the aggregation of facts under the hierarchy dimensions.

The requirements to the structure of the relational data source

The creation of a tabular model, which considerably increases the effectiveness of an analytical procedure, is provided with the use of MS Excel 2013/2016 and Power Pivot, which implies the possibility of connection to the relational base and import of some of its tables (Bekarevich and Pushkina 2015; Bekarevich and Pushkina 2017). The tabular model gives the analyst a convenient presentation of the data in a table of facts and in connected dimension tables, corresponding to the database structure. The model can include several table with facts, referring to the same or to different dimension tables (Bekarevich and Pushkina 2016).

To build a Tabular model based on facts and dimensions, required for a certain research, one should define a database subschema, which is the source of the model's data and which includes corresponding tables and attributes. To build a model, relationships should be established between the imported tables. The relationships between the tables would be imported together with the tables, but they are included into the model under certain conditions only. This could be explained by the fact that Excel and Power Pivot tabular models do not implement a table relationship under composite keys. So, to establish relationships within the model, all the keys of dimension tables should be simple. Fact tables may have composite keys, since they are located on the lower level of the model hierarchy and do not have any subordinated tables with which it is necessary to establish connections.

To provide an automatic construction of the relationships within the model, it is necessary to enter surrogate keys into tables measurement of the database, which contain the composite key. In this case to support the integrity of real composite keys of measurement

tables it is necessary to define unique indexes. To connect the tables, the surrogate key of a measurement table should be included into the subordinate table as an external key. Surrogate keys are supposed to be entered into dimension tables only, which allows preserving table connections of the database under the real keys where it is possible, and avoiding the storage of additional table indexes in the database. These actions are sufficient to automatically construct a tabular models during the import of tables and their relationships.

Case study 1

Figure 1 shows an example of Power Pivot tabular model designed for an operational data analysis. The model was automatically received during the import of tables and their connections from the relational database on the contracts and product shipments to customers.

This tabular model includes two fact tables − Contract_spec and Shipment, and a number of measurement tables, a part of which is common for the both fact tables (Bekarevich and Pushkina 2016). The model provides the indicator analysis for each fact table and a comparative analysis of facts from both tables. At the same time one could make necessary calculations, developed by Key Performance Indicators (KPI) for managers and employees, corresponding to the purposes of business analysis. The facts to consider here include the quantity of the ordered product and the order total cost in the Contract_spec table. The dimensions, influencing the specified facts, are Product, Contract, Customer, Month_year. Using the specified facts and dimensions, one can receive answers to various questions on shipments, contracts and their violations. For example, how many times and in what quantity the product was ordered by each customer, for some contracts ordered a particular product, how many products is ordered within a certain period, for which periods the orders were placed, which products were not ordered by customers, which customers did not conclude any contract.

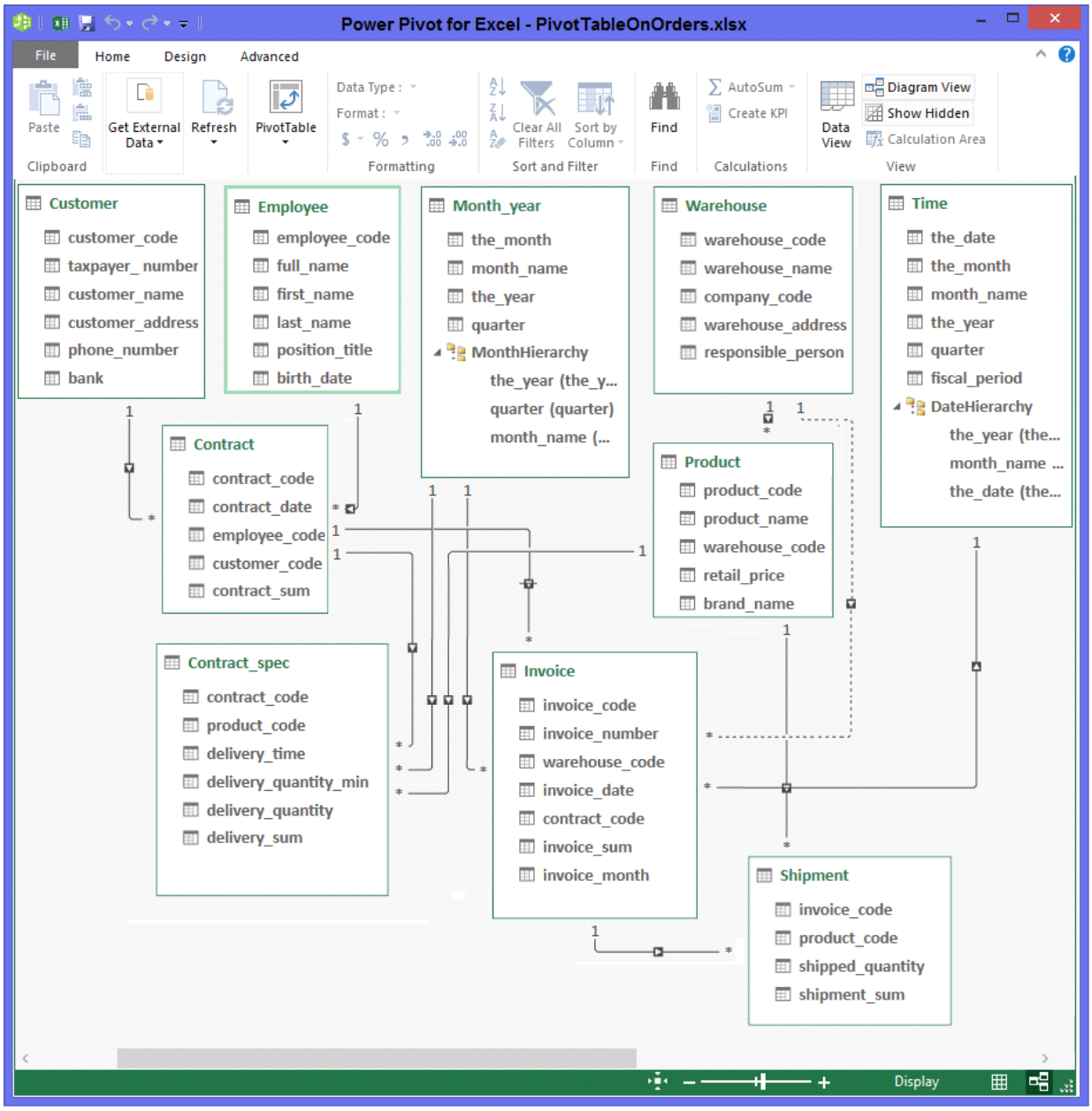

Figure 1. A tabular model for the analysis of the contract supplies

Aggregation of data

During a data analysis on the models containing the data of normalized tables of the relational base, aggregates are calculated based on the available detailed data (Tim Hill 2016; Jelen 2013). To decrease time spent on the calculation of the aggregates in the process of getting the results of the analysis, the corresponding SQL queries could be created beforehand, during the preparation of a model. These queries should serve as a base to create tables with most frequently used aggregates.

An additional effect may be received by excluding the fields from a model table, which are not used in the analysis. Such projection operation may also be performed by means of queries. Moreover, the model characteristics could be improved by joining the fields of various interconnected dimension tables into one tables of a lower level of the dimension hierarchy. It is possible to create hierarchies with the fields of the joined table, which makes the data analysis in aggregated reports more convenient.

It is possible to create queries with aggregated data on the database side or the Power Pivot client's side. The queries would form the content of fact tables with aggregated data into

models. When creating queries in the database it is possible to define their relationship to the tables of the database. Excel imports a query table together from the data source with data base tables. It is not necessary to save a query table in the database. During the import of queries with data aggregates and tables connected to them into the Excel book, a more compact tabular model will be automatically received.

Case study 2

Figure 2 for the initial Tabular model, presented in Figure 1, shows a transformed data model with tables queries Plan, Fact, containing aggregated quantities and the sum of contracts and shipment of goods.

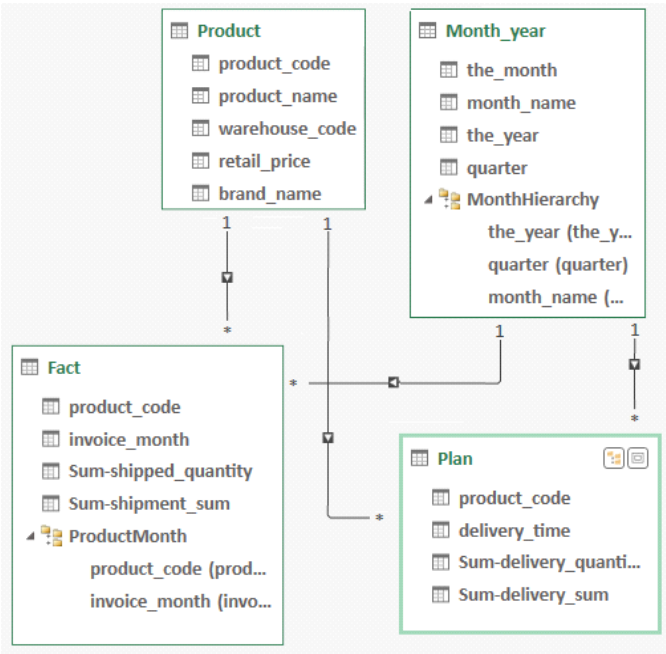

Figure. 2. Tabular model with the aggregated data in tables the Plan and Fact

Updating the tables of the data analysis model

An apparent advantage of a tabular model, created with the discussed tools, is a possibility to update of separate tables of the model directly from the operational source (Bekarevich and Pushkina 2013). As a rule, the updates require fact tables. Dimension tables are created, first of all, on the basis of reference books and conditional-constant data. They updated much less frequently fact tables. By establishing the selective updating policy for the tables of the model, one may significantly minimize the time required as well as the volume of the data being transmitted.

The operational analysis can be performed based on one non-normalized table, containing all the data necessary for the analysis. Such an option is provided in earlier versions of Excel, where you can import only one table. Such table could be prepared with SQL queries directly within a database. The use of this type of table, which is actually a multidimensional data model (Bekarevich and Pushkina 2013)], requires updating the entire data model.

Newer versions of Excel allow importing several tables simultaneously, and based on them to create a data tabular model for an analysis. In this case selective updating is provided.

Cloud storage of a data analysis model

The management personnel may be situated far away from a company's infrastructure. Modern technologies allow performing analytical tasks from remote computers and mobile devices through a browser and Internet access (Nam Hun Park, Kil Hong Joo 2014).

If a model was created with the methods described above, it is sufficient place the corresponding Excel book in the cloud (for example, OneDrive for business) to get a remote access to it. To ensure the possibility to update model data and model modifications, the database, to which the Excel book is connected, must also be placed in the cloud. When working in Excel from a remote computer, the analyst has access to the whole set of tools for the construction of a model and for data analysis. One can construct a model automatically by connection to the remote relational database and by importing tables and relationships. Using Power Pivot Excel tools, the model will be represented graphically as a data schema (see Fig. 1)

It is necessary to say, that if the analyst does not have Excel, but works through a browser, he/she may use Excel Online to perform data analysis in pivot tables. At the same time, depending on the type of the device, which the user operates, he/she has access to various features. For instance, a smartphone only allows operations with a prepared pivot table. A tablet with sufficient resources allows creating new pivot tables, as well as showing and modification of a tabular model.

Conclusions

The article formulates proposals to prepare a relational source for a tabular model; these proposals would provide automated creation of models in Power Pivot environment during the import of tables. This creates the conditions for an analyst to prepare models by himself/herself without any advice from an IT-specialist. The article shows the reasons for a limited use of surrogate keys in tables, which allows preserving the relationships of a normalized database under real keys. As a consequence, this preserves the possibility to automatically support a database integrity.

To raise the effectiveness of an analysis procedure, to reduce the required memory volume and time to get answers to the analyst's questions, the authors offer to use queries for the preliminary aggregation of detailed data in the new model tables.

The recommendations are given to allow a significant limitation of the volume of data, transferred from the source, due to the selective update of separate data model tables.

The location of a tabular model in the cloud will provide a remote analytical operation from any place. At the same time, to keep the data up to date, it is necessary that the source is also located in the cloud.

References

Bekarevich Y., Pushkina N. (2013) "Models for analytical processing data" Conference of St.-Petersburg State University of Economics. June 19-20, 2013 г. Information Technology in Business. Collection of Scientific Articles. SBR.:. Publisher Info-da, p. 28 – 35.

Bekarevich Y., Pushkina N. (2013) "User tools for analytical processing data" Conference of St.-Petersburg State University of Economics. June 19-20, 2013. Information Technology in Business. Collection of Scientific Articles. SBR.:. Publisher Info-da,. p. 35-45.

Bekarevich Y., Pushkina N. (2013) On-line Analytical Processing Technologies. – Study Guide. – SBR.: St.-Petersburg State University of Economics. 109 p.

Bekarevich Y., Pushkina N. (2014) Self-Help Manual Microsoft Access 2013 SBR.: BHV-St. Petersburg Publisher, –. – 450 p.

Bekarevich Y., Pushkina N. (2015) “Characteristics of the On-line Analytical Processing when using a tabular models” // Journal of Research and Practice «Problems of Economics and Management», № 2(42), Izhevsk, p. 120-126

Bekarevich Y., Pushkina N. (2016) “Questions on the Effectiveness of the Tabular models for data analysis” // Journal of Research and Practice “Problems of Economics and Management”, № 2(54), Izhevsk, p. 76-83

Bekarevich Y., Pushkina N. (2017) “POWER PIVOT EXCEL tools for solving analytical problems” Collection of scientific articles: Contemporary perspective on the future of science: priority directions and tools for development, St. Petersburg, p. 68-72

Bekarevich Y., Pushkina N. (2017) “The technology of preparing the model of operational data analysis based on relational sources” // Journal of Research and Practice «Problems of Economics and Management», № 3(67), Izhevsk, p. 76-83

Ferrari A., Russo M. (2014) Microsoft Excel 2013 Building Data Models with PowerPivot (Business Skills). Microsoft

Hassana A., Ravat F., Testy O., Tournier R., Zurfluh G. (2014) “Multidimensional database modelling with differentiated multiple aggregations”. Journal of Decision Systems, Vol. 23, No.4, 437-459

Jelen B. (2013) PowerPivot for the Data Analyst. Gardners Books

Jelen B., Collie R. (2014) Power Pivot Alchemy: Patterns and Techniques for Excel. Holy Macro! Books

Kimball R., Ross M. (2013) The Data Warehouse Toolkit: The Definitive Guide to Dimensional Modeling. Third Edition. John Wiley and Sons

Lidberg S. (2013) “Getting Started with SQL Server 2012 Cube Development”. Packt Publishing

Nam Hun Park, Kil Hong Joo (2014) “Query Processing on OLAP System with Cloud Computing Environment” International Journal of Multimedia and Ubiquitous Engineering, Vol.9, No.5, pp.169-174

Russo M., Ferrari A., Webb C. (2012) Microsoft SQL Server 2012 Analysis Services: The BISM Tabular Model. Microsoft

Sarka D., Lah M., Jerkis G. (2014) Implementing a Data Warehouse with Microsoft SQL Server 2012, Training Kit. Microsoft

Tim Hill. (2015) Excel 2016 for Windows Pivot Tables. Create Space Independent Publishing Platform