BUILDING A VISUALISATION DATA MODEL IN POWER BI

Abstract. Data is now crucial in every industry, and its role is especially important in the context of finance-related processes. In addition to collecting information from various sources, an equally important issue is its effective analysis and visualisation. Special software, such as Microsoft Power BI, is used for this purpose.

Power BI is a business intelligence service created by Microsoft. It is a response to the needs of users. Using a simple interface, it provides interactive visualisations and business intelligence opportunities, allowing you to create reports and dashboards. Effective use of this type of solution requires, first of all, properly selected data sources, which then need to be analysed to obtain useful business information and conclusions. It is also worth noting the main advantages of using BI systems:

- adapting data visualisation and presentation to the needs and expectations of a given organisation, department or specific user,
- integration with different data sources,
- information support for managers,
- improving the efficiency of the company based on current data and information available on the Internet.

The data visualisation opportunities offered by Power BI are useful for small companies, even one-person businesses, as well as large organisations. However, scalability is only one of the many advantages of this system. Also worth noting is the use of intelligent tools that automate much of the process, as well as support for
many security solutions, including privacy labels and file encryption. A huge number of available templates and easy-to-use graphical elements and charts mean that presenting any type of data and drawing valuable conclusions is no problem.

**Keywords:** data model, visualization, reports, business analysis, building connections.

Гавриш Богдана Михайлівна кандидат технічних наук, доцент, доцент кафедри інформаційних технологій видавничої справи, Національний університет «Львівська політехніка», вул. Степана Бандери, 12, м. Львів, 79000, тел.: (0322) 58-21-11, https://orcid.org/0000-0003-3213-9747

Шовгенюк Йордана Василівна студентка кафедри інформаційних технологій видавничої справи, Національний університет «Львівська політехніка», вул. Степана Бандери, 12, м. Львів, 79000, тел.: (0322) 58-21-11

Міщерякова Влада Олександрівна студентка кафедри інформаційних технологій видавничої справи, Національний університет «Львівська політехніка», вул. Степана Бандери, 12, м. Львів, 79000, тел.: (0322) 58-21-11

**ПОБУДОВА МОДЕЛІ ДАНИХ ВІЗУАЛІЗАЦІЇ В POWERBI**

Анотація. Дані зараз мають вирішальне значення в кожної галузі – їхня роль особливо важлива в контексті процесів, пов’язаних з фінансами. Крім збору інформації з різних джерел, не менш важливим питанням є її ефективний аналіз та візуалізація. Для цього використовується спеціальне програмне забезпечення, наприклад Microsoft Power BI.

Power BI – це служба бізнес-аналізу, створена Microsoft. Це відповідь на потреби користувачів. Використовуючи простий інтерфейс, він забезпечує інтерактивні візуалізації та можливості бізнес-аналітики, дозволяючи створювати звіти та інформаційні панелі. Ефективне використання такого типу рішень вимагає, перш за все, належним чином вибраних джерел даних, які потім необхідно проаналізувати для отримання корисної для бізнесу інформації та висновків. Також варто відзначити основні переваги використання BI-систем:

- адаптація візуалізації та представлення даних до потреб і очікувань даної організації, відділу або конкретного користувача,
- інтеграція з різними джерелами даних,
- інформаційна підтримка керівників,
- підвищення ефективності роботи компанії на основі поточних даних та інформації, доступної в Інтернеті.
Можливості візуалізації даних, які пропонує Power BI, корисні як у невеликих компаніях, навіть і з однією особою, так і у великих організаціях. Однак масштабованість — лише одна з багатьох переваг цієї системи. Варто також відзначити використання інтелектуальних інструментів, які автоматизують значну частину процесу, а також підтримку багатьох рішень безпеки, включаючи мітки конфіденційності та шифрування файлів. Величезна кількість доступних шаблонів і простих у використанні графічних елементів і діаграм означає, що представлення будь-яких типів даних і отримання цінних висновків не є жодною проблемою.

Ключові слова: модель даних, візуалізація, звіти, бізнес-аналіз, побудова зв’язків.

**Problem statement.** Creating a data model is essentially the process of creating the necessary relationships between the various data sources used in the model. Without an effective data model, the emergence of BI solutions would be impossible. Historically, traditional BI projects focused on building data models could take months or even years to complete when implemented within the rigid structure and constraints of a corporate business intelligence environment. Unfortunately, research shows that about 50% of all corporate BI projects fail. Such projects can cost hundreds of thousands to millions of dollars due to infrastructure, licensing and labour costs. Another reason for the low success rate is that businesses and end users often do not see visible results for many months, which can lead to growing frustration due to the lack of project "visibility" in key business areas. Extended project timeframes are often associated with the time required to gather requirements, develop a complex data model, and prepare and cleanse the original data sources. Many completed BI projects are often unable to deliver the expected results and often lack the components needed to perform the analysis that the business needs.

**Analysis of recent research and publications.** Business intelligence as it currently exists in Ukraine can no longer meet business needs due to the growth of large amounts of data (Big Data) that can no longer be processed without automation. The implementation of a business intelligence system completely changes the concept of company management: as a rule, most business managers give recommendations on how to resolve crisis situations, while the introduction of a comprehensive analytical platform Microsoft Power BI will combine, systematise and streamline all management and analytical reporting of the company, and automate the analysis and visualisation of data from various sources.

**The purpose of the article** is to study data models in Power BI Desktop, ways of processing and modelling complex relationships, as well as the performance of queries and processing of data models.
Summary of the main material. Data models in Power BI are designed specifically to extract analytical value from data and make informed business decisions based on them. Therefore, the data should be modeled in such a way that enables the creation of accurate and efficient reports. In Power BI, three types of data models can be distinguished: flat models, star schemas, and snowflakes.

A flat model, or in other words, a fully denormalized model, is a model consisting of a single data table without any auxiliary tables. In such a model, all measured elements and descriptive attributes are located in one table. This model is very common because many Excel users import data into Power BI directly from Excel tables. However, this approach is highly inefficient and has several drawbacks:

- The flat model does not scale very well; as the number of records in the table increases, such a data model will consume significantly more resources due to data redundancy and an increase in the number of columns;
- The flat model is not flexible and poorly adapts to future changes;
- The flat model is not "smart," meaning it can sometimes be confusing and difficult to understand;
- Calculations involving time series analysis, such as year-to-date (YTD) or year-over-year (YoY), are much more complicated to perform.

Another type of data model is the star schema, which is the best approach to data modeling in Power BI. The term "star schema" comes from dimensional modeling. Dimensional modeling is a method that businesses and organizations have been using for over 30 years to design their analytical data warehouses. The reason why such data models are called star schemas is because dimensional tables surround a fact table, somewhat resembling the shape of a star (fig. 1)

![Star Schema](image-url)
Fact tables store the measured elements in the data model, such as sales, tax amount, call duration. Dimension tables store descriptive attributes that help explain the modeled elements. Attributes related to specific elements are typically stored in separate tables. For example, product attributes such as name, color, size, weight, and others are stored in the product table.

Creating a data model based on the star schema in Power BI Desktop may require a bit more time, but it has several undeniable advantages (it's worth noting that the list below certainly doesn't cover all the advantages of the star schema):

- the data model is scalable, meaning it will smoothly grow with the addition of new data;
- the data model is flexible, so additional tables can be more easily integrated into the existing solution and support new analytical requirements that may arise in the future;
- data models based on the star schema are transparent and easy to understand;
- data models based on the star schema facilitate the execution of calculations related to time analysis.

The last type of data model is the snowflake schema. As the name suggests, the visual representation of this model more closely resembles a snowflake than a star. In dimensional modeling, most data models start as star schema models, but over time they may evolve into snowflake schemas, allowing them to support more advanced analytical requirements. A simple example of a snowflake could be breaking down a product dimension into several separate tables, such as product category, subcategory, and product itself.

The star schema is a method of data modeling that simplifies report creation and analysis. It makes understanding and working with the model as a whole easier. This modeling approach is part of a much broader topic called dimensional modeling. In dimensional modeling, two types of tables are considered: fact tables and dimension tables.

Facts are events that need to be measured. These can be product sales, student attendance, contacts with organization members, and any other events. The type of event can vary depending on the business unit or industry.

On the other hand, dimensions are what describe the facts (events). If a company earned $5 million, you can inquire about the year, the country, the manufacturers, or sellers involved in the transaction. Each of these highly relevant data inquiries indicates which dimension tables are needed in the model. Here, four separate dimension tables can be identified:

- date (year),
- geography (country),
- employee (salesperson),
- product.
Relationships in Power BI determine how different tables are related to each other. For example, the "Customers" table may be linked to the "Sales" table based on a column containing customer identifiers. It can be confidently stated that building relationships is the most crucial aspect of working in Power BI Desktop. Mechanisms such as automatic filtering of visualizations and reports, ease of creating measures and indicators using Data Analysis Expressions (DAX), and quick linking of various data sources all become possible thanks to well-built relationships in the data model.

Sometimes, Power BI Desktop creates relationships automatically. However, it's always necessary to check such automatically detected relationships to ensure they are correct. There are several characteristics of these relationships:

- Some relationships may be defined automatically.
- There can only be one active relationship between two tables.
- There can be an unlimited number of inactive relationships between tables.
- Relationships can only be built on one column, not multiple columns.
- Relationships are automatically filtered from the first to the next page.

![Fig. 2: Overview of connection elements](image)

A relationship is a line between two tables indicating that there is a connection between them.

Direction refers to the arrow indicating the direction of filtering.

First page - the number 1 indicates that the "Customer" table represents the first page of the relationship. This means that the key values on the first page of the relationship in this table are always unique.
"Other" page - the asterisk (*) indicates that the "FactInternetSales" table represents the other page of the relationship. The key (customer identifier) is stored in the sales table for each transaction, so the same key value can appear in the sales table multiple times.

The first thing to do after importing data is to check if all automatically detected relationships have been created correctly.

![Fig. 3: Starting the connection editor](image)

On the screen, you'll see the Manage Relationships editor window. The Relationships Editor is one of two places where you can create new relationships, as well as edit or delete existing ones. Figure 3 shows the appearance of the Manage Relationships editor.

In the section of the relationships editor window (Fig. 4), it is evident that there are three relationships, and all of them are active.

![Fig. 4: Verification of relationships](image)

The first row in Figure 4 displays the relationship between the FactInternetSales table and the DimCustomer table. The relationship between these two tables was created automatically using Power BI based on the CustomerKey column in each table. In this scenario, Power BI correctly chose the appropriate column names. However, if the relationship was created incorrectly, it would need to be manually modified.

**Conclusions.** Power BI Desktop removes nearly all barriers that have been the cause of the high failure rate of traditional BI projects. Power BI Desktop provides a much more flexible approach to building data models, so project timelines are often measured in days rather than months or years, costs are exponentially lower, and any missing components can easily be added if necessary.
Well-defined data relationships combined with simple yet highly effective usability improvements allow for building a data model that is both cohesive and flexible. In the past, creating BI projects required significant time and money investments. With Power BI Desktop and its self-service approach to business analytics, users now have the tools needed to create their own BI projects in a matter of hours and quickly obtain a data model with the information necessary to make informed business decisions.

References:

Література:
