Selected problems of designing modern industrial databases

Dariusz Karpisz[1,*] and *Anna* Kiełbus[1]

[1]Cracow University of Technology, Faculty of Mechanical Engineering, al. Jana Pawła II 37, 31-864 Krakow, Poland

Abstract. The paper presents problems of designing databases for various branches of industry. The development of information technologies and in particular of object-oriented programming has caused a change from data modelling to the modelling of applications. The increase of unstructured Big Data in Industry 4.0 era and requirements of sharing data model between many applications needs a reversion to data analysis and design and it is presented in the article.

1 Introduction

Modern ICT systems (Information and Communication Technologies) to support the operation of enterprises are based on computing tools such as ERP, CRM as well as CAD/CAM modelling systems [1, 2]. The databases are a specific part of these applications. Both approaches to the design of any applications and the design of databases is associated with a well-recognized theory [3].

However, there are problems related to the development of the programming method itself from the method of data analysis earlier to OOP (Object-Oriented Programming) currently. The common belief of programmers in the possibility of creating any application without data modelling is unfortunately still a fact. Currently, most databases are generated from application class models. They are used for more complex ORM (Object-Relational Mapping) modules to generate and utilize data models. In addition, companies have entered the Industry 4.0 era, the essence of which is access to a huge amount of previously unavailable, unstructured data referred to as BigData. Some of this data may be stored in relational systems, others in databases known as NoSQL (referring to non relational data models).

The combination of classic databases and fast unstructured databases is often not implemented by ICT applications, which slow down the process of loading data and their subsequent analysis or is implemented in many separate applications. The authors' experience shows that the preparation of applications for the industry often begins from the preparation of data model. This also relates to ensuring adequate performance and safety level as shown in articles [4].

---

[*] Corresponding author: dkarpisz@pk.edu.pl

2 Methods

There are many methodologies available to design applications at different levels of specification. One of the first methodologies was Information Engineering Methodology (IEM) [5] shown as a simple diagram in Fig. 1. The IEM assumed that focusing on data can identify the information needed by business. Invention of the relational data model [3] together with the theory of normalization helped to simplify the process of searching for a minimal data set of data model. The methodology emphasized in particular the modeling of data and the functions of their processing, which then forced the design and implementation of the application. System analysts (SA) and Data Base Administrators (DBAs) have thus influenced functional requirements (steering by data) and non-functional requirements (e.g. performance and model adaptability to changes) to applications (today most common associated with Front-End).
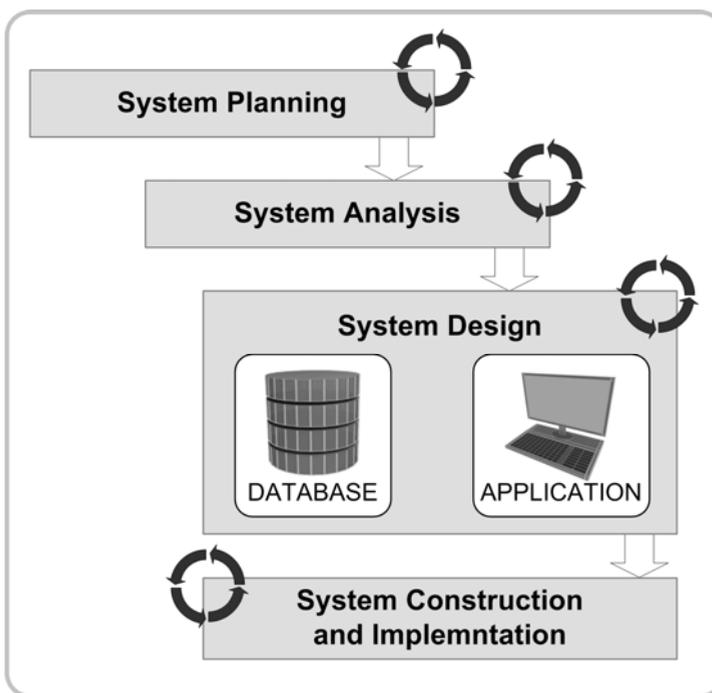

Fig. 1. Information Engineering Methodology.

It was important that the "System design" step assumed the parallel creation of both the data model and the application model by using multiple iterations. The database was often designed with overgrowth of tables, so that it contained tables and columns that can be used in subsequent versions of the application. The client application (as the presentation layer and also the logic layer) contained functionalities defined in functional and non-functional requirements defining appearance, available functions and their interrelationships (known as business logic). In order to support the design and implementation of software in IEM style, many tools and methods have been created under the common definition of CASE (Computer-aided Software Engineering) [6].

In response to the demand of the managerial staff, Business-Driven IEM variant was created. It gave the opportunity to better, proactively react to possible business changes in the organization. IEM has been completed with Strategic Planning to identify information required by management in strategic directions of organization changing.

As shown in Fig. 2, knowledge engineering was originally used to build ICT systems in many areas of the market according to the bottom-up design & development methodology. The evolution of programming languages and the acquisition by programmers competences of database designers, DBAs and data analysts, resulted in a general transition to the top-down application development model. It caused a drastic decrease in the quality of data. The data model created from the application class model in OOP style is compatible only with one application or its module. The resulting model is not optimized for data storage and search outside of the application. Correct data aggregation and analysis are also not possible due to the lack of proper data connections in the model - to correct data integrity, operation of the application from which the model was generated is always required.

Currently, especially in the *Industry 4.0* era, the importance of the data has grown again. When preparing modern production systems, the data model is analyzed again and this changes system design from top-down to bottom-up method.
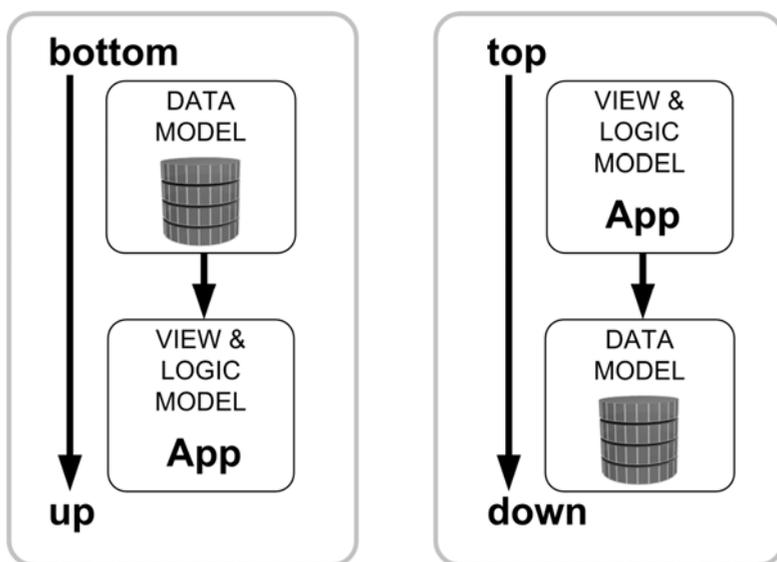

Fig. 2. Concepts bottom-up and top-down of ICT modeling.

3 Results

During the implementation of ICT systems projects for the machine industry, it turned out that precise determination of standards and factors are the most important ways of describing various technical and process objects. Both individual targets and their collections were analyzed.

The simplified ERD diagram from Fig. 3 shows the proposed way of storing information about any technical and non-technical objects. Each of the objects can be described by any number of associated standards with the joining entity "OBJECT Standards". The "Standards" entity can collect information related to the entire knowledge of the company e.g.:
- Official standards,
- Internal standards,
- Manufacturing experience,
- KnowHow.

In this way, management support was obtained using Knowledge Based Engineering (KBE) and also Knowledge Based (Expert) Systems. Entity "Standards" must be decomposed to the Physical Data Model (PDM) of database with these requirements:
- the possibility of cataloguing any standards, together with presenting both the vendor and the area of applying in manufacturing environment,
- the building of the hierarchy of standards/norms,
- the storing of the optional number of documents connected with standards.

A similar approach as presented for standards is necessary to determine the features of an object. These can be both descriptive and numerical features collected as a set of coefficients. The "Object" entity is described by a set of factors and it could be divided into many classes e.g. [4]:
- Nominal Factors - nominal monovalent,
- Range Factors - range from/to,
- Fuzzy Factors - described as fuzzy values (e.g. set: low, middle, high),
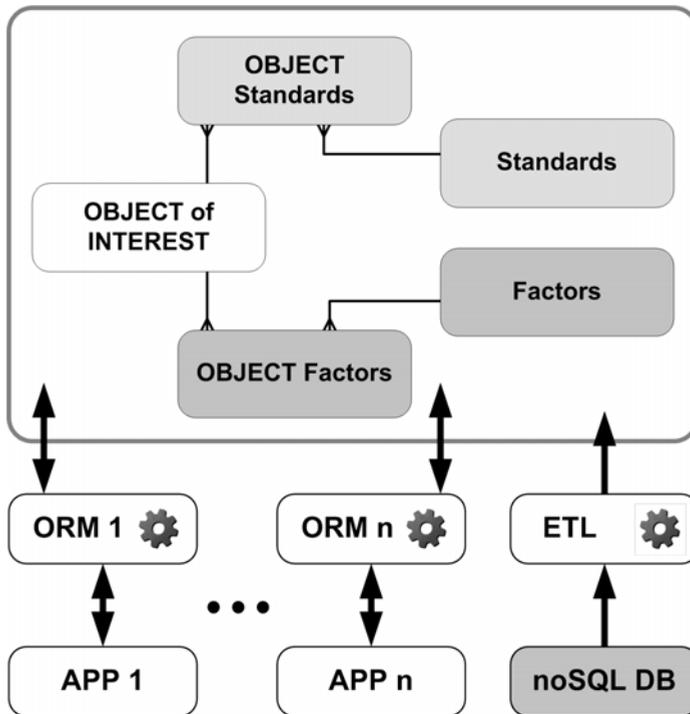- Timed Changing Factors - time-varying values.

Fig. 3. Universal data model to use for various applications.

The "Factors" entity is a universal dictionary for gathering information about any factors. The "Factors" should implement the following requirements:
- the building of the hierarchy of factors,
- the possibility of cataloguing any factor with a set of units of measurement,
- the possibility of units conversion,
- the possibility of connecting factor with any norms/standards to it description in industrial environment.

The relational combination of a set of factors dedicated to the description of specific objects is accomplished by the entity "Object Factors" as shown in fig. 3. The diagram does not show additional entities for "Standards" and "Factors" and their connections.

The proposed data model has been used to build industrial databases for the application of experimental machining of hard-to-machine materials [7], databases of the distributor and manufacturer of tools for the furniture industry and for the construction of a data model for the application of a distributor of tools and clamping systems for machining metals. In all three databases, a similar physical model was used (PDM). The target databases have been connected to dedicated applications written in the OOP style and to production and inventory management modules of MRP and PLM systems of enterprises.

4 Discussion

The proposed common data model for various industries has given the opportunity to build ICT systems optimized for data, not for programmers or IT companies. In fact, the target physical model as a set of tables of relational databases are much more extensive. With the "Standards" entity, tables for storing any files are additionally connected. The "Factors" entity has been expanded with a dictionary of units and a table combining a coefficient with two units and a conversion rule. It is possible to use advanced database mechanisms such as virtual column used to automatic computing. This brings the possibility of using active database mechanisms [8].

Databases built on the basis of the model shown in Fig. 3 can cooperate simultaneously with many applications. The box market by ORM1 is a connection of the target database with the client application such as WWW portal, similarly ORM2 mapping data from database to macros for CAM System. This gives the opportunity to share data for very different technical applications.

An interesting application of the proposed model was to execute data insertion attempts from documents data model database (NoSQL) to relational database. Data from the sensor with fast data acquisition to the database on the MongoDB server, after the completion of the acquisition, were automatically transferred to the relational database (Oracle 12c) using the mapping marked ETL in Fig. 3. This is the application that directly addresses the problems of processing and storing BigData in *Industry 4.0* concept. The ETL refers to technology known from data warehousing meaning:

- extract – involves extracting the data form NoSQL database,
- transform – involves transformation data to relational structures with changing of data types,
- load – store the transformed data to relational database.

In this way, the collected data can be processed using any analytical tools (e.g. build-in analytical SQL or BI data tools).

The problem in using the bottom-up approach is the habit of IT companies to quickly create applications with OOP from well proven applications frameworks. A good application programmer will not be a good data analyst at the same time. It is difficult to reconcile contradictory goals of user interface and data model designing. The programmers' assignment is to take care of the quality of the application. This is not connected to the quality of the generated data model, because in the top-down style, Information Engineering does not apply. Therefore, it is important to support the IEM approach for industrial applications.

The well-designed and well-implemented database may be very useful tool for the users located far from the software industry: image analyzers with their huge datasets [9, 10], especially in materials science [11-13] and medicine [14], the phytochemistry and the biotechnology [15, 16] or large-scale companies activity [17-20].

References

1. U.S. Access Board, *Information and Communication Technology (ICT) Standards and Guidelines, Notice of Proposed Rulemaking*, 36 CFR Parts 1193&1194, United States Access Board (2015)

2. R. Tongia, E. Subrahmanian, V. S. Arunachalam, *Information and Communications Technology for Sustainable Development, Defining a Global Research Agenda* (Allied Publishers, Bangalore, 2005)

3. E. F. Codd, *Commun. ACM* **13**, 6 (1970)

4. D. Karpisz, *Technical Trans.* 113, 73 (2016)

5. C. Finkelstein, *An Introduction to Information Engineering: From Strategic Planning to Information Systems* (Addison-Wesley, 1989)

6. A. F. Case, *ACM SIGMIS Database*, **17**, 35 (1985)

7. J. Gawlik, A. Kiełbus, D. Karpisz, *Solid State Phenom*. **223**, 35 (2015)

8. D. Karpisz, *Technical Trans*. **108**, 179 (2011)

9. A. Gadek-Moszczak, *Image Anal. Stereol.* **36**, 151 (2017)

10. A. Gadek-Moszczak, P. Matusiewicz, *Image Anal. Stereol.* **36**, 207 (2017)

11. J. Pietraszek, A. Gadek-Moszczak, A. Szczotok, *Solid State Phenom*. 197, 162 (2013)

12. J. Korzekwa, M. Bara, J. Pietraszek, P. Pawlus, *Int. J. Surf. Sci. Eng.* **10**, 570 (2016)

13. R. Dwornicka, N. Radek, M. Krawczyk, P. Osocha, J. Pobedza, *Proc. 26rd Int. Conf. on Metallurgy and Materials*, Brno, Czech Republic, May 24-26, 1252 (2017)

14. A. Gadek-Moszczak, J. Pietraszek, B. Jasiewicz, S. Sikorska, L. Wojnar, *Studies Comp. Intell.* **572**, 127 (2015)

15. E. Skrzypczak-Pietraszek, J. Pietraszek, *Acta Biol. Crac. Ser. Bot.* **51**, 62 (2009)

16. E. Skrzypczak-Pietraszek, I. Kwiecień, A. Goldyn, J. Pietraszek, *Phytochem. Lett.* **20**, 443 (2017)

17. P. Osocha, *Adv. Mat. Res.-Switz.* **874**, 89 (2014)

18. R. Ulewicz, J. Selejdak, S. Borkowski, M. Jagusiak-Kocik, *Proc. 22nd Int. Conf. on Metallurgy and Materials*, Brno, Czech Republic, May 15-17, 1926 (2013)

19. T. Lipinski, *Solid State Phenom.* **163**, 183 (2010)

20. T. Lipinski, A. Wach, *Proc. 23rd Int. Conf. on Metallurgy and Materials*, Brno, Czech Republic, May 21-23, 738 (2014)