Conceptual Modelling Styles

Mladen Varga
University of Zagreb, Faculty of Economics
mladen.varga@efzg.hr

Abstract. *Three steps in data modelling include conceptual, logical and physical modelling. The conceptual modelling models data from the users' perspective independently from any logical and physical implementation. The paper discusses different conceptual modelling styles that may apply in modelling transactional and analytical information system. They are distinguished by the modelling intention, the type of system decomposition and the way of modelling. The developed models may be equally expressive if they use the same but differently structured data which fulfil their different purposes*

Keywords. conceptual modelling, transactional system, operational system, data warehousing, analytical system

1. Introduction

Two main parts of an enterprise's information system are the operational (transactional) system and the decision support or the data warehousing (analytical) system. The paper discusses the problem of conceptual data modelling and investigates the main characteristics and differences of modelling data in the mentioned systems.

The *information system* is responsible for gathering, storing, processing and distributing an enterprise's data. Two main *information systems functions* are:

- transactional processing functions - *executing of transactions in order to assist enterprise's business processes*
- analytical processing functions – *assisting decision support functions by supplying information to decision makers*

The *transactional system* is part of the information system responsible for the executing of transactions of the enterprise's business processes. In the transactional system, many transactions are executed, most of them concurrently. A transaction inserts, updates, or in any other way processes the data in a database. In many occasions a transaction is an integral part of the business process. For example, a bank transaction is a business transaction and at the same time a transaction of the transactional part of an information system.

The *analytical system* is part of the information system responsible for the decision support functions or the data warehousing functions.

2. Data Modelling

Data modelling is an important task in the overall development of an enterprise's information system. Data modelling is a process of gathering the users' data requirements and building a database that satisfies these requirements. This process is usually decomposed into three steps shown in Fig. 1.

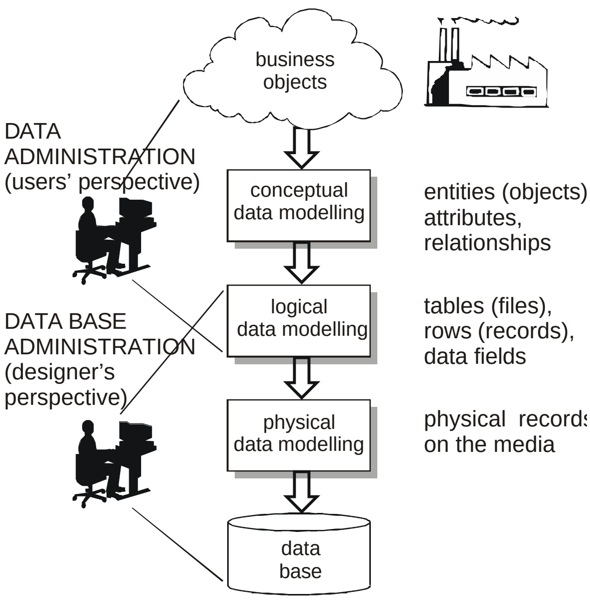

Figure 1: Data modelling steps

The first step is conceptual modelling that yields the conceptual model which is independent of any further logical or physical implementation. The conceptual model models users' data requirements in a conceptual manner using widely accepted concepts such as entities or objects, attributes and relationships. In this modelling paradigm

- A concept is an idea or notion that we apply to things, or object, in our awareness.

- An object (or entity) is anything to which a concept applies and it is an instance of a concept [3].
- A relationship defines the way various objects can be connected or linked.
- An attribute shows a property of an object or a relationship of objects.

The well-known entity-relationship and object models belong to this modelling paradigm. The conceptual model may be transformed into a logical (implementational) model such as the relational, the dimensional, the older network or hierarchical model, and even flat files models. Nowadays, the logical model most often used is the relational model. Its terminology includes terms such as tables (relations), columns (attributes) and rows (tuples). The logical model describes the logical implementation of the conceptual model. Furthermore the logical model is implemented through the physical model whose terminology uses such concepts as physical records and pointers. The physical model describes a real implemented database. Consequently, the database is a proper implementation of the physical model, its logical and conceptual model as well.

Conceptual modelling differs from logical and physical modelling greatly. The conceptual modelling is a task, which is done in the analysis phase, responsible for gathering, understanding and negotiating the users' data requirements. It belongs to the users' perspective. The conceptual model must reflect the users' view on the data used. Both logical and physical modelling are done in the design phase responsible for designing an optimal database. They belong to the designer's perspective and reflect the designer's view on the data used. Two different roles are associated with these tasks. Conceptual modelling is associated with the data administrator while logical and physical modelling is associated with the database administrator.

3. Conceptual Modelling Styles

A conceptual model is dominantly supposed to be entity-relationship model or object model. As the concepts on the data level are rather similar, we shall hereafter use the classical entity-relationship model and its terminology.

We shall consider the *conceptual modelling styles* of *transactional* and *analytical systems* and compare them in several aspects:

- modelling intention,
- system decomposition, and
- way of modelling.

3.1 Modelling intention

The most important element in distinguishing the modelling style of the transactional system and the modelling style of the analytical systems is the *modelling intention.* By intention we mean *what is modelled* i.e. what is basic aim we are trying to fulfil in the modelling process.

3.1.1 Modelling intention in transactional systems

The intention of the conceptual modelling of the transactional system is to model the *executing of the business processes*. The transactional system services the business processes by executing their transactions. How is a transaction involved in the business process? There are two contributions a transaction may fulfil in the business process:

- A transaction can be incorporated fully into the business process as its integral part so that it cannot be separated from it. For example, the money withdrawal at an automatic teller machine is done by the money withdrawal transaction.
- A transaction manages (inputs, changes, processes, outputs) and memorizes the business process data so that the status of the business process is transparent at any moment. For example, the money withdrawal transaction memorizes the data such as date, time, account and amount in a database.

The transactional system model must describe *the structure of data in executing the business processes*. The modeller of a transactional system models the flow of data of the business processes. The model is *the business process flow model* as the data in the flow of business processes are modelled. In other words, the intention of the conceptual modelling of the transactional system is to model the flow of the transactional data so that the model is capable to trace the execution of each transaction in the system, i.e. each transaction has to be traceable throughout the conceptual data model.

3.1.2 Modelling intention in analytical systems

The intention of the conceptual modelling of the analytical system is to model the *effects of business processes*. By the effects we mean consequences of the business processes represented by the data. They are generated from many transactions executed in the transactional system. For example, the cost of hospital admission is an effect that shows the price a patient has to pay for his/her admission in a hospital. As opposed to the transactional system, in the analytical system the trace of the transactions need not be visible. Only the effects of their execution have to be visible. The analytical model is therefore an *informational model* as *the data of the effects of the business processes are modelled.*

3.2 System decomposition

Modelling styles used in the transactional and the analytical system differ in the way of decomposition of the system into its components, i.e. *how the system parts constitute the whole system*.

3.2.1 Decomposition in transactional systems

In the transactional system the *whole business process* is decomposed into a set of smaller business processes. The number of decomposition levels is arbitrary. It depends on the size of the system and the way of decomposition. At the lowest level of decomposition are elementary processes that cannot be further decomposed. Usually, an elementary business process is implemented by a transaction. For example, the elementary process of the cash withdrawal at the automatic teller machine is implemented by the cash withdrawal transaction.

3.2.2 Decomposition in analytical systems

In the analytical system the decomposition is not connected with the business process itself but with the *business effects*. The business effects are decomposed into a set of elementary business effects. For example the Medicine effect in a hospital may be decomposed into the Medicine used and the Medical treatment effects.

What is the relation between the transactional and the analytical decomposition? In the transactional system the object of decomposition are business processes while in the analytical system the objects of decomposition are business effects. Are there any similarities in the process of decomposition? In practice, the business process suitable for conceptual modelling is in most cases an elementary business process. Modelling such a process yields the smallest integral data model. So the elementary process, which is the smallest integral business process, yields the smallest integral transactional data model. Such a data model can be treated as an integral conceptual data unit. At the same time an elementary business process produces the elementary business effect which yields the smallest integral analytical data model.

The conclusion is that *an elementary business process produces an elementary business effect*, i.e. there is strong relation between elementary business process and an elementary business effect. If, for example, the business process positioned on the lowest level of process decomposition produces two or more business effects this is the sign that it can be further decomposed into smaller processes.

3.3 Way of modelling

After the decomposition is determined conceptual modelling begins from the elementary level. At the elementary level there are elementary business processes in the transactional system while in the analytical system there are elementary business effects.

3.3.1 Way of modelling in transactional systems

In the transactional system where *the flow of business process is modelled* the elementary processes are considered, i.e. the execution of each transaction must be traced. The entities that are involved in the execution of transaction must be incorporated in the model.

As a rule, the entities modelled first are fundamental (strong, independent). The weak (dependent) entities that are paired to them are modelled next. For example a purchase order is a fundamental entity that is modelled first, a purchase order item is a weak entity dependent on the purchase order that is modelled next. Associative entities or relationships between fundamental entities are modelled after fundamental and weak entities are known.

The structure of the transactional model shows *the relationships between the fundamental entities that are important in the execution of transactions of the business process*.

3.3.2 Way of modelling in analytical systems

In the analytical system where *the effects of business are modelled* the elementary business effects are considered. The effects are represented by various business measures or facts, which are usually, but not necessarily, numeric. In statistics these attributes are summary attributes. The measure is associated to some perspectives or dimensions. These associations to perspectives or dimensions determine the meaning and significance of the measure. The measures are always contained in associative entities. This is the reason that associative entities are modelled first in the analytical system. Fundamental entities, that show the perspective or dimensions of the measure, are modelled next.

The structure of the analytical model shows *the relationships of dimensional or perspective entities*, which are attributed by one or more business measures, or *the relationships of associative entities to fundamental (dimensional) entities*.

4. Example

The transactional and the analytical modelling are shown in the following small example of a typical hospital information system.

4.1 Transactional model

The decomposition diagram of the business Hospital process is shown in Fig. 2.

- Hospital process
  - Patient admission
    - Patient registration
    - Patient examination
    - Patient transfer to ward
  - Patient treatment
  - Patient discharge

Figure 2: The Hospital process decomposition

The Hospital process process is decomposed into the Patient admission process, the Patient treatment process and the Patient discharge process. The Patient admission process is further decomposed into the Patient registration process, the Patient examination process and the Patient transfer to ward process. The Patient registration process describes an administrative registration of the patient into the hospital that is the prerequisite of first hospital examination. The Patient examination process includes medical examination of the patient, which is done by the admission doctor. The doctor diagnoses the patient, decides if the patient must remain in the hospital and directs him to the hospital ward. The Patient transfer to ward process describes the situation after the patient comes to the ward. He is examined there by the ward doctor and diagnosed one or more diagnoses. It is decided whose doctor will cure the patient. The processes the Patient treatment and the Patient discharge are not further decomposed. The Patient treatment process describes various medical treatments of the patient and the usage of medicines, which are required by the hospital protocol. The Patient discharge process is responsible for the hospital cost calculation based on the number of days spent in the hospital, medical treatments and medicines used. This process states the result of the hospital stay. At the same time these processes are the transactions executed within the Hospital information system.

In Fig. 3 the transactional model of the Patient admission process is shown. The Patient registration entities with their relationships are PATIENT, TOWN, ADMISSION and DOCTOR (admission doctor), the Patient examination entities are PATIENT, ADMISSION, DOCTOR (diagnosis doctor), DIAGNOSIS and ADMISSION_DIAGNOSIS, and the Patient transfer to ward entities are PATIENT, ADMISSION, DOCTOR (doctor in cure), WARD and ADMISSION_WARD. The transactional model is depicted in the form of the entity-relationship diagram.

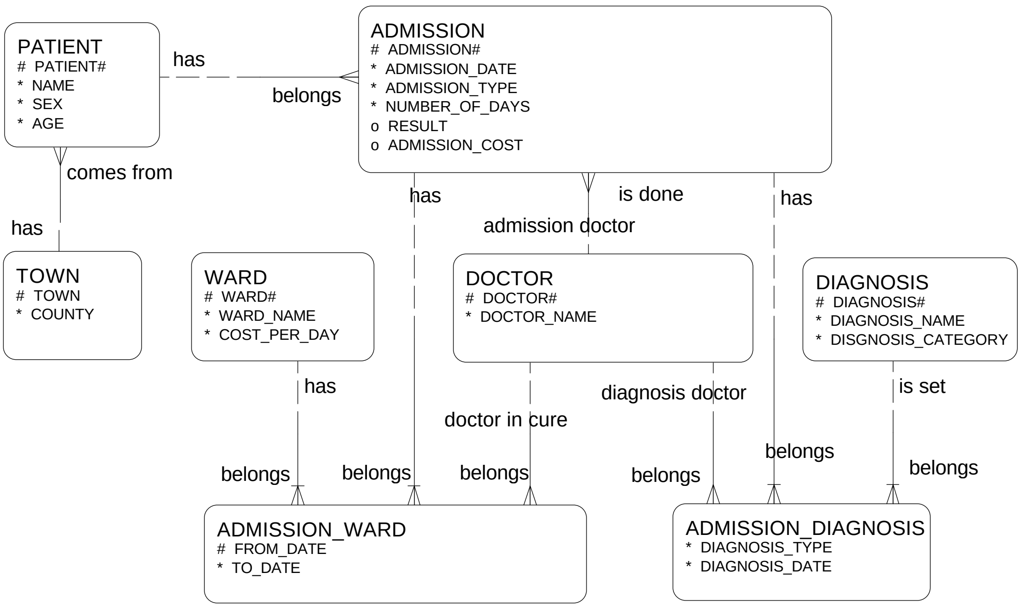

Figure 3: The transactional model of the Patient admission process

4.2 Analytical model

The decomposition of the Hospital effects is shown in Fig. 4.

- Hospital effects
  - Medicine effects
    - Diagnostics
    - Treatments
    - Medicines
  - Accommodations
  - Costs

Figure 4: The Hospital effects decomposition

The overall hospital effects may be decomposed to Medicine effects, Accommodations and Costs. Furthermore the Medicine effects are decomposed to Diagnostics, Treatments and Medicines. The Diagnostics effect describes the diagnosing of the patient. The Treatments effect describes the medical and nonmedical treatments applied to the patient. The Accommodations effect describes the accommodations of the patient in the various hospital wards. The Costs effect describes hospital cost of the patient's stay in the hospital.

The analytical model of the Accommodation effect is shown in Fig. 5 in the form of the entity-relationship diagram. The central entity is ACCOMMODATION that is the associative entity connected to a series of dimensional entities: AGE, SEX, TOWN, PATIENT, WARD, ADMISSION, DOCTOR (admission doctor, doctor in cure) and TIME (admission date, accommodation date).

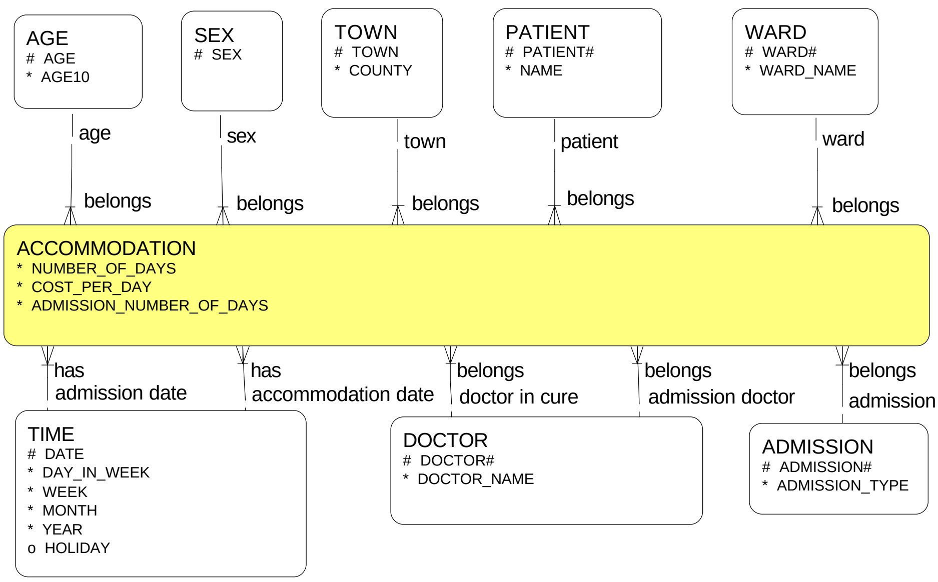

Figure 5: The analytical model of the Accommodation effect

4.3 Comment on relation between transactional and analytical decomposition

The following business processes yield business effects: the Patient registration process yields no business effect, the Patient examination process yields the Diagnostics effect, the Patient transfer to ward process yields the Accommodations effect, the Patient treatment process yields the Treatments and the Medicines effects and the Patient discharge process yields the Costs effect. Since the Patient treatment process yields two business effects it is not an elementary business process and it is a candidate for further decomposition. Obviously the Patient treatment process may be decomposed to two processes: the Patient medical treatment process and the Patient medicine usage process. Each of these processes yields one effect: the Patient medical treatment process yields the Treatments effect and the Patient medicine usage process yields the Medicines effect. The relation between the transactional process decomposition and the analytical effects decomposition is depicted in Fig. 6.

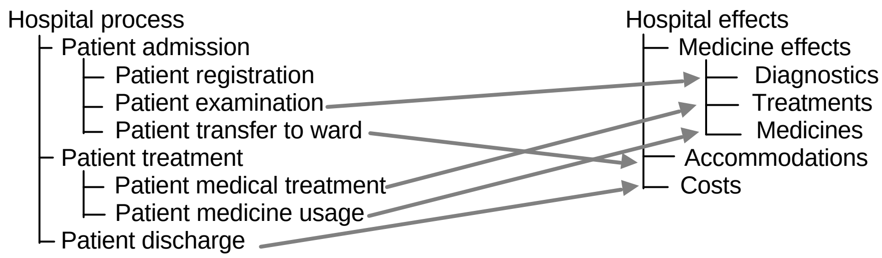

Figure 6. Relation between transactional and analytical decomposition

5. Equivalence of analytical and transactional model

Several questions may be raised:

- Can an analytical model represent a transactional model or vice versa?
- Is it possible to transform a transactional model of the enterprise into an equivalent analytical model?

The answer on both questions may be positive but some conditions must be fulfilled. The first question introduces the problem of equivalence of the analytical and the transactional model. In the previous discussion we have confirmed that an elementary business process yields an elementary business effect. So the basic issue is the equivalence of the smallest (elementary) transactional data model and its corresponding analytical data model.

The first and obvious condition of the equivalence of the analytical and the transactional data model is the same set of data in both models, i.e. both models must be equally expressive. The hypothesis is the existence of two models of the same set of data that are equally expressive but with a different purpose and a different structure that fulfils this purpose. The expressiveness of both models will be retained only if associations of the same quality between the corresponding entities (objects) exist in both models.

The second question is raised because of the usual need for transforming a transactional model into an equivalent analytical model. The answer depends on the quality of the transactional model, i.e. on the relationships used in the transactional model [1]. If there are associative entities, n-ary relationships or many-to-many binary relationships in the transactional model it is possible to define the corresponding analytical model. Only the presence of the mentioned types of relationships makes the transformation of a transactional into an analytical model possible. But, if the transactional model has no such relationships then there is no direct possibility of its transformation into an analytical model. The next question is: Has a transactional model with no associative entities, n-ary relationships or many-to-many binary relationships data suitable for successful transformation into an analytical model? Probably not if there are no attributes that have the characteristics of the measure attributes [2] of the analytical model. Without such attributes, business performance cannot be evaluated and the primary purpose of the analytical model used in the data warehouse cannot be fulfilled.

In the correct transactional model the entities containing measure attributes are associative entities or many-to-many binary relationships or n-ary relationships. The measure attributes usually do not belong to any fundamental (strong) entity but are properties of associative entities or n-ary relationships or many-to-many binary relationships among the fundamental entities.

The aspect that makes the transformation of the transactional into the analytical model more difficult is the rule used by many designers of transactional models. This rule says that n-ary relationships are rare in the real world and that they have to be avoided. Nevertheless, analytical modelling relies on modelling n-ary relationships.

In the correct transactional model, which is intended to be transformed into the analytical model, all entities must fulfil their roles entirely, that is every fundamental entity must be fundamental without any characteristics (attributes) of the associative entity and every associative entity must be only associative entity without any characteristics (attributes) of the fundamental entity.

The procedure of transforming the correct transactional into the analytical model is presented in [4]. The procedure shows that measure attributes of transactional models are settled in entities regardless of their type (fundamental, weak or associative) while measure attributes of analytical models are exclusively settled in associative entities. Measure attributes of an elementary data model are also concentrated in one associative entity in analytical models and possibly dispersed in more entities in transactional models.

The main steps of transforming the correct transactional into the analytical model are:

1. Identify associative entities, n-ary relationships or many-to-many binary relationships in the relational model.
2. For each many-to-many binary relationship, n-ary relationship or associative entity:
   a) Identify the measure attributes.
   b) Identify the corresponding dimensions, i.e. strong entities involved in the relationship. For each dimension identify dimensional attributes. Some attributes may become measure attributes.
   c) Identify the attributes of associative entities or n-ary relationships that define new dimensions. The time attribute (date, for example) is frequently used in the associate entity or in the n-ary relationship. Open the new dimension (time, for example) and define other dimensional attributes.
   d) Identify additional dimensions from distant entities. A distant entity is not directly involved in the many-to-many binary relationship and neither in the n-ary relationship, nor is it directly associated to the associative entity. It is reachable from the associative entity via the relationship of cardinality 1, showing that the associative entity functionally determines the distant entity. The distant entity can be reached by the same rule from the previously found distant entity. The attributes of each distant entity must be carefully examined while their attributes can apply both as dimensional attributes and as measure attributes.

The steps applied to the transactional model in Fig. 3 to transform it, beginning from the ADMISSION_WARD associative entity, into the analytical model in Fig. 5 are:

1. ADMISSION_WARD is the associative entity of the transactional model.
2. For ADMISSION_WARD associative entity:
   a) No measure attribute exists directly but the measure attribute NUMBER_OF_DAYS, showing the number of days patient spent in the ward, can be computed from FROM_DATE and TO_DATE.
   b) Attributes that identify dimensions are WARD#, DOCTOR# and ADMISSION#. Their corresponding dimensions are WARD with the dimensional attributes WARD# and WARD_NAME, DOCTOR (in cure) with the dimensional attributes DOCTOR# and DOCTOR_NAME and ADMISSION with the dimensional attributes ADMISSION# and ADMISSION_TYPE. COST_PER_DAY from WARD becomes the measure attribute. The attribute NUMBER_OF_DAYS from the ADMISSION entity becomes the measure attribute as well. Since it is the second attribute with the same name it is renamed in ADMISSION_NUMBER_OF_DAYS. Also, a new dimension is identified by ADMISSION_DATE attribute. This is TIME (admission time) dimension.
   c) The new dimension identified from the attributes of the ADMISSION_WARD associative entity is TIME (accommodation time). The dimensional attribute is FROM_DATE. The TIME dimension is

properly modelled with a set of dimensional attributes DAY_IN_WEEK, WEEK, MONTH, YEAR and HOLIDAY.

d) The new dimensions identified from distant entities are as follows: A distant entity not directly connected to the ADMISSION_WARD associative entity is PATIENT. The ADMISSION entity, which is functionally dependent on ADMISSION_WARD, determines the PATIENT entity via "belongs" relationship where its ADMISSION# key functionally determines the PATIENT's key PATIENT#. The attributes of PATIENT identify new dimensions PATIENT, SEX and AGE. The attribute NAME of the PATIENT entity is the dimensional attribute of the PATIENT dimension. Via "comes from" relationship a new distant entity is identified. This is TOWN, because PATIENT# functionally determines its key attribute TOWN. The COUNTY attribute from the TOWN entity is the dimensional attribute that belongs to the TOWN dimension and defines the TOWN dimension hierarchy: TOWN→COUNTY. From the ADMISSION entity via "is done" relationship the admission DOCTOR is identified and the new DOCTOR (admission doctor) dimension is specified.

6. Conclusion

For the same set of data it is possible to develop two different data models: the transactional model for modelling business processes and the analytical model for modelling business effects. Development of two different models is possible by using different modelling styles during the conceptual modelling. They are distinguished by the modelling intention, the type of the system decomposition and the way of modelling. Both models, transactional and analytical, depict the same problem and are equally expressive if they use the same set but differently structured data which fulfil their different purposes.

The main characteristics of the conceptual transactional and analytical modelling styles are summarized in Table 1.

	Transactional modelling	Analytical modelling
Purpose	Data modelling of transactional systems	Data modelling of decision support systems
What is modelled	Execution of business process → *process flow modelling*	Effects of business process → *process effect modelling* or *informational modelling*
System decomposition	Decomposition of business processes	Decomposition of business effects
Way of modelling	1) Definition of fundamental entities 2) Definition of associative entities or relationship between fundamental entities	1) Definition of associative entities and their attributes (business measures) 2) Definition of fundamental entities (business perspectives or dimensions)

Table 1: Difference between transactional and analytical modelling

7. References

[1] Firestone, J. Dimensional Modeling and ER Modeling in the Data Warehouse; 1998. http://www.dkms.com/dmerdw.html [12/12/2001]

[2] Kimball, R. The Data Warehouse Toolkit. John Wiley, New York; 1996.

[3] Martin, J., Odell J. Object-oriented Methods. A Foundation. Prentice Hall, Englewood Cliffs; 1995

[4] Varga, M. On the Difference of Relational and Dimensional Model. In Proceedings of the 12th International Conference on Information and Intelligent Systems IIS 2001, 2001 Sept 26-28; Varaždin, Croatia. Faculty of Organization and Informatics Varaždin, University of Zagreb, 2001. p. 245-251.