

Databases – evaluation module

1. Background and objectives

Many distributed systems or information services cannot function without databases or data pools. It is usual that an information service or application utilizes information stored in several separate databases. It is quite obvious that semantical correctness, consistency, integrity and availability of data are essential prerequisites to a successful implementation of an information service or a new system.

In many cases there is a need to evaluate the functionality of a database or a data pools, the quality of their implementation, the way the information has been modelled and the processes in which the information stored in the database is gathered, maintained and removed.

An evaluation of a database or a data pool can be made, when its implementation and processes which add, update or remove data are reasonably well known. An information service or a system can be evaluated before full-scale implementation or when an information service or a system has been realised.

This module can be used to evaluate the databases and data pools used to realise a particular information service or application regardless of the fact, who has the administrative responsibility or who produces the data stored in them. Questions concerning data protection or information security are mostly outside of the scope of this module.

2. Description of the evaluation process, defining the scope and choosing relevant points of view

The basic idea of the evaluation of databases is to compare the existing or planned databases against the information needs and other requirements set by the information service to be evaluated. By comparing the existing and planned databases to the requirements set by the information service, one can answer the question, whether the requirements have been or will be met and whether the needed information is available or not.

The process of evaluation is usually started by defining the scope of the evaluation. A good way to start is to define, to which questions the evaluation should give an answer. When using the databases evaluation module, it is fairly common that there is a need to include also other modules such as information security or revenue and finance in the evaluation.

A good point to start the evaluation is to define, what information is needed by the information service in question. An explicit definition of the needed information helps to recognize, what is the information the service is dependent on and where this information is gathered, stored and processed.

The second step is to find out, how the needed information is gathered, preprocessed, stored in databases and provided for use in practise. It is useful to know at least to some point the whole chain from raw data collection to processing and storage in particular databases. This means that one has to describe at least briefly the processes, which import, change or remove information in those databases. The way this information is obtained depends on the service to be evaluated.

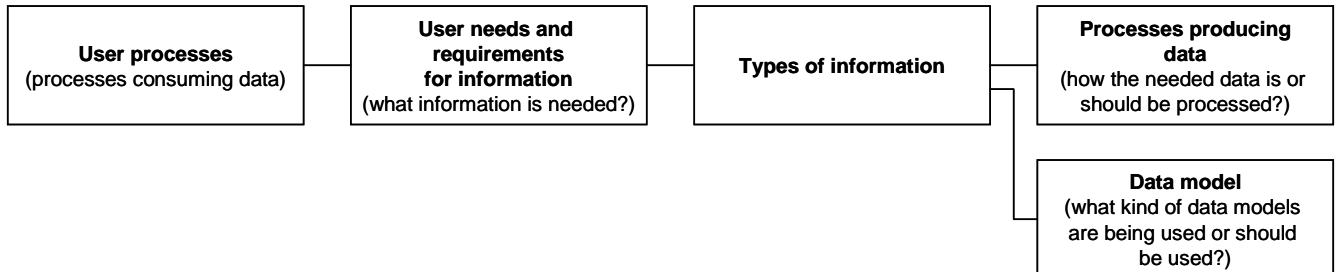
The data models, that the databases are based on, are then described using a suitable modelling method such as relational model or object-oriented modelling. In case of existing databases, these models can probably be obtained from the documentation of the database in question. The data models obtained are then compared against the phenomena modelled with the data model. When looking at the data models, one has to be careful with the implicit or openly written assumptions built in the data model. Comparison of data models to real-life phenomena has potential to highlight the obvious inconsistencies between data models and reality or uncover the assumptions that were made once but are no longer valid.

The fourth part of the evaluation process is to describe the processes that import, modify or remove information in the database or databases in question. Those processes can be described using specific process diagrams and tables.

The scope of evaluation should include also the temporal dimension in addition to availability, integrity and consistency. The time, in which the information is collected as raw data, transmitted, put into database and made available for use, is highly relevant question especially in case of information services providing real-time or almost real-time information.

3. Evaluation methods and metrics for evaluation

An overview of the evaluation method is the picture below.



1

1. Describe the processes that use the information stored in a database

The first part of the evaluation is to list the actors involved and to describe their different processes that use the information in the database. A good way to describe the processes that consume information is to write in a table for every process the name of the process, a general description of the process, relevant inputs, relevant outputs and general remarks if needed. The processes should be described at somewhat abstract level to cope with their inherent complexity. The processes should also be numbered so that referring to them in the latter stages of work is possible. If possible, the actors involved should be referred with abstract roles such as “ICT service provider” instead names of existing organisations such as “BitCo Ltd”.

When the processes that consume information have been recognised and described, the next step is to define the requirements that these processes have for information stored in the database or databases to be evaluated.

Number	Name of process	Description of process	Input	Outputs	General remarks

2. Define the requirements set by the processes that use the information stored in the database

When the processes that consume information are known, it is possible to list the needs and requirements of the different processes. The simplest way to do this is to list all types of information needed by the process in question, and do this for all relevant processes. Depending on the scope and objective of evaluation, these processes may be, for example, processes needed to implement a particular information service or processes consuming information stored in a system to be evaluated.

For every type of information needed by a process, the number of the row, the name of the process, the type of needed information, a description of the information content and possibly some general remarks will be filled into the table below.

Number	Process	Actor	Needed information	Description of information	General remarks

The result of this stage of the evaluation process will be a list of the different types of information needed by different processes of different actors involved.

3. Describe the types of information

When there is a list of the types of information needed by the different processes, it is possible to list shortly, what are the main types of information needed. The result will not be a data model as such. The objective behind compiling the list is to summarize the needs of different processes into the same table. The types of information the different processes need may be very similar or completely different. For every type of information, a short description, the source of information and possibly some remarks are entered into the table.

Numero	Type of information	Description	Source of information	Remarks

4. Describe the processes that produce information

When the types of information have been defined, one can describe the processes in which the information is produced and imported to a database or databases. The method is to first describe the objective, how the different processes should produce the needed information and import it to a database, and then describe the existing processes: how the information is produced and imported to the database at present. The third step is to analyse the gap between the objective and the current situation.

The first step is to outline, what kind of processes are needed to produce the needed types of information and import it into a database. These processes can be listed into a table, which includes the names of the processes, their descriptions, inputs and outputs and possible special remarks.

Number	Name of process	Process description	Inputs	Outputs	General remarks

The second step is to gather information on the existing processes or processes in a plan subject to evaluation. This information can be collected in a workshop organised for the purpose, from written documentation or guidelines for operation or with expert interviews. The existing or planned processes should be listed in a table in the same way as processes outlined in the previous step.

The third step is to make a gap analysis between the existing processes and the objective defined in the first step. The analysis should include all processes that produce and import information in the database as well as processes which alter or remove this information.

The flow of information between processes can be visualised with a process map which consists of the processes included in the study and data flows drawn as arrows between them. The data flows can be listed in a separate table. Every data flow between processes has a name, a general description and description of its contents and two processes between which the information flows.

Number	From	To	Name	Description	Transferred information

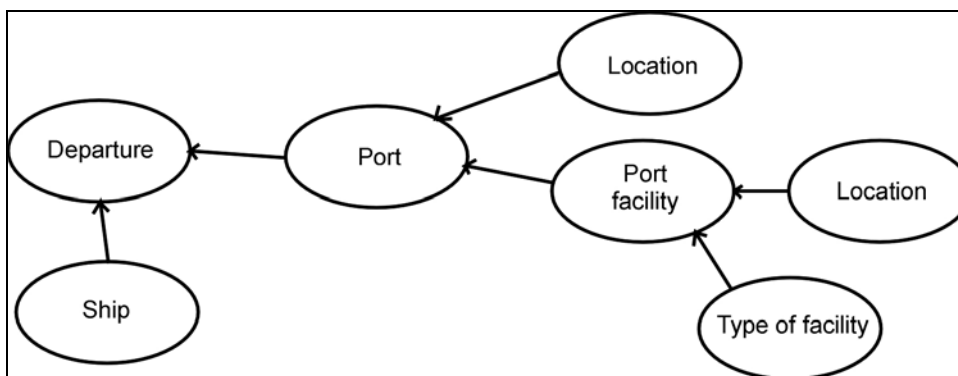
When analysing real-life systems or services, one has to consider practical but often overlooked questions such as, how well and reliably the previously described processes work in real life, what will be the quality of data they produce, and will the information be up to date when it is available for use. For example, different organisations may compile statistics in a different way, or the information stored in a database may become outdated unless it is maintained and updated appropriately.

5. Formulate the data model on the basis of requirements

When the types of processed information are known, it is possible to formulate a data model. The first step is to formulate the data model on the basis of real-life phenomena the processed information is related to. The important questions are, what data modelling techniques to use, what objects to include, how the chosen objects are related to each other, what kind of attributes they have and what kind of assumptions can be made. Common data modelling techniques such as relational model and object-oriented modelling can be used in most cases.

A conceptual model can be formulated to describe, how the different entities and attributes relate to each other. The first task is to recognise relevant entities or objects and identify their attributes. The second task is to identify the relationships between the different entities. The third task is to choose the data modelling techniques one wants to use for presentation and analysis.

A relatively simple method to present the entities and their attributes in the same picture can be seen in a picture below. Attributes that cannot exist without the entities they are related to have been connected to the entities with arrows. Entities that cannot exist without other entities have been connected with arrows to the entities they are related to.



Kuva 2. An example of conceptual data model.

The second step is to obtain or draw the data model of the database or databases to be evaluated. The data model of the existing database or databases to be evaluated can be obtained from written documentation in some cases. If the data model isn't available in written or printed form, it has to be generated on the basis of other available information such as guidelines for operations or plain text source code. Also in this step, one is recommended to choose the most suitable data modelling techniques. For example, a relational database can be best described with a relational model.

The third step is to compare the data models formulated on the basis of real-life phenomena to the data models used to implement the database or databases to be evaluated. The first check is to compare the needed types of information listed in stage 3 to the data model of the existing or planned databases under evaluation. This comparison gives some guidance, whether the needed information can be obtained from database or databases under evaluation or not.

The second question is, whether the data model the database is built on is consistent with the outside world. This consistency check can be made by comparing the data model generated on the basis of phenomena to be modelled and the data model the database has been built on side by side. This comparison usually reveals problems related to the way information has been modelled during the planning or implementation of the database under evaluation.

6. Practical implementation

Evaluation of a database should also include an evaluation of the practical implementation. The overall question is, whether the hardware and software platforms used to implement a database or planned to be used are suitable for the purpose. Suitability for a practical purpose can be described as a combination of adequate capacity, acceptable performance, scalability for future needs, fault-tolerance and ease of maintenance.

When evaluating an existing system, it is possible to measure some things such as response times and percentage of failed transactions. Others such as the details of the hardware and software used can usually be obtained from the documentation of the system to be evaluated.

Future needs such as growth in the number of users and the volume of data have to be taken into account when estimating the adequate data storage and processing capacity and needed bandwidth for data communications.

4. Results of evaluation

The evaluation answers the question, how well the databases and data pools, data models and processes, in which information is produced and stored in databases, respond to the needs of different users. The evaluation may give recommendations for future actions or improvements to make. The data model, on which the database is built on, is also documented and visualised as well as data flows between different processes which produce, alter and remove data.

Possible recommendations for future action are, for example, evaluation of the system with other Evaserve modules with different points of view, improvements or changes in the evaluated database or databases or improvements to processes that generate or update the data stored in the database.

5. Reporting of the conclusions and recommendations

The results of the evaluation are reported as text in suitable format. Tables, figures and other material produced during the evaluation process can be included in the document. Basic information of the author and the organisation which funded the work must be included in the evaluation report as well as the objectives of the evaluation, methods used, possible limitations of scope and the points on which the evaluation was focused on

Text included in the evaluation report must be clearly written and comprehensible for anyone with needed skills in written English – not just readers with their background in data processing or communications engineering. Sources, from which information has been obtained, must be marked in the evaluation report. The authors own judgements and conclusions should also be clearly separated from facts or views presented in oral interviews or written sources.

6. Connections to other modules

Information security

Information security has its own module in Evaserve evaluation system. Information security module of Evaserve contains most aspects of information security such as evaluation of risks related to information security and engineering of secure systems. The evaluation of information security is most useful in the design and development part of the life cycle of a system or an information service. Making changes to existing systems or organisational structures may turn out to have high costs.

In case of databases, a practical question of information security is recovery - how the system recovers from different types of failures or attacks.

Revenue and finance

Every information service has its own economics. Running large databases with high availability is costly, and even more costs are caused by updating the contents. The only thing that is certain, that someone has to bear these and other costs related to production, storage, processing and dissemination of data.

Evaserve module Revenue and finance can be used analyse the commercial and socio-economical profitability of an information service and business models related to it.

System analysis

An individual database or data pool is usually a part of larger information architecture. Information architectures have been discussed in the Evaserve System analyses module.

7. Literature and web pages

Entity-Relationship Model <http://www.cs.sfu.ca/CC/354/zaiane/material/notes/contents.html>

Object Role Modeling <http://www.orm.net/>