

Present and Future Possibilities of Database Technology

G. M. NIJSSEN

Nijssen, G. M.: Present and Future Possibilities of Database Technology. *Medinfo 1974*. pp. 1-4.

The trend in information systems is towards databases. The concept database is defined and the major requirements for database management systems are shortly discussed, as well as the basic concepts of these systems. The future possibilities of database management systems are deried, mainly by extrapolating the present trends. The future possibilities are very promising, specially for the non-computer specialists. No longer is he forced to learn a lot of cumbersome computer jargon in order to be able to use the computer. However, the advanced education of database technologists is a necessary condition for success.

G. M. Nijssen, Control Data Europe, Brussels, Belgium

1. Trends in information systems

In the past few years the database concept has received an increasing degree of attention. The fast increasing degree of attention of the last two years alone is a signal for many people that the database concept offers some important possibilities.

The following two trends can be observed in information systems:

- (1) Independent information systems, covering non-isolated subsets of a larger organization, are integrated in order to get better control over the real operations.
- (2) There is an increasing need for many non-computer specialists (which may be specialists in other areas such as doctors, managers, clerks, secretaries, nurses, etc.) to have direct access to (the computer and) databases, and to formulate their requests directly to the database system without intervention of programming specialists.

2. Database

The database is the common basis for many information systems, serving many applications and many users. Fig. 1 contains a picture which illustrates the database concept.

A database contains data which describe a certain part of the real world, viewed however through a certain filter. The filter is associated with a certain model, in

which only certain aspects of the reality are considered of interest; and all others are filtered out. In this way it

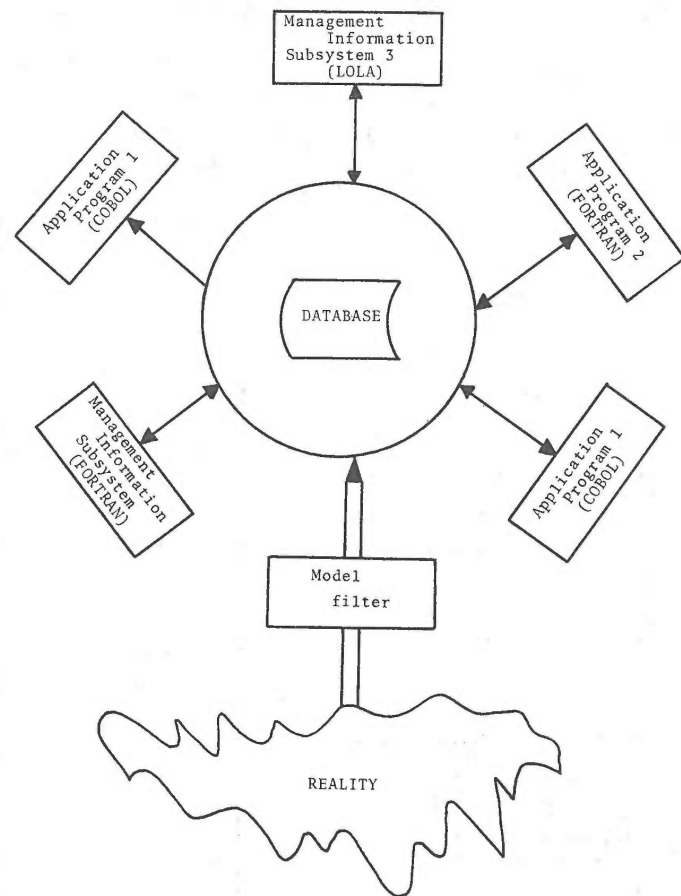


Fig. 1.

is possible to have several different models (and associated databases) which are related to the same reality.

Just as in every model, the degree of difference between the model and the reality is controllable. In other words, the number of entity classes, the relations between entity classes and the number of attributes of an entity class, are variable, and selectable in each case.

Connected with the database in Fig. 2 are five different programs; application programs written by professional programmers in languages like COBOL or FORTRAN, or programs to serve management needs written by professional programmers in COBOL or FORTRAN. However, there is also connected a program to serve management which is written by a manager or his delegate in a Layman Oriented Language, LOLA [1].

In other words, many different people use the database, and each via his own language.

We now quote the definition of a database as given in [1]: "A database is:

- a collection of data,
- which is used by several individuals
- and which is associated with a model."

3. Requirements for database management systems

The collection of software modules (written by database technologists) that provides the user with the languages to describe and access the database (as defined in section 2) is commonly referred to as a DBMS or DMS (Data

Base Management System). The requirements for such a DBMS may be derived from the definition of the concept of database as given in section 2. A DBMS should provide:

(a) powerful structuring tools (in order to be able to work models which can reflect a selectable number of aspects of the reality);

(b) many different languages both for (i) non-computer specialists and (ii) computer specialists;

(c) a great degree of flexibility between user programs and the way the database is organized in a machine (in database jargon, this is often referred to as data independence); this will avoid much reprogramming as well as other disturbances as databases evolve to continually reflect the reality;

(d) tools to enable the responsible people to implement privacy and integrity rules;

(e) tools to enable the database administrator to monitor, restructure and reorganize the database;

(f) hardware independence in order to give the user freedom to select any hardware (an independent organization CODASYL (which designed COBOL and still maintains it) has greatly contributed to user independence from hardware by releasing specifications for database languages) [2], [3], [4], [6].

4. Database management system concepts

The DBMS concepts might best be illustrated by a very short description of the "only proposal of stature for a

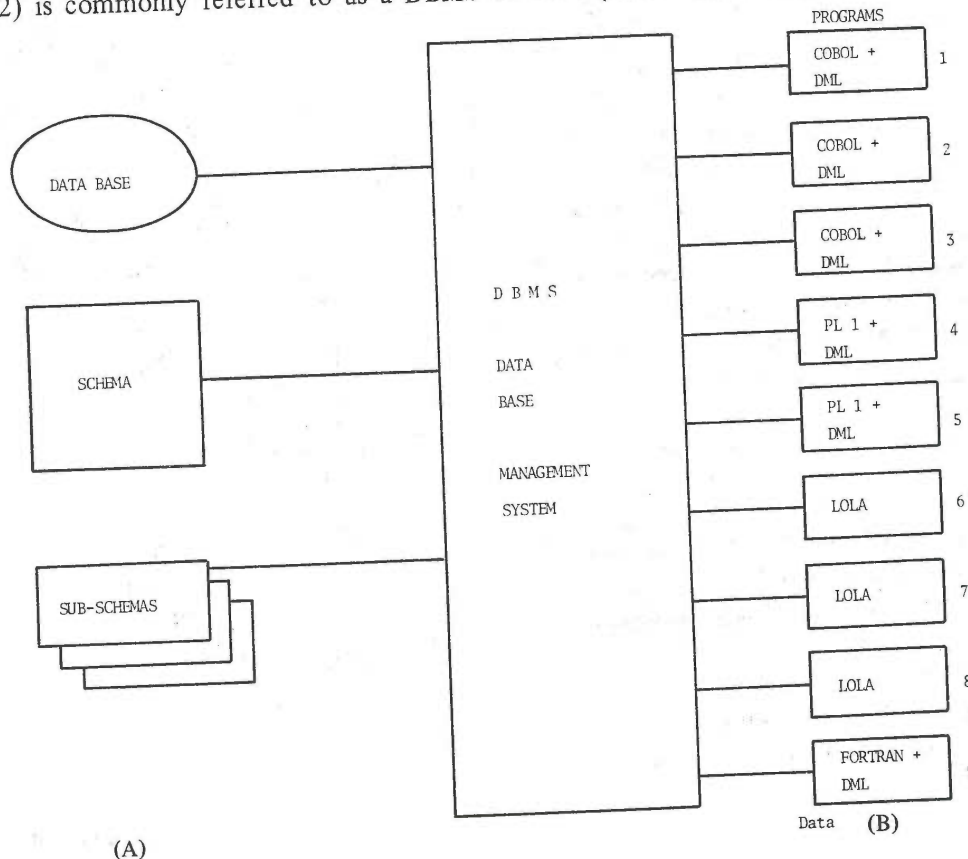


Fig. 2. (A) = Data base administration
(B) = Data base users.

common architecture for data base management systems" [2]. This quotation refers to the DBMS proposals of the independent CODASYL organization.

We illustrate this proposal by Fig. 2. The model of the reality is described in a language called the Data Description Language (DDL). The description of the entire model is called SCHEMA. Not every user wants (nor is authorized) to see the complete model, and often users want to apply a slightly different view (or submodel).

This user view is described in a so-called SUB-SCHEMA data description language, and the description is called a subschema. There may be any number of subschemas associated with a database. Both schema and subschema are (in the present commonly available database management systems) written by a computer specialist. Many different users may communicate with the database, each in his own language, through the intermediary service module, which is often referred to as the BDMS. CODASYL has issued specifications for languages to write the schemas and subschemas as well as the database extensions for COBOL (DML = Data Manipulation Language). A more detailed description can be found in [3], [4], [5].

5. Present possibilities of database technology

At present, there are available several database management systems for computers of the major mainframe manufacturers. Although these systems vary greatly in capabilities, a few of the following capabilities are available in nearly all systems.

5.1 *Data structuring or modelling.* The majority of presently available systems offer possibilities to describe fairly complex models, such as hierarchy and network data structures. However, the modelling is (at least in the seven mostly used systems) quite dependent on the physical aspects of the database, in other words application model building and physical implementation are not fully independent.

5.2 *Multi-language interface.* The majority of the presently available systems offer a multi-language interface. However, the languages are mainly for programmers such as COBOL, PL1, FORTRAN and assembler level languages.

Some systems provide a language which is aimed at the non-computer specialists, but many of these are still too "computer-oriented" and some of them require too detailed knowledge of the physical database characteristics.

5.3 *Flexibility (data independence).* User programs may be dependent on many aspects of the database

description or schema. A typical database is changing forever. This means that the schema is changing forever. Present systems offer a degree of data independence which is much higher than available before the advent of present database management systems. This means that reprogramming efforts because of changes in the schema, will decrease, where technically feasible.

5.4 *Privacy and integrity.* Most database management systems offer some degree of privacy and integrity, which enable the responsible people to maintain privacy and integrity which are commensurate with today's standards.

5.5 *Hardware independence.* Some of the present database management systems run on more than one vendor's computer system. A real step forward in hardware independence (like the step introduced by COBOL) will be possible if any industry standard for database languages is accepted.

6. Future possibilities of database technology

In [7], Whitney makes the following statement:

A fourth generation data management system will continue the development of all these trends towards generality, flexibility, and modularity. Other improvements will result from theories and concepts now being tried in experimental systems. Much greater degrees of data independence (from user program changes, from data description and storage changes, from new relationships among the data) will be common; user languages will become much less procedural; and data manipulation facilities for use in writing application programs will become much simpler and more powerful. Concepts from set theory and relation theory will become more widely used as the advantages of a sound theoretical basis for information systems become more widely appreciated. Increasingly, information management systems will make more of the optimization decisions relating to file organization and compromises between different user requirements. The trend to bending the computer towards user requirements rather than bending the user to the requirements of the computer will continue resulting in progressively easier to use systems.

Although this quotation gives a good insight in the next generation of database management systems, we want to add the point that not only set theory and relation theory will be used in the next generation of database management systems but also formal logic and linguistics. One result of the use of set theory, relation

theory, formal logic (these three are often referred to as modern mathematics) and linguistics will be a spectrum of languages, the great majority aimed at non-computer specialists, which at one extreme of the spectrum will consist of languages which permit a database user to make use of his native tongue. The database may contain data of a part of the reality which may be extremely complicated and one may describe this complicated model without any physical clutter.

This will very probably result in a much broader use of computerized databases than would be the case by straight extrapolation of past computer usage patterns.

Requirements for user languages which is a big step forward from languages for computer specialists, but still far away from the extreme mentioned above, are issued by GUIDE [8].

CODASYL has recently set up a new working group for end user languages, called End User Facility Task Group, EUFTG.

Other research and development efforts indeed indicate that the next generation of database management systems will, (for the first time in the computer history) offer the whole community a new versatile tool and not only a small group of specialists.

“The movement towards the database is in its infancy” [2].

“Computer data bases: the future is now” [9].

“Data base technology is complex and is far from fully understood”, [10].

“The concept (database) is real, viable and beneficial” [9].

The database can leave its infancy, right now, and be very beneficial, but only if education in database technology will make a big step forward.

With the majority of today's installed databases, running with today's database management systems, major misuses are almost always to be attributed to lack of the right education of database technology of the computer specialist. Many features of today's database management systems are not fully used and this can only partly be attributed to the unnecessary complexity of present systems.

In order that the next generation of database management systems, with all its promising possibilities, may be successful, it is necessary that the education problem of database specialists will be solved.

References

1. Nijssen, G. M.: Common Database: Key to successful information Systems. *In Proceedings of the First World Conference on Informatics in Government*, held October 16–20, 1972, in Florence, Italy, pp. 524–29.
2. Canning, R. G.: (Introduction to) 1973 ACM Turing Award Lecture. *Communications of the ACM*, vol. 16, No. 11, p. 653, November 1973.
3. CODASYL: Data Base Task Group April 1971 Report. Available from ACM, BCS or IFIP IAG.
4. CODASYL: Data Description Language Journal of Development. June 1973 report. Available from ACM, BCS or IFIP IAG.
5. Nijssen, G. M.: Common data base languages. *In Database*. A quarterly newsletter of the ACM SIGBDP Winter 1971, No. 4, pp. 7–12, Winter 1972.
6. Bachman, C. W.: The programmer as navigator (1973 ACM Turing Award Lecture). *Communications of the ACM*, vol. 16, No. 11, p. 653–58, November 1973.
7. Whitney, V. K. M.: Fourth generation data management systems. *Proceedings ACM National Computer Conference*, pp. 239–44, 1973.
8. Guide: Requirements for a user Language. Prepared by the members of the Data Base User Language Project of Guide International, 1973.
9. Nolan, R. L.: Computer data bases: the future is now. *Harvard Business Review*, pp. 98–114, September–October 1973.
10. Canning, R. G.: The cautious path to a database. *EDP Analyzer*, June 1973.