A SURVEY OF THE USE OF DATABASE MANAGEMENT SYSTEMS IN ACCELERATOR PROJECTS

J. Poole, P. M. Strubin
CERN, Geneva, Switzerland

Abstract

The International Accelerator Database Group (IADBG) was set up in 1994 to bring together the people who are working with databases in accelerator laboratories so that they can exchange information and experience. The group now has members from more than 20 institutes from all around the world, representing nearly double this number of projects. This paper is based on the information gathered by the IADBG and describes why commercial DataBase Management Systems (DBMS) are being used in accelerator projects and what they are being used for. Initially introduced to handle equipment builders' data, commercial DBMS are now being used in almost all areas of accelerators from on-line control to personnel data. A variety of commercial systems are being used in conjunction with a diverse selection of application software for data maintenance/manipulation and controls. This paper reviews the database activities known to IADBG.

I. Evolution

The use of commercial DBMS in accelerator projects started in the early 1980s, at which time it was already suggested[1] that the data for the LEP project should be maintained in Oracle tables. In the early days, DBMS were used to handle the data required for equipment specification and this was soon extended to include measurements made on components after delivery. The administrative sector followed shortly afterwards and started to use them for personnel and purchasing information and soon their use had become widespread throughout laboratories.

The inter-relationship of the data in various sectors has been recognised as a significant feature and one which requires careful attention to avoid duplication of effort and data. It is clear that making copies of data goes against the fundamental principles of database design but until the performance of the DB kernel and engines attained suitable levels, control system users had little choice but to feed flat files for use in the control system from off-line databases.

In the early part of the 1990s huge improvements in performance and reliability were made and on-line commercial DBMS were used for the first time. Since then, there have been developments to handle very large volumes of data and more recently object databases have been implemented.

Today, most accelerator projects are using some sort of commercial DBMS but the technology (which has only been around for a little over a decade) is still developing rapidly and interest in the field indicates the volume of data (which has probably reached the tera-byte level) will rise dramatically in the next few years. Interest in database technologies and the lack of a forum for the exchange of ideas sparked the formation of the IADBG[2] in 1994. The group, which now has members from more than 20 institutes and continues to expand, will hold its first workshop following this conference.

II. Commercial DBMS being Used

In the relational field, Oracle and Sybase are used more that other systems but Informix and Ingres are also represented in the IADBG (see Table 1). CEBAF is the only facility known to be using object databases and they are running ObjectStore.

A number of accelerator laboratories are still using in-house systems but are very interested in developing new facilities based on commercial systems.

III. Information Analysis and Database Design

It is the opinion of many IADBG members that analysis and design should be based on formal methods and supported by CASE tools. A good database design will lead to robust application software which can be developed rapidly, but on the other hand a simple error in design can easily cause a DBMS to stop.

It has been noted[3] that a detailed analysis of the information structures is essential before an effective database design can be established. The procedure requires formal methods like Structured Analysis and Structured Design (SASD) [4]. Powerful data modeling techniques are incorporated in SASD but they target the data involved in the software processes and therefore do not model project-wide data. For information structures like the equipment or personnel data, other techniques, such as the Nijssen Information Analysis Method (NIAM) [5] or Entity Relationship Modeling, are more appropriate.

IV. Functionality around the Database

There is perhaps more effort devoted to developing applications around the databases than there is actually creating and maintaining the data in them. The data maintenance is usually done using proprietary applications built from tools provided by the DBMS vendor. Such applications are typically graphical interfaces to the tables which contain the data. This is not the case, however, for ObjectStore where the only means of interaction with the data is through C++ programs.

Recent advances in database technology include referential integrity - this means that the DBMS itself will control data validity. This is achieved through the use of databases triggers and the ability to define constraints within the database itself. Rules and checks have only to be defined once now and it is no longer necessary to include them in every application.

Apart from graphical applications and command line interfaces, there are many applications which use embedded SQL to interact with relational data. These types of interface open the door to many automatic procedures and the possibility of driving control applications from the database. Examples of such applications are the automatic production of layout drawings and management of connections between equipment modules.

Data analysis is an important function in the world of database activities. DBMS vendors currently offer a number of interfaces to commercial data analysis packages like EXCEL. The advantage of this kind of interface is that the analyst does not have to have any knowledge of the database, just his own analysis application. Linked to this area is the problem of data visualisation; as the volume of data that can be handled conveniently has risen it has become increasingly difficult to display the information graphically in a meaningful way. There are a number of packages on the market and accelerator laboratories have often developed their own because the commercial world has different requirements.

V. Types of Data

In this section a brief description is given of the various data to which database technologies have been applied. A more complete analysis will be given in the proceedings of the IADBG Workshop[6] which follows this conference.

Personnel: databases containing employment and career details and personal information about the employees. A large cross-section of people need access to employees addresses and phone numbers and it is often necessary to have this information available within a control system to enable experts to be contacted when equipment fails. The databases are also being used to record intervention times for people who have been called in to repair equipment, so that overtime payments can be made automatically.
Financial and Procurement: details of the purchasing, budget controls and internal orders are of concern to the finance divisions and the equipment groups concerned. The finance database will require links to personnel data and possibly equipment specifications.
Planning: throughout the whole life cycle of an accelerator it is necessary to plan the installation, maintenance and de-commissioning of equipment. Planning these activities may require information from equipment databases and accelerator design. Relational databases have been used to store activity descriptions, durations and dependencies since the early 1980's.
Equipment Specification: this area was one of the first to be explored and is still one of intense development. It is perhaps no longer the first to be developed in a new project but is essential at all stages of the accelerator's life. It will provide the specifications to the suppliers in the early part of the project and the basic parameters concerning the behaviour of the equipment to the control system until the end of the project.
The data describes the physical properties of equipment and the required performance (for example, in the case of a magnet, steel specification, lamination thicknesses, type, size, weight, orientation, field strengths and so on).

Equipment Performance: generally concerned with calibrations (current to field, volts to frequency, etc.).
Inventory: concerns the equipment and financial data; maintenance of these data may be performed by someone outside the equipment groups. The data attributes specific items to functions, locations, personnel etc. and is tightly bound to maintenance data.
Maintenance: closely related to the equipment data but concerns the planning and history of maintenance operations.
Off-line Control: much of the data required for the control of an accelerator can be derived from the equipment information (equipment names, types, calibrations, magnet connections into families etc.). Certain fairly static data like lattice parameters, network data - addresses, topology, etc. may be prepared off-line for loading into the control system. It is appropriate to maintain these data in this way because they remain available for general use. Much of the off-line data is replicated in the control system where it can be used either from another database or within other data management systems (files, shared memory etc.).
On-line Control: concerns all of the settings, the machine configuration, calibrations etc. The term on-line implies that these are updated to reflect changes in the accelerator's state, or indeed to change the accelerator's state.
The application software of the control system requires certain data which are artifacts of the software mechanisms. An example of such data would be 'reserved' flags or information which controls the options within algorithmic procedures.
Logging and Measurements: every accelerator has some form of archiving system for machine settings and measurements from the instrumentation. Recent developments have allowed very large databases, which can hold a year or more of history recorded at frequencies ~1Hz, to be built. The distinction between logging and measurements is the frequency at which data is updated - the measurement database will reflect the state of the machine at a given moment and therefore requires rapid updates. These data are essentially date/time stamped values for the various accelerator parameters.
Alarms: alarm messages are generated in the control system but their treatment requires information about the equipment, personnel and logic to enable consequential analysis to be done. The alarm databases therefore contain hierarchical dependencies, alarm history and information describing actions to be taken for given alarms.
Accelerator Design: it is necessary to describe several versions of a machine so that studies can be made - mechanical design, modelling of beam behaviour, installation planning etc. The parameters in the database describe component structure (components of components), connections between components and their location. Accelerator design information is tightly bound to the modelling programs used for nearly all accelerator studies and the database is used to keep definitions of various versions of the accelerator and to generate input to the model program.
Design Drawings: management of the numerous drawings required for the construction of an accelerator. It is tightly bound to the accelerator design, which can be used to automatically produce drawings of new versions of the machine. In addition to the conventional catalogue-type data, information relating real pieces of equipment to drawings (drawing number and/or image) can be used.
Survey and Alignment: keeps track of the required and actual position of the accelerator components.
Documents: relational databases are well suited to documentation catalogues because they offer good facilities for searches, they are therefore often used at accelerator laboratories.

VI. Survey of DBMS Usage

Table 1 shows the information collected by the IADBG about the use of databases by the members. Although the table is certainly not complete, it can be seen that the variety of uses is quite large.

Configuration management is a function which encompasses several of the data types described above. It is a relatively new way of looking at the problem and involves describing how components are put together and provides the links to all the other information about the equipment. In a sense, this function is the heart of an accelerator project because it is in this area where one builds the definitive description of the way in which the components are connected.

The ever increasing amount of data requires data management tools which are more global in nature than classical DBMS. Such tools, which are now commercially available and known as Engineering Data Management Systems (EDMS), are based on well known DBMS like Oracle and Sybase. Industry is beginning to use these systems and the aim is to create links between data which would traditionally be stored in several different databases. The applicability of these techniques to the next generation of accelerators is currently under investigation.

VII. Conclusions

The use of DBMS has grown rapidly in the last decade but expertise is still somewhat limited. The value of data as a resource in an accelerator lab is becoming more appreciated and this is leading to more effort being devoted to its

Project	Equipt. Spec	Equipt. Perf	Inventory	Maintenance	Purchasing	On-line control	Off-line control	Logging	Alarms	Acc. Design	Design Drawings	Alignment	Documents	Personnel	Type Dat.Mangmnt
NSLS/BNL	X						X			X					Informix
RHIC/BNL										X					Sybase
CEBAF/CEBAF						X	X			X					ObjectStore
PS/CERN	X					X	X		X						Oracle
SPS/LEP/CERN	X	X	X	X		X	X	X	X	X	X				Oracle
LHC_Study/CERN		X						X				X			Oracle
VACUUM/CERN	X	X	X	X	X	X		X			X				Oracle
Admin/CERN					X							X	X	X	Oracle
SRS/DRAL						X									Private
ESRF/ESRF								X							Oracle
TEVATRON/FNAL						X	X	X	X						Sybase
MTF/FNAL		X								X			X	X	Sybase
GANIL/GANIL	X		X			X		X	X						Ingres
KEKB/KEK															Sybase
ALS/LBL							X								Dbase
PEPII/SLAC	X	X		X						X	X	X	X	X	Oracle
SLC/SLAC	X	X													Oracle
SPring-8/SPring-8	X	X	X			X	X		X						Sybase
SRRC/TAIWAN	X					X	X	X	X						Private
ELETTRA/TRIESTE	X	X				X	X								Oracle

Table 1: IADBG Survey of Database Usage in Accelerators

management. The problems encountered in an accelerator environment are quite different to those commonly found in the commercial world and therefore commercial solutions are not often available. The formation of the IADBG will hopefully go some way towards helping people to learn from the experience of others and lead to more effective developments in the future.

Experience with DBMS during the last few years has clearly shown their potential to contribute to greater efficiency throughout an organisation and to better accelerator performance.

References

[1] Databases in the LEP Control System, M.C. Crowley-Milling, CERN LEP Note 375, May 1982.

[2] International Accelerator DataBase Group (IADBG), an informal association of interested parties, open to people working in the field. For more information see the WWW pages (https://www.cern.ch/IADBG/Welcome.html) or E-mail the organiser, John.Poole@CERN.CH

[3] J. Poole, 'Databases for Accelerator Control - An Operations Viewpoint', Proceedings of the 1995 Particle Acelerator Conference, Dallas, May 1995.

[4] T. DeMarco,'Structured Analysis and System Specification', Yourdon Press, New York, 1982.

[5] G. Verheijen and J. Van Bekkum, 'NIAM: An Information Analysis Method', 1982.

[6] Proceedings of the IADBG Workshop on Databases to be published on WWW through the IADBG pages (see previous reference).