UNIT 14 DBMS IMPLEMENTATION AND FUTURE TRENDS

Obejctives

After going through this unit, you should be able to :

- identify the factors causing resistance to the induction of new DBMS tools
- enumerate the advantages of a distributed processing environment
- identify the factors (both technical and administrative) which motivate the move to acquire a DBMS.

Structure

14.1 INTRODUCTION

Having covered the conceptual and theoretical framework for DBMS, this unit focuses attention on the practical aspects of the implementation. Computers in general, and data base systems in particular, have the usual feature that a lot is expected from them. It is often the expectation of the management that when these two are in place, all relevant information would be available "at the press of the button."

It must, however, note that the establishment of a data base management system in an organisation is part of a process of implementaion and radical change in the organisation. When computerisation is attempted to be introduced in an organisation, where hitherto manual systems have been availed, although there is resistance, it can be overcome on the grounds of efficiency and increased productivity. In the case of introduction of data base management system, the resistance comes from the erstwhile computing department itself, because it tends to diminish its role as the controller of greater information.

The recent developements on hardware and software have made distributed and departmental computing a reality, so that the whole EDP department would have to delegate many of its functions from which it used to derive a lot of importance or power. This unit, therefore, deals with issues which managers would have to cope with in successfully establishing a DBMS system.

Having decided to go in for a DBMS, the selection and acquistion still cause major problems. As some say, today, it is in many ways easier to cope with the technical problems associated with establishing a DBMS system. Some of these are covered in section 14.5 of this unit. But, more important and probably the most

difficult to cope with is the problem of "human awareness". Since the DBMS approach itself is very new, the kind of management issues arising in relation to personnel involving in the establishment of management of DBMS is still not clear because of lack of sufficient experience and technical studies. In this unit, many of these issues and points are identified and raised, although it is not possible to give any unequivocal answers at the present moment.

14.2 ORGANISATIONAL RESISTANCE TO DBMS TOOLS

There are several points of resistance to new DBMS tools :

* Resistance to acquiring a new tool.
* Resistance to choosing to use a new tool.
* Resistance to learning how to use a new tool.
* Resistance to using a new tool.

The selection and acquisition of a DBMS and related tools is one of the most important computer-related decisions made in an organisation. It is also one of the most difficult. There are many system from which to choose and it is very difficult to obtain the necessary informaion to make a good decision. Vendors always have great things to say, convincing arguments for their systems, and often many satisfied customers. Published literature and software listing services are too cursory to provide sufficient information on which to base a decision. The more difficulty in gathering information and making the selection is one point of resistance to acquiring the new DBMS tools.

The initial cost may also be a barrier to acquisition. However, the subsequent investment in training poeple, developing applications, and entering and maintaining data will be many times more. Selection of an inadequate system can greatly increase these subsequent costs to the point where the initial acquisition cost becomes irrelevant.

In spite of the apparent resistance to acquisition, the projections by the crystal-gaters in the industry are forecasting a multi-billion dollar industry by the year 2000. Even though an organisation may acquire a DBMS, there are still several additional points of resistance to overcome.

Simply having a DBMS does not mean that it will be used. Several factors may contribute to the lack of use of new DBMS tools.

* Lack of familiarity with the tools and what it can do.
* System developers used to writing COBOL (or other language) programs prefer to build systems using the tools they already know.
* The pressure to get new application development projects completed dictates using established tools and techniques.
* Systems development personnel have not been thoroughly trained in the use of the new tool.
* The organisation has not set up a program to train users of new DBMS tools.
* Users are reluctant to use a new tool because there is no one in the organisation to provide advice in its us and to help when problems arise.
* Tool is only known to a few specialists in the data processing organisation.

* No one in the organisation is actively encouraging, even compelling the use of new DBMS tools.
* DP management is afraid of run away demand on the computing facilities if they allow users to directly access the data on the host computer using an easy to use, high-level retrieval facility.
* Organisational policies which do not demand appropriate justification for the tools chosen (or not chosen) for each system development project.

14.3 DATABASE IN A DISTRIBUTED PROCESSING ENVIRONMENT

Computer at multiple sites linked together through a communications facility provides the basis for distributed processing environment. Distributed processing is driven by two main technological factors :

* Lower hardware costs for computer processing and data storage.
* Moderately higher communications costs.

and three main organisational factors :

* More responsive to local user needs by offering faster access and greater autonomy or control.
* Greater reliability through reduced operational dependence on a single, central site for data processing and data storage.
* Increased sharing and coordination in the use of common resources.

Distributed processing is not a panacea for achieving these organisational goals and reducing costs. A distributed processing environment adds complexity to the data management porblem compared to a single site environment. In spite of the claims of some vendors of networks or distributed DBMS, substantial problems remain to achieve the full potential of distributed data processing.

A distributed processing environment is characterised by a network of multiple computing nodes connected with some communications facility. A node, like a single site computer, can perform some combination of the following functions : execute programs, respond to a user request, run a DBMS, or store data.

A communications facility is the collection of processes and physical facilities which interconnect the nodes. It includes knowledge of the physical location of each node, the physical connections or paths between the nodes, the protocols for sending data from it to one or more other nodes. A network access process (NAP) exists at every node and is the interface between processes at the node and the communications facility. The NAP is usually some combination of software and hardware. The hardware may consist of a standard input-output port on the computer or a special plug-in board to which a network cable is connected. The actual configuration or topology of the network and the communication protocol is secondary - the important point is that any node can communicate with any other node in the network.

14.4 EMERGING STANDARDS

Mention of standards often evokes strong feelings. For some it is what they feel should be the direction and content of a standard, for others it is despair at what is or is not happening in the standards arena. Nevertheless, standards are important to both users and developers of DBMS products. This section reviews the what and why of standards, the evolution and current status of activities in the development of database-related standards, and the direction, likely outcomes and consequences of these current activities.

Advantages and Inhibitors to the Development of Standards

The main purpose of standards is to foster interchangeability of products by defining standard interfaces and to foster the compatibility or coexistence of products by defining standard interchange formats. Interchangeability allows user to mix and match DBMS and related products. Transferability allows different products to be used together by enabling the transfer of data or other materials between them. Standards accomplish the following desirable goals :

* Guide manufacturers and vendors in the design and development of products. They encourage developers to build system with common interfaces.
* Open competition to more developers, particularly small ones, when they can build to a standard and when interchangeable parts are available from different vendors.
* Give conforming products an implicit stamp of approval
* Assist users in choosing from among competing products and result in fewer different products from which to choose.

  Increase portability of software and applications across people, machines, and operating system environments.

  Reduce the cost of training and increase the productivity of personnel who use the standard products.

All of the above assume that each standard is a "good" standard and is generally accepted. People, however, can disagree on what constitutes a "good" standard. The vendors of existing products will not want to see a standard go beyond or be different from the capabilities of their products. Users may feel that a standard does not go far enough to adequately meet their needs.

The major disadvantage of standards is that a premature standard can inhibit innovation in product development and incorporation of the latest technological advances. Other forces will also inhibit the adoption of standards. A vendor which dominates an industry or a particular type of product may delay or disrupt the standards development process, preferring to see their own position or product become a de facto standard.

The Development of Database - Related Standards

Standards typically follow industry developments, and are therefore strongly influenced by vested interests. A standardization effort begins with some concrete proposal which may be in the form of an operational system or language, or a detailed specification for a system or a language. Furthermore, any proposal should have reasonably wide industry willingness to at least entertain cooperative action toward developing a standard. Wide industry willingness may stem from cooperative industry development of a proposal or from widespread use of an existing system or language.

In 1970, Ted Codd, working for IBM, published a paper introducing the relational data model. That paper generated considerable interest, as evidenced in many subsequent papers of primarily theoretical and academic interest, and spawned several experimental system development projects.

From the perspective of the data structure the differences are few though not insignificant. The major difference being that the earlier network DDL permits nested repeating groups within a record whereas the relational data structure does not. Many people, including those favouring the earlier network approach, argue that good database design means flat files in third normal form (at least initially --including nested repeating groups of data items in records for implementation efficiency).

From the perspective of data manipulation the high-level language of the relational approach is clearly superior to the record-at-a-time-low-level languages.

Direction of Emerging Standards

The first database-related standard to come from ANSI was the Network Database Language. The major features of the proposed standard were :

1. Concepts, introducing the pieces, objects, and common elements of the language which is described in the rest of the document.
2. Schema Definition Language for declaring a network structured database and its integrity constratints.
3. Subschema Definition Language for declaring the user view of a database schema.
4. Data manipulation Language and module language for declaring the executable statements and database procedures for a specific database application.

Implementation of this standard can exist in an environment with :

* Application programming languages.
* End-user and natural language query facilities.
* Report generator and graphics output facilities.
* Data storage definition and device media control languages.
* Access control mechanisms.
* Database copy and reversion ("unload") facilities.
* Schema access facilities for ad hoc users and user-written application procedures.
* Schema manipulation language for revising a database definition.
* Database restructure and reorganization utilities to bring an existing database into conformance with a revised schema.
* Information resource data dictionaries.
* Tools for database design, database administration, and performance monitoring.
* Distributed database facilities in a Network DBMS.
* Interfaces to word processing, spread sheet, decision modelling, and statistical packages.

Development of a Reference Model for Database standards

The development of related but uncoordinated standard is cause for concern. There are several different areas within DBMS technology which are or could be the object of standards development.

* Data structure definition languages which are used to define database schemas and user schemas. Such languages are used in a DBMS, in programming languages, and in Data Dictionaries.
* Integrity control languages to define semantic constraints, validation criteria, access control, etc.,

* Physical storage structures, access methods, and device/media control.

* Data manipulation languages which can appear in a host-language interface to a conventional programming language, a self-contained DBMS language, an ad hoc query language, or a report generation facility. Retrieval and manipulation languages are needed to operate on stored databases but also stored schemas, userschemas, and data dictionaries.

* Data mapping and conversion languages for use in database reorganization, restructure, creation, reversion, and schema-user schema translations.

The development and coordination of standards across all these areas is a massive undertaking. It can only be done in stages; some priorities must be set. A comprehensive data structure definition language seems to be the most important first step, since a given data structure is the object of retrieval, manipulation, and conversion. To bring some order to standards development, there have been attempts to define a reference model which includes the major modules of a database system. A reference model establishes a common framework or background for a class of systems such as DBMS to which people may refer when talking about specific products, product design and development, or standards. It serves as a tool for the development and coordination of standards by identifying major interfaces between component parts of the overall system.

A database system has four major types of interfaces :

* End-user who is concerned with getting some job done within the organization. End users interface with the various modules or functions of a database system to access database definitions and to retrieve and manipulate stored data, both in writing procedures for deferred execution and in ad hoc, interactive use.

* Administrative support staff concerned with the design, development, installation, maintenance, and evaluation of database system products, and the training of end users. The database administrator needs to create and revise database definitions, to establish integrity controls, and to monitor performance.

* External interfaces to other computer system components (non-DBMS) such as the operating system and the network access process.

* Internal interfaces between functional parts or modules of the database system.

14.5 DBMS SELECTION AND ACQUISITION

Movement towards the database approach to data processing is not every organisation now. Discerning the right time requires an assessment of organisational need, capacity, and readiness.

Several factors may motivate the move the acquire a DBMS

* The need for more comprehensive data storage and retrieval capabilities.
* Faster response to ad hoc queries.
* Faster application development.
* Reduced data redundancy
* Transferability across hardware
* Reduced program maintenance
* Increased ability to respond to changing requirements

* More consistent data values
* Increased security
* Better audit facilities.
* More realiable backup and recovery
* Reduced operating costs.

Whatever the reasons, they should be clearly documented. This serves to set priorities on the objectives and provides the basis for making trade offs among conflicting objectives and in the selection of various DBMS features.

Selecction criteria are broken down into technical criteria and administrative criteria. Technical selection criteria relate to the functional capabilities. of the DBMS product. Administrative selection criteria include efficiency, ease of use, documentation, acquisition and operating costs, compatibility with the existing organisation and data processing facilities, and characteristics of the vendor.

The most important technical criteria concern the basic type of system required and the set of functional capabilities needed to serve the user community.

Recognising that any statement of selection criteria must be tailored to the needs of the organisation, certain technical criteria are often overlooked. This book gives a comprehensive picture of the functions of database management.

The major categories of technical criteria for selecting a DBMS are as follows :

MAJOR TECHNICAL CRITERIA

1. Database definition, including logical structure and physical stoage structure
2. Generalized retrieval capability - asking questions and getting answers
3. Generalized Updata capability - to add, delete, and modify the data in the database
4. Programmer interface.
5. Maintenance of Database Integrity
6. Modes of operation-online and/or batch, for above functions
7. Database revision and evolvability

The outline of features in these categories provides a starting point for an organisation to begin an evaluation and selection process.

Selection criteria not relating to functional capabilities are considered administrative criteria. These include the required configuration of hardware and software, operational and performance characteristics, vendor support, documentation, required staff support, and cost. While some technical capabilities (or lack of) have administrative implications, it is desirable not to confuse the technical criteria for evaluation with administrative considerations.

Aside from the sytem itself, the next most important, consideration is the vendor (or developer). Are the people genuinely excited about their product ? That is the most important question to ask. Visit their location, mingle with the people in sales and in development, and look into their eyes as they speak of working for this vendor. Are they proud to be on the team ? Are they enthusiastic about the future of this vendor and this product ? Whether

or not it's possible to get that close, a prospective buyer should try to get a feel of these questions about the vendor.

The following lists direct attention to various areas which may be important in evaluating the administrative aspects of a candidate system.

MAJOR ADMINISTRATIVE CRITERIA

* Vendor Characteristics and Product Stability-history, size, financial strength, years since product introduced record of enhancements, and number of users.

* Maintenance support - written agreement and are other users satisfied, responsiveness to user problems.

* Documentation and Training - readable, updated with changes, training classes and materials available at reasonable cost.

* Ease of Learning and Use-understandable system architecture, user-freindly interface, and online help facilities.

* Operating and Performance Characteristics - it actually works, and will handle the sized and throughputs required.

* Supporting Environment : Hardware, Software, and Administrative Staff-the system will "fit" into the organisation and the existing (or upgraded) hardware and software environment.

* Costs - direct acquisition or lease plus changes to existing hardware configuration, installation, training and documentation.

14.6 SUMMARY

We have pointed out in this unit that there are resistance to new DBMS tools. In spite of the apparent resistance, many more organisations are moving in towards the use of DBMS. It is also quite clear that organisations will have to move towards DBMS, especially those based on relational approach in order to maintain their competitive position in the emerging market places.

The availability of many more products as well as the appearance of many more features in these products causes a greater responsibility on the managers responsible for the implementation to adopt an approach which would lead to successfully meeting the information needs of the organisation.

It has been emphasised in this unit that apart from identifying the technically right approach and to provide the system on which it could run, other administrative factors must also be borne in mind, so that the new systems actually fit well into the organisation. This unit also briefly refers to the trends in the products and the emerging standards because while putting up a DBMS system, it must be realised that it will have to cope with the information needs of the organisation not only as they stand today, but also for sometime into the future. It has been usually observed that once computerisation is successfully maintained in any organisation, the needs of that organisation for data management grow exponentially in volume. It is, therefore, of paramount importance that the management takes a vision which is somewhat larger than just the needs of the moment, keeping in mind the future trends and emerging standards.

14.7 SELF ASSESSMENT EXERCISE

1. Identify the most important factors inhibiting an organization's move towards a DBMS.
2. Outline the basic sequence of steps to acquiring a DBMS. What steps are the most important ?
3. Describe the problems which may arise if the DBMS evaluation team consists entirely of persons from Systems and Programming in the Data Processing Department.
4. Identify and briefly describe some of the terms which should normally be included in a software acquisition contract.
5. Why should an organization be careful about placing over reliance on "benchmark" tests in selecting a DBMS ?
6. What are the advantages and shortcomings of obtaining information about a candidate system from each of these sources : vendors, current users, demonstration, introductory course, trial use, benchmark, listing services ?

14.8 FURTHER READINGS

1. Atre S. *Database Structural Techniques for Design, Performance & Management,* John Wiley & Sons, 1980
2. Date C.J. *An Introduction to Database Systems,* Addison-Wesley, 1981
3. Kroenke D.M. *Database Processing : Fundamentals, Design, Implementation* 2nd Edition, SRA, 1983
4. Everest, G.C., *Database Management : Objectives System functions & Administration,* McGraw Hill, 1986.
5. Ven Halle Fleming, *Handbook of Relational Database Design,* Wesley, 1990.