

STRATEGIES FOR CONVERTING TO A
DBMS ENVIRONMENT

Dave M. Durbin
Data Base Administrator
Ames Research Center

Ames Research Center(ARC) installed ADABAS on the administrative computer, an IBM4341 running VM/CMS. Prior to the installation of ADABAS, the only data base-like system used at Ames was a product called QBE (Query-by-Example). This product was installed and an extract of the payroll and personnel file was loaded as a table under QBE for use by the Personnel Office and Resources Management Office for ad hoc queries.

The strategy of converting to Data Base Management Systems processing techniques consists of three different strategies - one for each of the major stages in the development process. Each strategy was chosen for its approach in bringing about a smooth evolutionary type transition from one mode of operation to the next.

The initial strategy of what we call the "indoctrination stage" consisted of the following elements:

- * Provide maximum access to current administrative data as soon as possible.
- * Select and develop small prototype systems.
- * Establish a user information center as a central focal point for user training and assistance.
- * Develop a training program for programmers, management and ad hoc users in DBMS application and utilization.

Each of these strategies has one or more "sub-strategies" if you will. Each of these serves to highlight the purpose and intent of the main strategy.

By providing maximum access to current administrative data as quickly as possible, we hoped to achieve the following advantages:

1. Casually introduce on-line computing to the non-ADP professional with little impact on the current systems.
2. Provide training to the user community using the data they are familiar with and, thus, flattening the learning curve.
3. Decrease requests for AD HOC reports from the user as they gained proficiency in performing their own requests - resulting in more time available to the DP staff in developing ADABAS systems.
4. Providing a basis for determining efficiencies in file design based upon what type of queries the users performed.
5. Satisfy the users desire to have more access to their data.

The trade-offs were:

1. Possible system degradation due to less efficient file design, users performing inefficient queries, etc..
2. Too much freedom given to the users which may have to be taken away later.
3. Worst possible case - the user would be turned off immediately and develop a negative attitude toward on-line processing.
4. Enlarge the task of cleanup - to make data definitions conform to standards, etc..

To complement this approach, we developed a plan for establishing a user information center to serve as the central focal point for handling user training and assistance. The plan also called for developing training programs for DP personnel, management and ad hoc users.

Another part of the plan called for selecting and developing small "stand-alone" systems to provide needed computing capabilities where none existed.

Expected benefits of this particular strategy were:

1. Reduce the impact of non-experience by using less complicated systems.
2. Lessen the impact of system complexity in overcoming the learning curve.
3. Gain experience in developing on-line systems in DBMS environment.
4. Provide new service to user community and demonstrate the capabilities of the DBMS.
5. Development process could be used as initial guide to on-line systems development.

The last piece of the "indoctrination" strategy was to develop an effective training program for each of three groups of people.

1. The DP professional training consisted of specific training in the internals of the data base management system, courses in the procedural language and a seminar in applications development methodology.
2. The top and middle management training consisted mainly of a series of briefings and demonstrations.
3. The training for the Center ad hoc users was primarily an overview of the DBMS and specific instruction in the procedural language with emphasis on query. This training is provided on several levels from beginner through advanced. In addition, a selected group of users attended the seminar on PDM 80.

This strategy was rather straight forward and simple. There was little concern for efficient design, file layouts, normalization of entity classes, or any such buzzwords that you hear about today in the data base world. The emphasis was on experience and experimentation, learning what could be done with the data base and alternative ways of doing it. Without experience with a data base management system - and Ames had little of that - these "modern" concepts had little meaning and were difficult to relate to. These new ideas and concepts would become more important in the next stage of the development process, but the indoctrination stage was just that - indoctrinate the community to the new tool - the Data Base Management System. The fastest way to do that was to keep things simple. Before I go on to explain the components of the next stage of the development process, let me elaborate a little on the results of the initial stage.

The first step was to load the most used files from the current batch systems into the data base for access by the owners of the data for ad hoc reporting. The first files loaded were the personnel and payroll system file. The files were defined with little concern for standard data names and proper data base file design. The arrangement of the fields was based on a combination of the current record layout and a limited knowledge of what fields were most frequently accessed. The objective was to get the data into the data base. Some of the data was loaded and a few programs written to assist the personnel department in accessing the data. One such program was the individual employee record which basically displayed an individual employee's personnel status and action history on the screen. The other programs were a series of menu-driven which assisted the user in operating the system and developing queries without having to learn all the commands or syntax. The initial reaction was one of caution. The user was unsure of his capabilities to learn and effectively program queries with the new language NATURAL, which is considerably more complex than QBE which they had been using. However, it was only a matter of a month or two before they were using NATURAL exclusively. This was due to several reasons: Primarily because the file loaded into ADABAS was more complete than the one in QBE and the ADABAS file was reloaded every day that an update had taken place, whereas the QBE file was loaded only at the end of each pay period. Other factors were the quicker response in ADABAS and the personnel department had hired a summer student who was eager to learn NATURAL and had developed some rather useful reports using NATURAL which greatly assisted the organization. Therefore, after a few short months, QBE was quietly phased out and the lease agreement terminated. ADABAS and NATURAL were accepted by the users. In fact, the user community was becoming excited over this new source of information and wanted more - not only from the data processing department - but they wanted more knowledge about the data base and NATURAL. They wanted to do more for themselves.

The logical place for the users to go was to the user information center that was established when QBE was installed, but had seen little activity. The users came here for answers. No formal training was provided yet for

the users. The training was concentrated on the DP personnel. Until that training was completed, training of the users could not commence.

With the help of Langley Research Center, we loaded files from other systems with the Data Base and we obtained "on loan", a highly qualified instructor to provide advanced NATURAL training to our DP staff and to assist in developing a formal training program for the user community. Meanwhile, development of three small systems in ADABAS and programmed primarily in NATURAL was under way. The three projects were:

1. LIRS II - a legal information retrieval system for maintaining an inventory of legal documents relating to the 80x120 Foot Wind Tunnel litigation.
2. TPS - Treasury Payments System for processing of miscellaneous payments.
3. System Documentation Library for maintaining inventory and control of systems documentation.

The LIRS II System was developed in four weeks with an additional two weeks spent in enhancing the query portion of the system.

The Treasury Payments System was developed in eight weeks.

The System Documentation Library was developed over a longer period of time, approximately ten weeks, by a programmer who had little experience in NATURAL programming. This system was later scrapped and redone within a two week period as a class project in advanced NATURAL techniques.

These three systems were selected for development based on three factors:

1. There was an immediate need for the data processing application and the system did not exist before.
2. Expected development cycle was short and if the system had to be redone, it could easily be accomplished.
3. The systems were not complex.

These three systems are still in operation today and require minimal maintenance - in fact - a routine maintenance schedule has not yet been established.

These three projects led to the development of NATURAL programming standards and system design standards, but most of all, they served to show the capabilities of ADABAS and the productivity of the NATURAL programming language to the DP department and to the Center.

By the end of the first phase of development (and there is no definite cutoff point between the first and second phase), it was evident that the DBMS dramatically and permanently changed the concepts and attitudes toward computing at Ames.

The user information center conducts regular, formal training classes for the user community. The users are finding that they can meet some of their information needs without the assistance of the DP department. Communications between the DP department and the users is at an all-time high. The user has learned the importance of their involvement in system design and development.

There are some negative aspects to this strategy which I will list below; however, these shortcomings were recognized at the inception. Dealing with these issues would be a major part of the second phase in the evolutionary process.

1. Security issues. Security of the data against unauthorized access or destruction was dealt with on a very low key basis. Minimal security precautions were established primarily to protect privacy data against unauthorized access and to insure integrity of on-line systems which did not have a backup batch system. While these security procedures are not foolproof, reports of the data base activity have shown no violations and there have been no problems detected.
2. Data dictionary role in the DBMS. Very little attention was paid to the data dictionary and its role in managing the data base environment. Consequently, the first phase saw no standardization of data names and data definitions, data redundancy and data relationships were not defined.
3. Data base tuning and capacity planning. Obviously this was nearly impossible in the first phase of our strategy primarily due to lack of any knowledge in this area.

This issue became important sooner than anticipated. With the installation of the NEMS System, the data base performance ground to a snail's pace. Within a couple of weeks, however, through adjusting of buffer areas and some blocking factors of data base data sets, the performance was brought back to a tolerable level. Still, I was not satisfied and could foresee difficulties in the near future if performance could not be improved. The performance did not appear to be as efficient as some COMPLETE users were experiencing. To make a long story short, we finally came across the fact that CMS was single threading all of its I/O, and that Cornell University had developed BDAM emulation software which allowed overlapping of I/O. Within another couple of weeks, we had installed the software on the data base machine and were experiencing a ten-fold improvement in response times.

Overall, the objectives of the first phase of the strategy were well achieved. The users could now have access to their data in a realtime sense. They were well on their way to being able to ask the right questions and getting the right answers to those questions without waiting in line at the data processing department door.

The second phase is well under way. The main objective of the second phase is to develop an effective, efficient and viable methodology for developing, maintaining and utilizing the Center's information resources.

In this phase, the issues of security, data structures, file structures, and proper data definitions become important. To assist us and to get exposure to the latest methodologies in data base development, we hired UCLA professor, Dr. Alphonso Cardenas, to provide a seminar on his methodology "PDM 80" and to provide consulting services during the first development project using his methodology. That project started in January of this year and is scheduled for completion in August. It is planned that the results will be a system with a well documented, developmental methodology that will serve as a model or prototype for all future development efforts. The project encompasses the R&PM functions of the Resources Management Office.

Also scheduled in this phase, is the installation of NATURAL security and the development of required procedures to assure data integrity, protection of sensitive data from unauthorized disclosure, and protect the entire data resource from destruction.

Another important issue in the second phase is the development of a truly integrated data dictionary system. The objectives of the data dictionary are:

1. Document and define all the entities, attributes, classes and relationships that comprise the Center's information resource.
2. Standardize data names by establishing naming conventions.
3. Define systems, programs, files and data elements used in them.
4. Define the owners of the data.
5. Provide the capability for determining and subsequently managing data and system changes.

I might add that included in this data dictionary is all data considered to be part of the Center's information resource regardless of where it resides.

Finally when this is all accomplished, we are ready for the final phase in evolving toward DBMS technology and that is the incorporation of the data bases into a complete automated office management system. "Office automation" to use the buzz word of today.

Even today with the increasing use of the micro-computers as work stations with their own data base systems, word processors, and communications with the main-frame, office automation is a fact of life at Ames and is rapidly increasing.

Many of the organizations at Ames, especially in the administrative areas, have in one short year gone from computer-phobia to computer-literacy. The data processing department is evolving from that of the doer to the role of the consultant.

What's ahead for Ames? In the short term, SOFTWARE AG has offered to allow Ames to test their new ADABAS/VMS product for the VAX. Currently, Ames is loading financial data from the IBM Financial System onto one of the VAX machines and using INGRES to query the data.

The test period will be used to compare the two DBMS'. If ADABAS performs as well as INGRES, we would then purchase ADABAS/VMS which will allow us to transport the NATURAL programs from the IBM to the VAX.

Several menu-driven, ad hoc query facilities have been developed which provides the capability for the user to extract data from the data base files without seeing or using one word of NATURAL language. One of the systems allows the user to export data to his micro-computer for use by a micro-based data base system such as Lotus 1-2-3. These systems have been a real hit with the users and more of these are in the planning stage.

Conclusions:

Starting small is important.

Several small systems developed implemented will achieve immediate recognition and subsequently, strong support for the capabilities of the system and the DP department. Starting small will allow your DP professionals to gain experience, and at the same time, achieve some demonstrative successes.

Access is the key to success.

Provide the user with the access to his data and with a small selection of software products or tools for the user to manipulate the data. Use the information center to train and assist the users; to answer their questions; to develop their skills in writing their own programs.

User information center is vital to success.

At first, if the center is nothing more than that of a solution center, i.e., providing answers to users questions on how to write a query, it will still serve to nurture the non-DP professional in finding his own solutions. In fact, in some instances, you may even find the information center personnel benefiting from the approach of some users. It seems as

an exchange center - an exchange of ideas and techniques between the administrative professional and the data processing professional. Result: a more solid knowledge base for developing applications and an atmosphere of cooperation and common interests. It serves to gradually bring about a change from the old method of the users tasking the DP department with an application or problem, and the DP department providing a solution which often the user is not completely satisfied with, to a new attitude.

1. One of working together to achieve a common purpose.
2. One in which the user shares in the responsibility and credit for the success of a project.

Nearly all of the modern DBMS development methodologies today, stress the importance of a well defined, comprehensive integrated data dictionary. I strongly agree with that concept. It should be the map of the data base system, but don't waste effort on attempting to standardize data names of existing files being loaded into the DBMS. Load the file with the existing names, define them as well as reasonably possible, but remember to start - the key is access. Often using names that the user is familiar with is the best course in getting them through the indoctrination phase - and the advanced user is an excellent source for developing names and definitions of the data during the later phases of development.