Data base administration—Classical pattern, some experiences and trends

by JEAN-PAUL De BLAVIS
Centre d’Enseignement Supérieur des Affaires
Issy-les-Moulineaux, France

and

THOMAS H. JOHNSON
The Wharton School
University of Pennsylvania
Philadelphia, Pennsylvania

ABSTRACT

This three part paper covers the growth and classical patterns of data base administration along with a survey of some currently practicing data base administrators. As a result, some trends and evolutions of the data base administrator position are presented.

After a brief historical introduction, the paper first sets forth basic definitions and organizational considerations for the DBA function. Interfaces, both internal and external, are defined and the associated problems are discussed. The functions and responsibilities of the DBA are presented along with the tools available for carrying them out as recommended by various committees including CODASYL, GUIDE/SHARE and other reports.

Then, a survey of organizations utilizing data base administrators is presented. The survey looks at organizations of varying sizes and commitments to data base technology. The results of the survey try to show how the previous recommendations are reflected in reality. An attempt to rationalize the plans and the actual status of the DBA in an organization is carried out.

Finally, from the survey results and from projections provided by the data base administrators themselves, some trends in the evolution of their functions are outlined. In light of these projections an attempt is made to review some of the recommendations currently put forward.

INTRODUCTION

Does the data base administrator portrayed in the current literature exist in practice? What are the differences, and what are the trends for the DBA function? This paper attempts to answer these questions by briefly reviewing the literature, discussing actual practice, and outlining the trends in the data base administrator's role.

We developed a central thesis from pre-screening interviews and literature search to focus our research on the data base administrator. In testing this thesis, we uncovered areas for future research and application emphasis. After posing the central problem, we reviewed the definitions and functions of the DBA as defined in current literature to frame our discussion. Next, we focused on the current state of the DBA, summarizing interviews of over twenty practitioners. We present results in the form of evolutionary trends occurring in this area and conclude with opportunities for research suitable to the support of practitioners.

The central thesis is:

The data base administrator now, and in the future, is an individual who performs the function of planning, designing, operating and controlling the data base of an organization at both the policy and operational level.

In order to frame the results from our study, we must first establish the definition, function and organization proposed for data base administrators. We have synthesized proposals and reports from both committees studying the area and authors publishing in the area. Committees include GUIDE-SHARE, CODASYL, and ANSI-SPARC. They are very active in the DBA area, especially GUIDE-SHARE which has advocated recent expansions of the DBA role to data administrator and put forward techniques for focusing on data resource policy.

DEFINITIONS AND FUNCTIONS

Definitions of the data base administrator's function

The individual introduced in both Codasyl and Share Reports, called the data base administrator, is a person meant to solve many of the problems in file integration and in maintenance of any organizational data base. He is...
supposed to be familiar with computerized systems, with data management, and above all, with every aspect of the corporate data base. The DBA function has been previously defined as a human function, responsible for the coordination of all data related activities. The following is an extended definition specifying some of the generic areas of the DBA’s functional responsibility. The data base administrator is the individual providing the coordination, perspective, and administration of the data base by exercising specific responsibilities. His responsibilities should include the definition, organization, protection, efficiency, and documentation of the data base. He should also be responsible for defining the rules by which data is to be accessed and stored.

To decide where to place the DBA involves trade-offs almost always because of the distinct organizational characteristics in which every systems environment operates. The answer provided by the literature is generally that the DBA should report to the highest full-time information systems manager. Practically, however, there are very few DBA’s who are aligned this way, as shown by some field studies.

**Interfaces**

The data base administrator generally interfaces directly with three groups within the organization. The systems administration group is the first. This group is concerned with the operation, maintenance and performance of the information systems equipment. This includes the performance of data base management systems, as well as security, recovery and re-start of these systems. They are also concerned about the throughput of the system and the running of production systems in the most effective manner.

Another interface exists between the data base administrator and systems development. Systems development is considered to include the planning, analysis, design, and implementation of application processes. The data base administration would participate in the design effort and make determination of technical and economic feasibility in seeking to satisfy the data requirements of the applications processes. The data requirements of new systems may be satisfied by using data already collected and stored or by extending the definition of the data base to include new data requirements.

In cooperation with users, the data base administrator is supposed to seek to determine what data to collect and store and the criteria to use in validating input data and stored data. It has been suggested that the correctness of data is the responsibility of the user, while the protection of data is the responsibility of the data base administrator. A Diebold Research Report notes a trend towards “placing responsibility for data accuracy, validity, and so forth, in the hands of the user who will be served by the data base.” The report suggests the appointment of a “Prime Responsible Authority” (PRA) for each data base from its primary user division. The PRA interfaces with all users on one side and with the data base administrator on the other. In other words, the “prime responsible authority,” working in a user division rather than in the data base administrator’s organization, would be responsible for content, integrity and use of the data base with respect to all users within the organization, both within and outside of his own using organization.

An Auerbach document lists ten areas of functional responsibility for the data base administrator as it can be seen in Table I. Rather than discussing those ten areas, we shall focus on the unique functions of data base administration, which are: definition, storage and update of data; making the data base available to the using environment; informing and servicing users; maintaining data base integrity; and operations and performance.

**Definition, storage, and update of data**

The process of data definition begins in response to stated data requirements from the using environment. The first step in definition process is to design the logical data structure, incorporating as much as possible of the natural structure inherent in the data. In a sense, a logical data structure should model selected aspects of the operations and entities as they really exist. This is pointing out that to date very little work has been done on developing formal methodologies to aid in the process of data base design which is really needed though, with perhaps the exception of some current efforts in that direction being in progress. Once the logical data structure is developed, it is formalized in the Data Definition Language (DDL) of the particular data base management being used. Things that the data describe in logical data structure will eventually be stored physically in the computer system. Having defined some data to the system, the next step is to set up the mechanisms to acquire the new data and to bring it into the system.

Finally, the last step in the process is to execute management policies regarding update of data. Even if the value of data diminishes over time, it is not desirable to aggregate older data. An updating mechanism should be set up to store the older in off-line archival files for example.

**Establishing data availability**

One of the functions of the data administrator is to assist users in their search for data to satisfy their application needs.
requirements. He should maintain a Data Base Directory (DBD), in which are recorded the record types, and set types currently available to users. The DBD will then be the initial source for information relative to data availability. If it happens some data elements are not available within the confines of the existing data base, the data administrator will arrange the interface with the necessary data sources to satisfy the demands of the users. Such demands, of course, should remain within the cost constraints controlling the user and the data base administrator.

Some additional factors to be included when considering data availability are the following:

1. Present form and location of data.
2. Access techniques to be used.
3. Intended use of data in relation to its present accuracy, completeness and timeliness.
4. Need for modification of data.
5. Present authorizing agent for use of data.
6. Cost of providing the data.

Maintain data base integrity

The protection of the data base is an essential responsibility of the data base administrator. The data base represents a large dollar investment, and data contained in it is vital to all who use it. Alteration, destruction or disclosure of the data base may represent an enormous and irrevocable loss in time and money. Although complete integrity protection is never possible, a high level of protection should be kept as much as possible.

The general problem areas in data base protection are:

1. Data base access and manipulation;
2. Data base integrity;
3. Safe recovery/restart.

Data base documentation

In addition to being an organizer and an administrator of the data in the data base, the DBA is the prime documentarian and educator with regard to the Data Base Management System (DBMS) at his installation, and he should provide for the recording of procedures, standards, guidelines, and data base descriptions for proper use of the data base.8

STUDY RESULTS

Over twenty Data Base Administrators’ groups have been studied to date. Their companies ranged in size from $3 billion to $20 million in revenues and data bases ranging in size from 100 billion characters to a few hundred thousand characters. Despite the range of application and size of companies, we found remarkable similarities among approaches, problems, and successes. The results are presented here in summary form to support the main thesis. The detailed survey is being published in working paper form.14

Recall the primary thesis we set out to prove:

The data base administrator now and in the future is an individual who performs the functions of planning, designing, operating, and controlling the data base at both policy and operational levels.

Operations vs. policy

Critical in the thesis statement are the dual tasks of operations and policy making. This conflict, we found, is one area where the position statement breaks down. The data base administrator is said to be responsible for the definition, storage, and update of data; making the data base available to the using environment; informing and servicing users; maintaining data base integrity, operations and performance. The data base administrator is also said to be responsible for corporate data base policy; access authorities; definition of data base content and organization; selection of data to collect and store input; criteria to use in validating input; and conflict mediation among users, system designers, and computer operations.22

These latter functions required the DBA to be high enough in the organization’s management, so that the position has authority to set and enforce policy. The former functions, more operational, required less top management involvement and usually meant a lower position in the organization. Furthermore, the operational dimension caused the DBA to bias the broader policy issues affecting many groups in favor of his own operations.

Evolution of the DBA

As the study began, we uncovered another framework which shed considerable light on the DBA. A definite maturing of the DBA function occurs within an organization. The newer DBA’s were often focusing on problems that the more established DBA’s had already resolved. DBA’s at different stages of development did not articulate the stages that we observed. They related primarily to the current problem and to a continuum they were currently traversing.

Our formulation of these evolutionary steps made observation of DBA’s much simpler and more logical. First, it helped separate the progress of a particular DBA group and the progress of the field as a whole. By classifying each DBA group according to their development stage, we could study that group’s state relative to the stage and not to the general population. For instance, the DBA just getting started with a three person staff has significantly more support than the DBA in the height of development with three people. We then could ask if the problem definition
phase is receiving greater attention instead of observing the
gross state of the group.

We outlined five stages in the evolution of the data base
administrator function: Introduction, initiation, integration,
operation, and maturation.

Introduction generally took the form of either a study
group or a manager’s individual recommendation and deci­sion to go with a data base management system. Although
we did not take much specific data on this phase, most
earlier introductions seemed to discover a need for a data
base administrator after using a DBMS, while more recent
introductions also discovered a need for the DBA to
parallel or precede a DBMS.

Next, the initiation of a DBA consists of developing one
or two data base systems. As much as total system design
or top down approach may be touted, the corporate data
base was never built in this phase, it was always started as
one or two application data bases. Most often the DBA’s
speed of success rested on how well these applications
were performed (we say speed of success because despite
severe negative benefits in some systems, data bases are
becoming essential to most business organizations and the
question is generally how fast). It is in this stage lasting one
or two years that the DBA spends a considerable amount of
time and money to establish a base for future growth.
Correct choice of system, personnel, applications, etc. are
critical here.

Integration, the joining of several systems together into
the corporate data base, is a development phase of several
years duration. It really never stops, but we have estab­lished an (arbitrary) turning point into maturity when the
DBA has control of data base’s definition, design, access,
and standards along with having a large majority of DP
application systems on the DBMS. This points out the two
maturation phases going on during integration phases. One
is the actual integration of data into the data base, the other
is the acceptance of the DBA as the group in control,
mainly operational control, of the data base.

Maturity sees the focus of effort move from development
to operations. It also means a focus of the DBA role as the
center of data base systems responsibility from design to
operation. Some organizations might believe they have
matured with only one or two subsystems operating, but
until the DBA is involved from the beginning in all data
base work and also has the final power to make the
operational decisions, the group has not matured. In fact,
during our survey we found very few mature groups.

General experiences

For the study, we outlined several questions to help us
establish a more logical pattern to our research. Some
questions were posed directly to the subject; others were
broken down in more detail and synthesized here.

• Is there a measure of the size of the DBA’s organiza­tion?
• What are the qualifications needed to fill the DBA
position?
• Which of the ten functions does the DBA perform?
• Where does the DBA report in the organization?
• What is the cost of the DBA and where are the
economies?
• What is the biggest problem faced by the DBA?

Size of DBA function

The size of the DBA group is proportional to the size of
the data base up to a point. With one exception, the DBA
group increased with the size of the data base. The increase
was not linear, because the high initial cost to support the
system causes a rapid growth initially followed by a more
gradual but noticeable increase. The largest group inter­viewed had 14 people in the DBA group; the smallest had
one. Economies of specialization and scale took over as the
data base grew, and the staff became organized, usually functionally, but sometimes in project teams. We learned
that the start-up effort can be substantial. The analyst/programmer training consumed a large portion of the
DBA’s time during the first two years of operation, but fell
drastically after user acceptance. In many cases, training
and marketing DBMS went together, so the workload
further increased start-up staff size.

Once application systems arrived at an operational level,
the group did not shrink because of required data base and
program updates and other support functions. Surprisingly,
all those interviewed who had been operational for some­time, sighted more man-hours consumed on system prob­lems than data structure updates. As major applications
became operational, the role of the DBA staff grew more
operational and tended to increase, usually in support
control.

There was one strategic milestone for staffing which
passed quickly in some cases and never in others, and that
is commitment to data base technology. With the commit­ment given, staff size no longer was a problem. Without it,
the DBA found mustering resources, like staff, difficult. A
few groups were able to overcome this obstacle by per­forming well despite resource constraints; one group dis­banded, several more are not beyond this milestone.

Two final notes: first, staff size is not a measure of the
quality of a data base group, but it can be a measure of the
data base size. There may be a relationship between the
size/complexity of a data base and the size of the DBA
group. Next, there appears to be little relationship be­tween the type of DBMS and the size of the DBA

What are the requirements for a DBA

It was felt by those interviewed that the primary require­ments for a person to be a DBA or join the DBA group
were both technical skills and knowledge of the company, administrative capability ranked a weak second.

Sixty percent of the current administrators had lengthy experience within their companies' DP department, not necessarily with DBMS. The others were DBMS experts brought in for the job. The overwhelming majority of the administrator's non-clerical staff were technically trained in DBMS. The large majority of administrators suggested that their replacement should have both technical training and a minimum of two to five years with the company. Their staff members were required to have a strong technology background before being considered for hiring or transfer from an application area being implemented under the DBMS.

The technical skills requirement was even more pronounced in the more mature groups. They were finding a greater demand on the part of analysts and users for more detailed systems assistance. In all of the mature groups, the DBA group found themselves spending a large amount of time in the system support function. Frequently, they were supporting applications programmers as systems programmers because the applications people considered the data base part of the operating system.

Administrative skills were secondary qualifications to three-fourths of the DBA's. A person with a Master's in computer sciences and with company acquired skills is much preferred as a DBA over someone having a Master's in Business with technical skills. It was felt that administrative skills could be learned on the job. This opinion held for the very large DBA groups and the very small ones.

**DBA JOB FUNCTION**

The DBA function among organizations is remarkably similar in the long run, but start-up situations were crucially different. The actual job descriptions, the 30 percent we were able to see, came directly from the DBA literature.

We were very surprised that generalization about the DBA function could be made between different-size companies using different DBMS products. Each company and package have difficulties unique to itself, but histories, problems, and operational organizations were similar. The literature explained the functional areas for setting up operation, and the DBA's generally read these major sources for advice. The DBA groups we found were primarily operational functions whose objectives were technical support for applications analysts. The organizations tended to grow more support-oriented with time.

Of the ten tasks outlined earlier, implementation, access, DBMS enhancements, education, and vendor enhancements were the five major areas of concentration found in young groups, in the initiation phase. In this phase, data definition and design were not relinquished by the application analysts. Control, documentation and monitoring were not institutionalized very well in the DBA functions. Operations were almost always delegated to the company's operations department under advisory relationships with the DBA. Often backup recovery was a major headache for the newly initiated data base group until operations had accepted responsibility for its execution.

These young DBA groups were brought in at the last stages of applications development to "make this a data base system." The applications analyst would have done all the work, built the system, and as a last touch, attached the DBMS. Education, in the form of data base techniques and data base design principles, became a major consumer of the DBA group's time. The education served three purposes: sell DBMS as a concept, sell the need for involvement of the DBA early in the application design, and teach the analyst how to use the DBMS.

The evolution towards involvement of the DBA in application design became evident in the large majority of mature systems. The groups generally gained this involvement via user acceptance, not by fiat. As the groups matured, internal controls, documentation and monitoring became a part of daily operation. Backup and security was one area of responsibility serviced from the start. The DBMS generally voided normal file backup operations, yet, the operation was required. The DBA set up procedures for the operations group to follow to back up the DBMS.

Vast differences did occur in initiating the DBA function and bringing it to mature operation. These differences appear in organizational consideration, top level commitment to the DBA's, and interfaces with operations, applications, development and users. There is genuine need for further research in this area to support the start-up and evolution of a DBMS and the DBA, especially relating to these areas.

**Where the DBA report in the organization**

Data base administrators rarely report to the highest systems' administrator, but more frequently report to the manager of systems analysts.

Current authors in an attempt to improve the DBA position's strength have an extensive list of policy roles for the group to play. These roles do not yet occur in the practical world. Being the mediator and direct advisor to the chief systems administrator on data policy has been recently proposed. In practice, we have not seen this occurring in the person of the data base administrator. The DBA is most frequently a senior analyst or group leader in applications or systems development organizations. In only two cases did the data base administrator report to the top level of systems management.

We recognize the need for better data base policy in organizations. But from the history of current DBA's, this policy maker would probably be another individual with the DBA serving as an important part of policy implementation. In fact, historically, there generally was an individual or group of individuals who took on this policy role. They instituted or supported the policy to use a DBMS and to initiate the DBA function. The creation of a position to plan and control the use of data resources is one way to institute data base policy. The formalization of a committee is
another way to institute policy. Either way, it appears to be a job separate from the DBA's job.

Cost and economies of the DBA

Two facts stood out in our survey in this area. First, the organizations had little idea what the direct cost of the DBMS or the DBA were. Second, the organizations recognized that they required a DBMS and the DBA to economically operate their data processing system.

It is astounding how poorly the cost and economies of DBMS are measured, for that matter it is astounding how poorly the computer resources are measured. In general, an astute applications group within these companies should immediately take advantage of the DBA and attempt to use the data base because the charges for these resources were not accounted for directly. In some cases operations cost, computer run time, was charged back to the user, and served as a deterrent to using the data base. But generally the DBA groups were an overhead item, a necessary expense to keep the system operational.

The rush to be "on the data base" did not occur. However, the costs or economies were rarely the reason given for not using the DBMS.

No good data about the economies of these data base systems came out of the study. At most it can be said that they are expensive to build and to run. A DBMS needs a group of experts, the DBA, to run and maintain it in order to keep the unaccounted for costs down to some unknown amount. Organizations are willing to pay some proportion of the budget to support the systems and the staff. And, the organizations who had committed to the use of a DBMS felt they were getting good results from the change.

The major problem stated by the DBA

Administrative problems, or organizational issues surfaced as the most important problems in the great majority of groups studied. After insisting on technical people to staff the group, the DBA manager cited his critical problems occurred in a variety of administrative areas. These problems were often a function of the point of evolution of the group.

The groups just starting up cited top management support, cooperation from user groups, and training or education as their most frequent problems. The more mature groups cited control and coordination problems along with evolving technology difficulties. The younger groups generally felt satisfied with the technology, but felt constrained by the environment, while the older groups although still bothered by administrative problems had technical difficulties as frequently mentioned as organizational ones.

For the DBA in the initiation phase the major problem revolved around either introducing a new technology into the organization or the creation of this cross organization body that the DBA represented. We expected to find these groups were highly critical of the technology or with an extensive list of needed advances. Instead we found general satisfaction with the DBMS product they used and frustration with its users. Sometimes the problems appeared to be the overanxiousness of the DBA who had been mandated to put in an application and was trying to install the total corporate system. More often the reasons were problems with the education and training of others outside the DBA on proper approach. Most typically the young group was seeking the responsibility for design and development instead of just the programming of the DDL or DML. This interface with the users, the establishment of educational programs, the convincing of top management was the slowest and most painful effort, yet generally a successful one.

The mature groups had achieved these responsibilities and were established as a necessary component in any new development. Here we did find complaints about technology: a lack of a data dictionary, or better recovery control, or a solution to the on-line update problem. We also found more administrative control problems surfacing. Their concerns about participating in system change decisions and about data control problems were more pronounced. There was a distinct interest in establishing responsibility for the data but at the same time an unwillingness to take the responsibility. They also expressed concern about control over changes to the data base and its structure. Finally they were searching for people with the qualifications and the ability to work in the DBA group.

CONCLUSION

The DBA is a necessary part of any data base management system effort. No corporation should attempt to form a "corporate data base" without this individual or group to manage its internal operations. The DBA is not however the policy maker that we see in the literature. Corporations do not really view data as a resource nor do they view the DBA as a manager of a valuable corporate resource.

The DBA's we surveyed are highly qualified technical teams or individuals that first bring a new technology to bear on data processing problems and then make the technology operate with the maximum effectiveness possible. They are in a unique position of bridging various applications and have an extraordinarily difficult time bringing the available capabilities to potential users, but they seem to perform this job reasonably well. The DBA group now and in the future needs well qualified technical people who can interface with their users and colleagues.

At the same time they need to attract those individuals who have a knowledge of the organization's problems and who can interface with the users from a different perspective.

The DBA is an entity with a future. We see it as a training ground for managers who need a broad view of the company's information system problems, yet who need to have the technical capabilities to manage some user groups. The DBA seems to naturally focus more and more on the technical issues which are within its control and which grow as the position grows. The DBMS technology is being
applied to more and more problems in more and more organizations. As the organizations continue to utilize information systems the DBA will provide the data base support necessary, probably on a more expanded technical front.

As far as data base policy is concerned, we see a different individual or group becoming concerned with this problem. The DBA would be the most important implementer of the policy, but not the policy maker. The corporate information managers will need staff or support people who can advise them on policy matters without the vested interest of the DBMS implementation. These policy makers would be concerned with the data resource as one of the corporate resources. They would draw on the DBA group for talent and individuals to move into the advisory wing, but this would probably not be the same group.

The DBA will be involved with technology growth problems as well as the growth of his data base. The systems studied showed that most DBMS systems operate in batch oriented systems. Of those that don't, only a few allow sophisticated on-line update and access to the data base. Most of the systems were not using the operational data for more expanded MIS purposes, either. We expect the demand for on-line update, and the demand for more MIS applications to grow along with the introduction of more sophisticated data base techniques like distributed data bases. The DBA's technology growth will also force the group to become highly technical in nature and highly specialized.

Further research and assistance is needed for the DBA and the user, who will be interfacing more with the DBA in the future. More automated design tools must be developed for the user to be able to present his design to the DBA in a clear manner both at initial development and at operation time. Very few of the DBA's had the time to monitor and optimize the use of the data base. More research into modeling and optimizing would aid in this area. The problem of an MIS still has not been solved in these systems. The data bases become so complex that managers are more lost than ever, and research on an effective way for true management use of complex DBMS systems is needed. Research in the economics of DBMS and their staffing is clearly needed. Organizational issues and administrative structuring seem to be a must for expanded research. For those who claim that data is a resource the last two issues are clearly a critical area to justify that position. Finally, the impact of the new technologies on the DBA must be a constant concern.

REFERENCES
