Stephen F. Crane has worked with UNISYS (Burroughs) mainframe computer software since the early 1970’s, specializing in Burroughs’ Data Management System II (DMSII). Working for the state of Michigan from 1972 to 1977, he helped pioneer the use of DMSII at the Department of State Police and later, working for the Department of Transportation, he joined a statewide task force to evaluate the use of “structured” methodologies for programming, database design, and systems design. He also spent three years as Director of Education for MIS, International, Inc., For the past ten years, Crane has acted (and, he freely admits, occasionally overacted) as Database Administrator for Koch Industries, Inc., a petroleum company headquartered in Wichita, Kansas. Koch is the second largest closely held company in the country and had revenues of $16 billion in 1987—the last year the company released a figure. Koch uses DMSII extensively, having two large UNISYS A Series mainframes and a large network of PC’s.

Interview by Mohammad Dadashzadeh

JDA: What are the major responsibilities of the Database Administrator at Koch Industries?
Crane: The Database Administration group has four overall responsibilities at Koch Industries. First, we assist in the design of databases for computerized application systems. Second, we install and maintain those databases, almost exclusively using UNISYS’ DMSII. Third, we provide the primary technical mainframe support for the application system design & programming staff... developing specialized utility programs, evaluating new software tools, answering questions concerning system software, performance, etc. Fourth, we are currently installing a new data dictionary system and end-user reporting tool for which we will also be responsible. And, of course, we perform an occasional miracle!

Our role is primarily that of a service organization for the system and program developers.

JDA: What do you consider to be the most important contributions of the DBA to an organization?
Crane: That is highly dependent on the company. No theoretical scheme can ever suit all organizations. One company may be highly centralized; another may have many widespread, fairly autonomous sub-organizations. And each company has its own ideal “information network”. But two fundamental areas where the DBA of any organization can contribute are in developing and maintaining a practical data dictionary and in maintaining an awareness of new technology.

The data dictionary is an important tool for a variety of reasons, the most important being that it can provide the foundation for all application systems development. After that, however, it too becomes highly dependent on the needs of the organization. An organization maintaining a relatively static transaction processing environment might be satisfied with a simple listing of entity definitions. Another organization might need a sophisticated ad hoc query capability utilizing distributed processing, requiring that the data dictionary contain not only the database entities and definitions, but also the complete database semantics, both obvious and subtle. Since the latter form of data dictionary is much more costly than the former, the organization, with the help of the DBA, must carefully weigh the costs and benefits.

The second major contribution, keeping aware of new technology, is more than challenging with the speed of today’s advances and the workload many DBA’s maintain. But it is an important process, since the data dictionary and the DBMS are so critical to most MIS organizations, and the MIS organization is critical to the operation of a company... no matter what your boss says.

Unfortunately, at Koch Industries, our information requirements have far outpaced our data dictionary’s ability to keep up. But, for the most part, this is not due to a lack of technology. I am the first to admit that we didn’t anticipate our true information processing needs until the users stormed the fort. We have just recently begun to invest the necessary resources to install and maintain what promises to be an effective data dictionary.

JDA: What is the computing environment at Koch Industries? Do each of the major groups at Koch have their own separate (and independent) database?
Crane: Koch Industries is highly centralized, with all mainframe development and computing performed in one location. We do, however, maintain a large network of terminals and PC’s, with many users wanting to select data from the mainframe and download it to their PC’s for additional processing of their own.

Anomalous with Koch’s highly centralized physical environment is its decentralized approach to designing computerized applications. While some applications have been designed for the company as a whole, most applications are separate, developed for a certain sub-organization, and some redundancy is evident. But even in
those cases, there is some sharing of data. The DBMS we use does not restrict programs from accessing multiple databases.

**JDA: What are some of the major computer-based information systems at Koch?**

**Crane:** Almost all of Koch’s information systems combine on-line transaction processing with batch processing... from our general ledger, accounts receivable & payable systems, to our manufacturing and purchasing systems. We process about 60,000 on-line transaction per day. We are not a financial institution nor are we the Department of Defense, but we do have our share of triumphs and tribulations.

Almost all of our systems use UNISYS’ DMSII, a network DBMS. We have 36 physical databases, consisting, in total, of approximately 900 data sets and 2,600 access paths (sets and subsets). A physical database might consist of multiple logical databases (applications) not necessarily related to each other. Combining unrelated applications within one physical database is usually the result of performance considerations, and generally has no effect on programming effort. So, when I refer to an application’s database, I am referring to the logical, not physical, depiction.

Our DMSII databases currently require about 20 gigabytes of disk storage. DMSII has proven to be very reliable and flexible, easily satisfying Koch’s DBMS requirements. But DMSII does have one considerable weakness; a weakness that has sometimes affected our ability to satisfy the end users’ needs. UNISYS has been unable to develop an adequate end user reporting/inquiry tool. They have tried... again and again... but the best they could develop was cumbersome, incomplete, and user-antagonistic. That is why we have recently installed, along with a new data dictionary, a very user-friendly reporting tool developed by another vendor, EDP Systems, Inc. (ESI).

**JDA: Does Koch Industries have an integrated database?**

**Crane:** If you mean a database reflecting the overall view of our information environment, no. We do, however, have certain databases that are “global” to Koch Industries. These databases might be maintained by a specific application (for example: accounts payable, accounts receivable, manufacturing, etc.), but contain data shared by other information systems. A good example is vendor information. While this data is maintained in our accounts payable database (one of our company-wide databases), it is used by many different information systems. Again, DMSII places almost no restrictions on a program’s ability to access data in multiple physical or logical databases... aside from any security we build in.

So, in the purest technical sense, we do not have an integrated database. Conceptually, however, we have a partially integrated database environment; and with having recently installed a new data dictionary, we will move even farther (conceptually) in that direction.

**JDA: What do you consider to be the major disadvantages in having separate databases?**

**Crane:** Technically speaking, with a DBMS like DMSII, there are few significant disadvantages. (Am I beginning to sound like a UNISYS marketing representative?)

Conceptually, however, several deleterious effects are caused by the intrinsic redundancy in separate databases. An entity stored in two places has to be maintained, secured, and validated for integrity in two places. Available machine resources, such as disk, CPU, and IO are depleted.

Unfortunately, until recently, Koch has done a splendid job of protecting redundancy’s right to proliferate. We can blame this on a lack of tools... but, in reality, a fully integrated database, minus redundancy, is not a realistic expectation in today’s business information processing environment. If my boss were to ask me to “eliminate redundancy in our databases”, I would reply, “Give me fifty more people, two years, a large mainframe system, and a blank check for the necessary software tools”. Then I would spend at least a portion of that two years updating my resume.

I may sound pessimistic, but in the real world an MIS organization must spend its resources wisely. Real world information processing is more complex than the theorists and the hardware/software manufacturers often realize. I remember sitting in the audience during a session at a 1986 CUBE conference (Cooperating Users of Burroughs Equipment), when the speaker joked about testing DMSII’s performance using a database with almost 300 physical structures, going on to say that she could not imagine a real database of that complexity. Many of us in the audience, from larger companies, have several databases at least that large. Needless to say, we were incredulous, especially considering the speaker was the head of DMSII development for UNISYS!

On the other hand, I welcome any new, cost effective tools that can help me to integrate databases and reduce that redundancy. I think our data dictionary system has arrived just in time.

**JDA: What tools would you consider or have at your disposal if you were to integrate the separate databases?**

**Crane:** The primary requirement would be a state-of-the-art data dictionary, such as the one we are installing, DATAPULSE, developed by ESI; or the data dictionary UNISYS is now touting, ADDS.

**JDA: Is Koch Industries planning to migrate to a rela-**
Related Content

Large Scale Graph Mining with MapReduce: Counting Triangles in Large Real Networks
Charalampos E. Tsourakakis (2012). *Graph Data Management: Techniques and Applications* (pp. 299-314).
[www.igi-global.com/chapter/large-scale-graph-mining-mapreduce/58616?camid=4v1a](www.igi-global.com/chapter/large-scale-graph-mining-mapreduce/58616?camid=4v1a)

Performance Evaluation of Parallel S-Trees
[www.igi-global.com/article/performance-evaluation-parallel-trees/3253?camid=4v1a](www.igi-global.com/article/performance-evaluation-parallel-trees/3253?camid=4v1a)

From ‘Flow’ to ‘Database’: A Comparative Study of the Uses of Traditional and Internet Television in Estonia
[www.igi-global.com/chapter/flow-database-comparative-study-uses/7987?camid=4v1a](www.igi-global.com/chapter/flow-database-comparative-study-uses/7987?camid=4v1a)

MECP: A Memory Efficient Real Time Commit Protocol
[www.igi-global.com/chapter/mecp-memory-efficient-real-time/20760?camid=4v1a](www.igi-global.com/chapter/mecp-memory-efficient-real-time/20760?camid=4v1a)