Celebrating IBM Data Management

White Paper

September 1998

Charles J. Bontempo
e-mail: bon@bestweb.net
Adjunct Professor of Computer Science
Polytechnic University

Contents

Celebrating IBM Data Management

Introduction

This is the 15th anniversary of DB2 for OS/390 and MVS, the 30th anniversary of the Information Management System (both IMS TM and IMS DB) and the tenth anniversary of the International DB2 Users Group (IDUG). It is also the year in which the IBM data management group has announced significant enhancements to their data solutions, especially in its database management systems (DBMS) and business intelligence (BI) offerings. Finally it is a year in which recent advances by IBM's DBMS and BI offerings have garnered an impressive array of awards, rankings and comments from the industry's press, consultants and analysts.

Perhaps the highlight of these awards is from Database Programming & Design which assigned IBM first position among the top twelve companies that "most influence the data-driven vision". Commenting on their selection of IBM, the editors noted that after its introduction in 1983, DB2 for MVS "immediately became the standard bearer" for RDBMSs. Commenting further on DB2 as a player on *non-IBM platforms*, the editors stated that "with the release of DB2 Universal Database Version 5, the data management group looks like it is once again fielding the standard bearer."

This DB2 anniversary also elicited the following congratulatory remarks from Chris Date, widely recognized database authority and contributor to the relational database approach:

> I am very pleased to congratulate the DB2 team on their achievement. Although it wasn't the first SQL product to market ---not even the first from IBM --- DB2 for MVS was clearly the product that put the seal of approval on the technology and set the stage for the modern scene.

In the data warehouse and datamart arena, the META group recently judged Visual Warehouse as "priced right, fully powered and partner-ready" and a "solid value in bundled DW functionality" that "includes certain industry leading features...". (META Trend, May 5, 1998.)

These data management anniversaries are an appropriate occasion for a review of IBM's most recent and longer term data management milestones. They can be used to point to the direction and promise of its future achievements in these key areas of data management.

Enterprise Level Data Solutions

IBM's user community consists of a wide range of industries and broad array of public service enterprises. This has given the company far-reaching experience in meeting the requirements of true *enterprise level* data management, *i.e.*, data management to support those information requirements that transcend the needs of a single business function, are critical in meeting strategic and tactical business objectives and provide access to data from users throughout the organization. Its solutions range from the smallest to the largest systems, operating with all of the well established system architectures - - - from local single site, centralized configurations to globally distributed, multi-tiered systems. Thus IBM can

convincingly claim no need for a one-size-fits-all approach in its data management solutions and it can offer one of the most comprehensive set of data management solutions available today.

In meeting the requirements to support the mission-critical business functions of this diverse and broad community, IBM has focused on quality offerings emphasizing reliability, robustness and system features that assure the integrity of its users' data. Since it has always developed data solutions on its own multiple platforms, it has met the challenges of network computing, interoperability, and the effective consolidation and integration of data in diverse operating environments at the enterprise level. *Thus IBM's data management history can be read as a continuing effort to extend and strengthen support for data sharing --- to broaden the domain of users as well as the domain of data that is managed as an enterprise resource*. In addition to the support for multi-vendor data sources provided with its Business Intelligence Solutions, examples of its success in meeting these challenges include:

- its Distributed Relational Data Architecture (DRDA)
- the unmatched advanced technology of DB2 DataJoiner including its support for heterogeneous (multi-vendor) replication
- the direct access to multiple DB2 databases from many different client environments now possible with Query Management Facility (QMF)
- the object/relational capabilities of its DB2 Family members

In providing its DBMS solutions on multiple platforms, it has also succeeded in exploiting the hardware and operating environments of each of its offerings. For example to support data sharing, DB2 and IMS take advantage of the MVS Workload Manager as well as the System/390 Parallel Sysplex Coupling Facility. IMS also exploits the Coupling Facility in its support of shared message queues. The database optimizer of DB2 Universal Database V5 (for OS/2, Windows, AIX and other Unix platforms) exploits the different capabilities of the many environments in which it operates. DB2 for AS/400 takes advantage of its single-level store architecture and its deep integration of key RDBMS functions within OS/400. For these reasons, industry observers generally acknowledge IBM's sustained success in providing well-conceived, sound enterprise-level data solutions.

IBM Advances in Data Management Solutions

IBM's many longer term achievements in these areas should also be noted. For example IBM research was the first to produce and document a cost-based database optimizer; research in this area continues with results applied to parallel and object/relational database processing. This is the first optimizer with *systematic query rewrite* capability that enables improved query performance for SQL generated by the many front-end interfaces currently in use with DB2. Its replication tools now provide "update anywhere" capability with automatic conflict detection/compensation and with support for mobile users. IBM was a leader in providing object/relational capabilities with DB2 Common Server --- capabilities that have been extended in its DB2 Universal Database V5 for OS/2, Windows 95, 98 and NT, Sun Solaris and HP-UX, and more recently in DB2 Universal Database Server for OS/390 V6. The latter RDBMS now provides the "beef" *e.g.*, stored procedures, triggers, ODBC and CLI interfaces, to support its use of the term *enterprise server*, indicating *true client/server database processing on mainframe systems*. DB2 for AS/400 is the first RDBMS to offer Encoded Vector Indexes, highly useful for business intelligence applications. In the longer term, DB2 for MVS was the first RDBMS to offer declarative referential integrity. DB2 and IMS have been leaders in providing high volume, highly available database processing, capabilities that are now crucial for data warehouse as well as OLTP applications. IMS recently

introduced *the first shared message queue capability for clustered systems.* For years these DBMSs have been supported by comprehensive sets of complementary support tools.

IBM's fundamental contributions are more obvious. SQL, the universal language of RDBMSs for the foreseeable future, originated in the IBM research labs, Of course the relational database approach orginated at IBM research. IMS was a pioneer in achieving transaction integrity for online database processing. More generally, DB2 and IMS have given real meaning to the term *industrial strength* database management.

In the Business Intelligence (BI) area, IBM developed and published its data warehouse architecture in 1993. Today its BI offerings, particularly Visual Warehouse, The Intelligent Miner Family, DB2 OLAP Server, and DataJoiner represent advanced technology for those customers who seek the benefits of affordable online analytical processing (OLAP), data warehouses, data marts, and proven data mining techniques, while leveraging their current data assets and their relational database skills
.

IBM's impressive number of patents in the data management area are well documented. Its research and development communities include an unusual number of IBM and ACM Fellows, IBM Distinguished Engineers, and technical staff who are highly prized throughout the industry as contributors to data technology. Also IBM established the Database Technology Institute (DBTI) to accelerate the adoption of proven research advances by its products. The company is a leader in proposing standards as an active member of the ANSI and ISO SQL standards bodies.

Innovation: Knowing What, When and How

Notwithstanding this record of advances in data management, IBM has avoided the pursuit of the "latest technology" *for its own sake,* as noted by Don Haderle, IBM Fellow and vice president for IBM database technology:

> We are always pursuing new avenues of data technology. But at the same time, we continuously evaluate them for the potential business value they offer our customers ---- for the contributions they can make in supporting their current and future business objectives.

Relational Database at IBM: A Brief History

The seminal work on the relational database approach by Dr. E. F. Codd while at IBM research culminated in the publication of his 1970 paper "A Relational Model of Large Shared Data Banks," which appeared in the *Communications of the ACM* (June, 1970). This and other publications by Dr. Codd and Chris Date, his colleague at IBM at that time, evoked strong interest throughout the growing number of DBMS users and developers in the industry.

Dr. Don Chamberlin and his colleague Ray Boyce, both at IBM research, went on to produce SEQUEL (Structured English Query Language), later renamed "SQL" by IBM. SQL replaced Codd's mathematical notation with a simpler keyword syntax and introduced operators for database updates and for grouping. It enabled users to write applications that are independent of access paths and provided a unified syntax for queries, data definition, data manipulation and control.

The Great Database Debate

IBM's early research on relational database was viewed by some as a serious threat to the standards efforts being made at that time for network DBMSs, based on the specifications of the CODASYL Database Task Group. To further understanding of the issues raised, the *ACM* conducted a workshop including a debate between the proponents of these two approaches.. The intensity of this debate produced a dramatic polarization into pro- and anti- relational camps. However, interest in the relational approach grew --- thanks to the efforts of its proponents in "spreading the word." (See *Data Models: Data-Structure-Set versus Relational*, ACM SIGMOD Workshop, 1974.)

The System R Research Project

While the SQL language development proceeded, IBM research began work in 1973 on System R, the most ambitious of several IBM experimental relational DBMSs. The goal of this project was to demonstrate the viability of a full function relational DBMS. Some of its key architectural features were: a two-subsystem structure consisting of the search engine, and a high-level component that performed SQL statement processing, optimization, view processing, security, computation, result set processing, and integrity. Many of the research results of this project were described in IBM and ACM publications. (See for example, "System R: Relational Approach to Database Management," *ACM TODS*, June 1976.) This research effort achieved significant results in many areas of relational *theory and practice*. It is probably best known as the first DBMS to implement SQL and for its pioneering work, led by IBM Fellow Dr. Patricia Selinger, in cost-based search optimization. Also its two-layered system model has been adopted by the DB2 Family. The project was concluded in 1979 after having been tested at three IBM customer locations. Research efforts then turned to distributed RDBMS with the R* ('R-star') research project and later to object/relational database research with the Starburst project.

.

Introducing the DB2 Family: Validating the SQL-based Approach

In 1980 IBM delivered the System /38, an integrated database server.. This was followed by SQL/DS for VSE and VM, whose original version was closely modeled on the System R research prototype. DB2 for MVS V1 was introduced in 1983. The announced goal of this version was to provide the simplicity, data independence and user productivity that this new approach promised. Later releases of DB2 focused on improving its performance, reliability, and capacity to levels required for serious OLTP (Online Transaction Processing) in the service of a wide range of demanding industries, *e.g.*, banking, customer service, distribution, transportation, brokerage. By 1988 DB2 for MVS had demonstrated impressive thruput for OLTP. It was supporting distributed database at the levels of *remote unit of work* in 1989 and *distributed unit of work* in 1993.

IDUG: The DB2 Extended Family

In 1988 The International DB2 Users Group was established as an independent, nonprofit user organization under the auspices of PLATINUM technology, inc. and KPMG Peat Marwick. Its aims are to provide quality education and services and to promote the effective use of the DB2 Family of database managers. Today it is made up of thousands of users, consultants, vendors and associate members. There are four annual IDUG conferences around the globe: North America, Europe, Asia-Pacific and the IDUG Technical Conference conducted in Canada. IDUG also serves as the liaison for many DB2 User Groups worldwide and publishes the *IDUG Solutions Journal*.

IDUG's first Annual North American Conference in 1989 exceeded the expectations of its program committee. According to Ken Paris, IDUG Vice President for Conferences at the time, the conference committee was "especially interested in those presentations which confirmed that DB2 was a full blown production DBMS to support mission critical applications." The important role that IDUG and its members play in the continuing development and success of the DB2 family is clearly acknowledged by Janet Perna, general manager of IBM database management:

> IDUG is a key development channel for DB2 and a great way to reach DB2 customers. I certainly view it as a key member of the "extended" DB2 Family, and I personally value highly the opportunities I have had to contribute to its programs.

Other important user groups for IBM data management include SHARE and GUIDE International and COMMON. The DB2 Technical Conference is a key event for the exchange of information on IBM's data solutions.

The DB2 Family: The Future is Now

The following sections of this paper describe some of the key features and capabilities of IBM data management solutions. The reader is referred to the many white papers available at www.software.ibm.com/data for more comprehensive treatments of a particular solution. It should be noted that all of the DB2 Family treated here provide features to support the client/server paradigm, including *nonproprietary* stored procedures (SPs), triggers and a wide range of client interfaces. IBM and other vendors offer many applications that access data managed by each of them. DB2 Connect can be used for direct client access to these DBMSs without need of a gateway. They are Year 2000 ready, Web-enabled, and IBM's Net.Data can be used for Web-based access to all members of the DB2 Family. They can be used with Visual Warehouse to support target data marts and data warehouses and as the data sources for Intelligent Miner operations. Finally, they offer strong RDBMS capabilities in support of Business Intelligence (BI) applications: high scalability with inter- and intra-query parallel processing with a wide range of system architectures --- shared data, shared nothing, and clustered configurations. They provide high capacity for multi-terbyte databases, specialized indexing techniques for BI applications and optimized performance for star query processing. To enhance query performance, their advanced query processing capabilities include SQL query rewrite capabilities. And their database optimizers employ an SQL "push-down" strategy that supports a basic axiom of database processing: process data as close as possible to its source.

Prospective DBMS users are well advised to consider the DB2 Family and IMS as part of an effective BI strategy of putting data to work to their business advantage. The synergy they offer with companion BI offerings from IBM and its partners can provide the best avenue to a competitive and profitable advantage for many customers.

DB2 Universal Database Server for OS/390 V6: Enabling New Applications

With the announcement of DB2 Universal Database Server for OS/390 V6, this keystone of the DB2 family celebrates its 15th year of unparalleled success as the leading RDBMS for mission critical, mainframe database applications *This new version establishes its functional parity with the object/relational features of DB2 Universal Database V5 for OS/2, Windows 95, 98 & NT, AIX, Sun Solaris, and HP-UX*. The goal of object/relational DBMSs is to enrich the semantics of the database and to equip the DBMS to assume a more active role in *informing* its users. These object oriented capabilities

enable more flexible data modeling for a new class of Web-based , business intelligence applications that require support for content management and knowledge discovery. *Without these new capabilities the costs of such applications can be prohibitive*. More generally, they provide many capabilities of the object oriented approach that substantially increase productivity: retrieval and manipulation of large objects, user-defined types (UDTs) and user-defined functions (UDFs), and triggers. V6 also provides new SQL built-in functions to manipulate this complex data and a set of IBM developed *extenders* for audio, image, voice and text data.

Version 6 supports Large Objects of up to 2GBs of data for a single column value and up to 4000 terabytes for all values of a single column. *DB2 Large Objects can be SQL-referenced and manipulated just as any other DB2 data type*. To control network costs, client applications need to read only those parts of the large object needed. UDTs are data types that are integrated into the database for better performance and encapsulated for integrity. They are defined as *distinct types* and can be used to prevent erroneous operations on data with incompatible application semantics. For example, defining distinct types for euro currency and German marks will prevent misguided attempts to add or subtract these distinct monetary values. UDFs, like DB2 stored procedures (SPs), can be written with widely used object programming languages. For example, a UDF might be developed in C++ to convert German marks to euro currency. To enhance performance UDFs execute on the server; to preserve integrity they run in their own address space. Also like DB2 stored procedures, their priorities are managed by Workload Manager.

Triggers are predefined actions that are "fired" automatically by a database event. They can be used to enforce business rules, update summary data (useful for BI applications), automatically generate values for newly inserted data or to invoke a UDF. They can also initiate an action *outside of the database*, like enforcing a business rule by e-mailing a manager that a purchaser has ordered parts from an unapproved supplier. DB2 triggers can also activate SPs that access non-DB2 data, *e.g.*, update DL/I data. These changes can be coordinated as part of the same unit of work as the triggering SQL operation.

Extenders leverage DB2 support for large mulimedia objects. They can be used to store and manipulate image, audio, video, and text data along with attributes relevant to each type. For example, the text extender could be used for a patent search application or to categorize, index and search a corpus of customer complaint correspondence.

Version 6 also strengthens its support for parallel processing of shared data for those heavy query applications typical of BI applications. It enhances the high availability provided with its data sharing capabilities, including markedly improved performance of DB2 utilities, which signals IBM's renewed stress on providing best-of-breed utilities. V6 also offers advanced support for Web-based access to its data bases --- databases that can be enriched with the multimedia data so important for effective e-business applications.

Current users of this system for OLTP who are planning to deploy a data warehouse should consider the flexible system configurations that are now possible with DB2 support on the S/390 Parallel Sysplex platform. One configuration combines both OLTP and data warehouse data in a single data sharing group. This allows efficient SQL joining of both kinds of data; dynamic system reconfiguration to accommodate fluctuating workloads; and most importantly, reduced costs of computing. Stress testing of mixed OLTP and large query workloads at the IBM Teraplex Center has demonstrated the effectiveness of DB2's collaboration with the Workload Manager in achieving balanced performance for such mixed workload.

Increased Availability, Capacity, Scalability and Performance

Version 6 offers several enhancements to increase the availability of its databases. With the Group Buffer Pool (GBP) Duplexing capability, DB2 can write changed pages to both a primary and a secondary GBP. If the primary GBP fails, DB2 can switch to the secondary GPB for fast recovery and higher availability. A common performance problem of parallel DBMSs occurs when data becomes unevenly distributed or skewed across partitions as a result of database update activity. Rebalancing partitions to achieve a more favorable distribution of data could require a complete reorganization of the data across all partitions. Version 6 addresses this problem with its new SQL ALTER INDEX statement that changes partitioning boundaries so that only the partitions affected by such a change are subjected to reorganization while unaffected partitions remain available for access.

Users will benefit from the performance enhancements and the increased capacity IBM introduced in other recent versions of DB2. (See the white paper "Bringing OLTP to Network Computing".) V6 supplements V5's JDBC support with SQLJ ---- JAVA embedded SQL code to provide the benefits of JAVA portability along with the performance and security of *static* SQL.

Maximum table size has been increased to 16 terabytes.. In addition, it now supports up to 256 gigabytes of virtual buffer storage in data spaces. It offers more page size options for more efficient use of storage and improved performance. DDF Connection Pooling enables DB2 to pool connections among thousands of TCP/IP or SNA connections.

To assist in controlling the occurrence of runaway queries, Version 6 supplements the DB2 resource limit facility with a new predictive governor that terminates what would otherwise be long running queries before they begin. The new Statement Cost Estimation facilitates statement tuning. Applications using three part names can now use the DRDA protocol.

DB2 V6 offers convincing evidence that an enterprise database server can play a key role in today's network computing environment for both OLTP and BI applications. In addition to its economies of scale that translate to lowers costs of ownership, it can be a highly viable strategy for consolidating, securing and managing enterprise-wide data and for making its data Web-accessible to the widest range of enterprise users involved in conducting electronic commerce.

DB2 Universal Database (UDB) V5

This is the RDBMS that put real teeth into the term *universal*. It has received plaudits throughout the industry for its object/relational capabilities including its support for those applications that retrieve and manipulate multimedia data. Of course it provides all of the object/relational features that have already been described for DB2 Universal Database Server for OS/390. It runs in all of the major UNIX operating environments, as well as OS/2 and Windows 95, 98 & NT. In providing what IBM calls "Universal Accessibility," it clearly demonstrates the level of IBM's commitment to real openness and portability. Like other members of the DB2 Family, it is fully enabled for Web-based access through the facilities of IBM's Net.Data. The "universal" idea doesn't stop here. This system is available in several editions, each of which is suited for a different community of users --- from those who operate in "personal" mode to workgroups, to enterprise users with network-based facilities, and to software developers.

Scalable to Very Large Databases

This RDBMS scales from uniprocessor configurations to MPP systems of clustered SMP machines. It also exploits 64-bit computing. Both UNIX and Windows users with interest in BI applications will find that it delivers parallel processing capabilities second to none. Its shared nothing architecture (no shared memory and partitioned databases) offers an economic, incremental growth path to support multi-terabyte databases.

This universal RDBMS "parallelizes"all major database processing operations for both retrieval and database update. It performs *intra-partition* and *intra-query* parallelism with both horizontal and vertical (pipeline) parallel operations. It can execute parallel operations at a granular level, *e.g.*, sort, merge, and join, for substantial reductions in elapsed query time. Parallel tasks execute asynchronously. It also parallelizes sub-Selects, joins (including outer joins), the familiar SQL set operations, database loads, backups, and restores. Database loads can be performed with or without logging to speed up processing.

Smart Search Optimization, **Enhanced Availability**

To enhance performance the system's database optimizer uses knowledge of its operating environment (including Windows NT) and of expected network costs in producing its optimal search strategy. (Its recent TCP benchmark results score it ahead of its competition in key categories.) Since the system supports intelligent partitioning, it also exploits its knowledge of data placement. Unlike some of its competitor's optimizers, it develops an *overall* parallel access plan, rather than a serial plan that then executes in parallel. This system boasts the Starburst optimizer with its systematic query rewrite capability which actually rewrites SQL statements in a form that achieves better performance than the original SQL formulations. This feature can compensate for SQL queries written by novice users or generated by front-end interfaces --- SQL statements that might otherwise incur poor performance.

Additional features that enhance performance include: multiple buffer pools with selective pool assignment of tables and indexes; asynchronous page cleaners; a new Star-join technique. Availability is enhanced with online, sub-partition backup and restore and an online Redistribute Utility to compensate for skewing and to accommodate incremental growth of the database with uniform data distribution over all processors, including those newly acquired.

Support for OLAP and Complex Queries

This RDBMS's new SQL CUBE and ROLLUP statements perform multidimensional analysis operations. It is a leader in offering the advantages of *dynamic bit maps* over the static bit maps provided by several of its competitors:

- good for *both high and low* cardinality columns
- simpler database design
- less storage required
- no special compression (decompression) needed for sparse bit maps
- no complex index update processing

It also offers Tables Functions for retrieving and preparing nonrelational data for SQL processing and it provides extended SQL for *advanced recursive query* operations.

Ease of Administration and Use

Its Control Center provides a single point of control to manage multiple, local and remote DB2 databases, tables, views, indexes, users, UDTs, UDFs, triggers, disks, node groups, partitions and stored procedures (SPs). Smart Guides are included to assist in system configuration, performance tuning, database backup and restore, and replication setup. It includes integrated replication that supports both push and pull replication modes.

Finally, this system also offers a rich set of extenders for multimedia data to assist in building applications involving content management. They deserve close evaluation by those who might otherwise be intimidated by the prospect of developing their own code in support of these applications. There is a rich store of functions packaged with each of its extenders.

DB2 for AS/400 - the Integrated RDBMS

This is an advanced 64-bit RDBMS It runs on AS/400 based servers that offer *64-bit implementation of the operating system and its applications as well as its RDBMS.* Its unique architectural features are its single level store and the deep embedding of DBMS function in its operating system --- even below the machine interface level. For example its radix binary tree indexes are scanned by the system at the microcode level. Its single level storage architecture enables automatic striping of data across all available disk devices. DB2 Multisystem for AS/400 (an orderable feature) can scale up to 32 SMP nodes of 12 processors each; this configuration can support databases up to 16 terabytes.

Like other members of the DB2 Family, it offers all of the expected functions of the client/server paradigm with support for ODBC, X/Open CLI, triggers, and nonproprietary stored procedures (any AS/400 programming language as well as ANS compliant SPs consisting of only of SQL). It also offers the advantages of an advanced cost-based optimizer to perform query optimization, and it performs major processing functions in parallel, including parallel index build and data loading. In addition to nested loop and sort/merge join methods, it supports hash joins.

DB2 for As/400 Data Warehouse-Data Mart Support

DB2 for AS/400 includes the following capabilities that are especially useful for Business Intelligence processing:

- dynamic bitmaps
- query rewrite
- Remote Journaling
- Encoded Vector Indexes (EVIs)

Encoded Vector Indexes , introduced in this system's most recent announcement, is the industry's first vector approach to bit mapped indexes. This IBM patented index object provides high performance, scalability and reduced storage requirements, especially important for business intelligence applications. With this new release, this RDBMS now offers Scalar Subselect, the ability to update one table based on values from another table, ALIAS support, and Server Mode which allows SQL interface to support multiple users, multiple connections and multiple transactions from the same job. Multi-threading of SQL allows multiple SQL requests to run concurrently in a single transaction.

Remote Journaling enhances data availability by replicating data to a second AS/400. This capability is also of special interest for data warehouse applications. Data can be transparently moved from the operational database to journal receivers on an AS/400 target warehouse for the subsequent conditioning, *e.g.*, data conversion and transformation, needed for BI applications. This minimizes the impact of these essential data warehouse functions on operational database processing.

Also of interest for BI applications is its Predictive Governor to control query run times. Execution is controlled based on a comparison of predicted run times and a configurable allowable run time. More generally, its new DB Monitor and Index Advisor can be used for performance tuning. Index Advisor suggests specific columns for permanent indexing to enhance performance.

Seamless network connection is supported for TCP/IP, APPC and APPN. This RDBMS supports DRDA at the *distributed unit of work* level

This member of the DB2 Family has achieved impressive stress test results at the Rochester Teraplex Center. .

Information Management System (IMS): Database For the Space Age and Beyond

IMS celebrates its 30-year anniversary in the current year. With IMS V6, it also celebrates an important recent achievement in message processing for clustered systems --- an enhancement to IMS TM that provides increased capacity, incremental growth, automatic workload balancing (within a Sysplex), enhanced reliability, and increased availability. This new capability, IMS Shared Message Queues (SMQ), utilizes the Sysplex Coupling Facility for shared queues that service multiple IMS Transaction Managers. A message can be shared and processed by any IMS sharing the queues. Shared queues can be used to distribute work across the Sysplex. For IMS Fast Path users, the new Expedited Message Handler provides the same function with similar benefits to users.

This system has auspicious origins. When President Kennedy focused U. S. space efforts on a moon launch,. a joint team of IBM and North American Rockwell technical personnel developed a DBMS to generate large bills of materials for the construction of spacecraft and an automated logistics system. The result was the precursor of IMS, the first commercially available online DBMS. Today, IMS manages over 12 million GBs (12 thousand terabytes!) of production data worldwide, and it processes 9 billion transactions daily for customers throughout the world, including 90 per cent of the Fortune 1000 companies.

System Milestones

Soon after its introduction by IBM, IMS became the DBMS of choice in several industries that could benefit from even higher DBMS performance. The response from IMS developers was IMS Fast Path databases that emphasize high performance for Data Entry and Main Storage Databases (another IMS "first"). These IMS databases expedite processing for applications with high volumes of simple transactions and with high availability requirements. Data independence and programmer productivity were enhanced further with Field Level Sensitivity.

IMS Today

IMS exploits n-way data sharing in support of *parallel transaction processing*. Its customers can benefit from reduced computing costs, improved performance, and incremental growth up to 32 IMS subsystems and 32 operating system images (nodes) running in a S/390 Parallel Sysplex configuration. *This parallel processing support is available for Fast Path databases.*. IMS establishes a monitoring environment for the MVS Workload Manager, enabling dynamic adjustments to workload distribution.

System availability has been enhanced with Shared Message Queues (already noted) that enable any capable IMS subsystem to assume the processing load automatically on behalf of any other IMS subsystem that becomes unavailable. Extended Recovery Facility (XRF) continues to provide a hot standby capability for local failures. It has been complemented with V. 5's Remote Site Recovery (RSR), which tracks an active, *remote* site's operation for takeover should the active site fail. RSR can reduce recovery time to only minutes for *extended* planned and unplanned (disaster caused) outages. Both data and messages are recovered automatically. IMS V.6 introduced Fast Database Recovery for fast access to shared data managed by a failed system in a Sysplex environment; all committed data on that system can be accessed without the delay otherwise incurred to restart the failed system.

Today many installations require extended data sharing with distributed database support for their geographically dispersed users. IMS V.6 provides Distributed Sync Point --- transparent 2-phase commit processing to provide integrity-assured, synchronous commit of distributed resources.

Broadened Access to IMS Transactions and Data: The WWW and Objects

The set of IMS Connectors provide open access to IMS applications and data from Windows, AIX, Sun Solaris, OS/400, S/390 and OS/2 environments. IMS WWW Templates can be invoked from a Web server on any APPC platform. The IMS TCP/IP Open Transaction Manager Access (OTMA) Connection provides enhanced communication linkages between remote workstations and IMS. IMS Web supports a multi-tier structure for Web access; it includes both a GUI interface for development and a runtime environment for TCP/IP environments. IMS Client for Java templates are used to access IMS S/390 applications and can be used from any Java Virtual Machine. IMS Object Connector enables deployment of object oriented applications that access IMS DB data. It generates C++ classes that wrap IMS DB data segments, with no DL/I coding required. Use of this Connector provides code reuse and enhances programmer productivity.

The IMS Role in Business Intelligence

Industry analysts frequently remind us of the prodigious amount of operational data currently managed by IMS. This is a virtual storehouse of information of enormous business value for typical business intelligence applications such as customer service, supply chain management, fraud detection and target marketing, These are just a few examples of the wide range of applications through which IMS customers can acquire and sustain a competitive edge. IMS has clearly opened avenues to its rich information sources from diverse environments using many of the key capabilities of the WWW. Note also that IMS data is readily accessible for transformation and delivery to DB2 managed data warehouses by IBM's replication tools, or as a data source for IBM's Visual Warehouse and for Data Joiner. Decision support applications can exploit IMS data with the aid of DataJoiner or IMS gateways provided by other vendors. Also IMS and DB2 data can be accessed from a single application using both DL/I and the SQL database APIs.

Into the Next Millennium

Throughout its 30-year history, IMS has enhanced this system with those features that have made it the leader for high performance, enterprise-level OLTP applications supported with hierarchic databases. Its success, according to its chief architect, Vern Watts, "is largely attributable to its responsiveness to real customer needs". The IMS users can reasonably anticipate further enhancements to its capacity, levels of performance and availability as well as more improved system management facilities. IMS's product management also anticipates even stronger support for object oriented applications in future versions, along with continued emphasis on features that facilitate Web access to IMS data. This DBMS is ready for the year 2000 and will benefit its users well into the next millenium.

The Intelligent Approach to Business Intelligence

The business value of business intelligence solutions is perhaps best demonstrated by the dramatic year-2000 revenue projections that industry analysts provide for data mart, data warehouse and data mining solution providers. In addition, recent studies show remarkable ROI measures for those who successfully deploy and exploit these solutions for their broad business applicability and enormous potential business value. Success stories abound in the areas of customer relationship management, supply chain management, fraud detection, risk management, and market and financial analysis. IBM is counted among the leaders in this arena by the same industry analysts. IBM's BI solutions, including the DBMS support of its DB2 Family, also provide a comprehensive infrastructure for those seeking the advantages of electronic commerce. Moreover IBM and its customers have a distinct advantage in deploying these solutions. Much of the operational data that "sources" BI operations is already managed by IBM data management facilities. The opportunities to benefit from the synergy among them should not be overlooked.

Visual Warehouse: The Integrated BI Solution

Despite the well publicized successes in this area, IBM has recognized the prolonged, complex and costly efforts often required for the successful deployment of these solutions and has responded with Visual Warehouse (VW), an integrated packaged solution that enables cost effective, rapid deployment and focused administration of the necessary data mart and data warehouse operations including: source data extraction from multi-vendor data sources; automatic data extract, transformation and cleansing based on business views defined with VW's graphical query builder; population of the target warehouse; user defined refresh cycles; automatic scheduling and monitoring of warehouse tasks; and the full capabilities of IBM's Data Guide as an information catalog that performs metadata interchange.

VW's agent technology provides scalability from one or more data marts to enterprise level data warehouses --- from LAN based to distributed warehouse environments. It can be used to deploy warehouses and data marts with operational data sources managed by the DB2 Family, by other vendor DBMSs, and by non-relational DBMSs as well as conventional file systems.

In addition VW provides seamless integration with Arbor Essbase for its best-of-breed, online analytical processing (OLAP) capabilities and with ETI*Extract for the wide range of data sources and sophisticated data transformation capabilities it offers. The Essbase APIs enable the use of a extremely wide range of front-end query and reporting tools.

DB2 OLAP Sever

An alternative VW offering includes DB2 OLAP Server (also available separately). This VW offering provides the best of two worlds of OLAP: *multidimensional OLAP* (MOLAP) and *relational OLAP* (ROLAP). DB2 OLAP Server offers the computational and analytic capabilities of Arbor Essbase, Version 5 for OLAP processing against DB2 star-schema databases. It provides:

- Accessibility using all Essbase ready tools as well as standard SQL
- Multi-user read/write capability
- Automatic star schema creation, with indexing
- Essbase analytic calculations
- Dynamic and precomputed calculations
- Essbase APIs for seamless access to multidimensional data by many ISV desktop tools and applications

Thus DB2 OLAP Server combines the rich multidimensional analysis functions of Essbase with the scalability, reliability and manageability of DB2 and SQL based access.

Information Discovery: Mining Data Intelligently

The recent announcements of the Intelligent Miner (IM) Family offerings represent impressive enhancements to what were already among the market's leading tools for information discovery: a new version of IM for data and Intelligent Miner for Text (MIT). The former includes enhancements to support the full range of mining processes ---from data preparation and analysis up to and including the assimilation of mining results. More specifically it provides::

- Nine innovative mining algorithms
- Automation of typical data preparation tasks
- Multiple visualizers
- Administrative Java-based user interface
- Repeatable and reusable sequences
- Increased parallelism of algorithms
- Task Guides and a mining Progress Indicator

IM for Text, a new IM Family offering, supports mining of the vast, previously untapped bodies of text, *e.g.*, e-mail, insurance claims, news feeds, customer correspondence, and Web pages --- all to derive business intelligence that is usually unavailable. It includes Text Analysis tools to categorize, and cluster documents and to extract significant lexical features. Text Miner, an advanced search engine, performs in-depth text analysis, advanced relevance ranking, hybrid Boolean and free text searches, fuzzy searches, and thesaurus support. GUI access is provided from any Java V1.1 capable browser to any interesting Internet location. IM also provides Web Tools for building global WWW or Internet search services.

Simplifying the Complex: DB2 DataJoiner (DJ)

Industry analysts and surveys report that many organizations currently use several multi-vendor DBMSs. Attempts to derive enterprise-level information from these diverse data sources are often hampered or totally inhibited by cost and other resouce constraints. DB2 DataJoiner (DJ), IBM's

multi-database server represents innovative technology that can support the production of highly useful business intelligence with substantial cost savings from the following diverse sources: DB2, Oracle, Microsoft SQL Server, Sybase and Informix databases, as well as non-relational data sources like IMS and VSAM. It provides a single site image of these heterogeneous systems and a global database view for both retrieval and update of this "live" (current) data. It also offers object/relational database processing on this data. Users employ a single SQL dialect that DJ globally optimizes and maps to the particular SQL of these multiple backend DBMSs, with support for *distributed unit of work* and *distributed request* (including distributed join) processing. Studies show that DJ's global optimizer sometimes provides better query performance than *direct* native access to its data sources. Web gateways and application tools can be used to access DJ's data sources in Internet environments.

Looking Ahead

Forecasting the long term direction that the software industry will take can be risky business. The infrastructure for the Internet and WWW was developed many years ago, but who could have predicted at that time the great changes (including the development of global market dynamics) it would precipitate for enterprises of all sizes. A *sound business approach to the future* looks for a solid a record of achievement from vendors who offer solutions with a foundation that will endure, provide room to grow and to innovate, and most importantly, to preserve and exercise strategic options in the implementation of data management solutions. Based on the solutions IBM offers today and the record of achievement documented in this paper, IBM customers appear to enjoy these advantages.

Don Haderle has indicated some of the directions and goals IBM is pursuing for both the near and longer term future. First is support for the concept of what he calls an "overall business transaction view." It provides *comprehensive* DBMS support for transactions that span different business operations, at disparate locations and within a mix of operating environments. Don provides the following example:

> Consider the purchase of a new automobile. Viewed at a global level, it involves the dealer, the dealer's manufacturer, the manufacturer's suppliers, the dealer's lending organization and others. They are all participants in such transactions. We need to preserve the integrity of all of the data involved in this global transaction --- *when it succeeds and when any component fails. The challenge is doing this online, in real-time for a wide range of business applications and data environments.*

Haderle points out that this is what IBM's DataLink technology is all about. This technology, expected from IBM in the near future, constitutes an important step in the direction of the *complete* DBMS support for the global transactions that Haderle describes
.

We can reasonably expect further enhancements to IBM's current object/relational capabilities. DB2 for AS/400 has already expressed its intent to achieve parity with the current *Universal* DB2 Family members. In addition IBM recently announced a new spatial extender that supports the retrieval and manipulation of geographic data. The ability to combine coordinate searches with other demographic data can be highly useful in BI applications like target marketing or determining the optimal locations for franchised operations.

To simplify and increase the efficiency of data mining functions, we may soon see their integration with members of the DB2 Family. As Don Haderle has noted, providing more function neatly integrated

within the framework of SQL-based DBMSs is the key to continued user productivity, performance and manageability.

Concluding Remarks

Support for enterprise wide data management requires robust, reliable and extensible DBMS capabilities. Current, as well as prospective users, should consider DBMSs and BI solutions that offer the required levels of performance, availability, scalability, manageability and costs of ownership. IBM's DB2 Family, IMS and BI solutions offer capabilities needed for enterprise level support. These criteria are key to the success of operational and informational applications. In addition *these criteria should be coupled with consideration of the durability and support that a vendor* can deliver. Finally the adoption of DB2 Universal Database systems throughout an enterprise offers the benefits of leveraging the common advanced functions of these database management systems, along with the common DB2 skills of their staffs to achieve cost economies along with increased productivity. As the nucleus of data management solutions, these offerings provide a solid foundation for moving successfully to the year 2000 and beyond.

Acknowledgments

The author gratefully acknowledges the assistance of several IBM staff members and others who assisted in the preparation of this paper by providing source information and review comments:

Nagraj Alur, (SantaTeresa), Mark Anderson (Rochester), Don Chamberlin (Almaden), Chris Date, Carolyn Dow (Chicago), Susan Grey (Santa Teresa), Don Haderle (Santa Teresa), Barbara Klein (Santa Teresa), Howard Katz (Somers), Roger Miller (Santa Teresa), Janet Perna (Somers), Bob Picciano (Santa Teresa), Pat Selinger (Almaden), Paul Sembrat (Somers), Mike Swift (Santa Teresa) and Pat Walsh (Waltham, Mass).