The Right Database for Your Growing Business

Ndo M. Osias
ndo_osias@hotmail.com

Abstract

As a business grows there is a need to upgrade both the hardware and software that form the company's information infrastructure. In today's information society, the role and importance of data for a business are crucial and the choice of an appropriate database to store data is the key to successful use of information. This paper will discuss various database technologies currently available based on the complexity of data. Strategies for choosing which vendor to use based on cost, time for implementation and future growth potential will also be addressed. The influence of the Internet on business will be approached from a long-term perspective. Specific technologies to be discussed include Microsoft Access, Microsoft SQL Server 7, Sybase Adaptive Server Enterprise 12, and Oracle 8i.

I. Introduction

In the last two decades of the twentieth century, tremendous amounts of information have been created and stored by organizations. Because of the importance of accurate information to businesses, there is a need for centralized management and control of data. For years, the method of electronic data storage consisted of filing systems. The file processing model comprised separate computer programs that were used to update independent data files which were used to produce reports required by each corporate application. The main weaknesses of the file processing model were the redundancy of data resulting from the fact that each department within a company maintained its own set of files and the inconsistency due to clerical errors. Additionally, the lack of centralized data repositories duplicated effort. To resolve problems of traditional file-based systems, the database approach was developed.

I.1. The Database Approach

A *database* is a collection of records accessible by various applications that use the data contained in it. It is a collection of data organized to serve many applications efficiently by centralizing the data and minimizing redundant data. Such a system is far easier to administer and helps to eliminate the problems of the traditional file-based system. A *database management system* (DBMS) is the software that manages database memory structures, creates and maintains a database and enables business applications to extract the data they need without having to create separate files or data definitions in their computer programs.

The database approach has several advantages to business environments.
1. **Data Consistency:** Enterprise data is stored in a central location and allows everyone in the company to see identical information at it was stored at a particular point in time.
2. **Restricted Data Access:** Users in different departments can be given different roles and privileges liked to their accounts and thus restrict access to sensitive company information such as financial data and salaries to authorized employees.
3. **Concurrent Access to Current Data:** The DBMS allows all users connected to the database to access the same current information concurrently; this results in increased productivity because employees do not need to wait for the data.
4. **Backup and Recovery:** The DBMS provides the means to back up data as often as needed and allows the database to be restored to its original state before failure.

II. Classification and Architecture of DBMS

Database management systems are classified on the basis on their data model. The most common types of DBMS commercially available today are:

1. **Relational DBMS:** The database is represented by of collection of tables. Data in tables is organized in rows and columns. Each row in the table represents a collection of related data values. Column names define the data values in each row and all values in a column have exactly the same data type.
2. **Object-Oriented DBMS:** The database is defined in terms of objects, their properties (or attributes), and their operations (or methods).
3. **Object-relational DBMS:** This data model is a combination of the relational and object-oriented models. The relational data model is extended with user-defined types and functions that model business objects and encapsulate business rules with data.

Most databases today use the relational model. The language used to manipulate data in relational databases is Structured Query Language (SQL). There are several ways to implement the relational data model.

1. **Process-per-client:** Each database client has its own process address space. The database runs in one or more separate background processes. The advantages of this method include the fact that it provides a protected environment where users do not interfere with each other and the database manager is protected from users. Processes can be easily assigned to different processors on a symmetric multi-processing (SMP) machine. However, with large number of users performance is poor because of the higher requirement for processor cycles and memory for each client.
2. **Multithreaded:** The database, the applications, and the user connections run in the same address space without using the operating system for scheduling. This approach uses less memory and CPU cycles than the process-per-client architecture. It also functions well with large number of users to perform short transactions. On the other hand, the whole DBMS can crash as a result of an application failure and long queries perform poorly because of their consumption of system resources.
3. **Hybrid:** This architecture comprises multithreaded network listeners that assign each client to a dispatcher which in turn places messages on a message queue. Message responses are then removed from the queue and sent back to the requesting client. Reusable shared server worker processes pick the work request from the work queue, execute it, and place the response on the outbound queue. The advantage is a protected environment that does not require assigning permanent processes to each user. Queue latencies and resource contention are potential problems with hybrid databases.

III. The Major Database Management Systems

According to Dataquest, 1997 revenues for the SQL database market were $6.6 billion with Oracle leading the pack at $1.789 billion, IBM at $1.775 billion, Microsoft at $997 million and Sybase at $297 million. Generally speaking, with the exception of Access which has limited provisions for concurrent users, the major DBMS have similar features to support business solutions. The main difference being that SQL Server 7 is only available on the Windows OS while Oracle 8i, DB2 Universal Database version 6, and Sybase Adaptive Server Enterprise 12 can be implemented on a variety of operating systems. The other DBMS also have the reputation of being more stable and robust than SQL Server is. The Internet is changing today's business environment. The trend is for companies to implement large-scale e-commerce sites and Business-to-Business Internet applications. It is important to consider the

Internet support provided by each DBMS. The main features of each database are discussed in this section.

1. Oracle

Oracle8i is the latest version of Oracle Corporation's DBMS and was released in 1999. It is an object-relational DBMS that can store text, documents in various formats such as Word or PDF, audio data, video data, images, and geographic location information. Oracle 8i is available on more than 90 platforms including about 60 versions of Unix, Windows 95/98/NT/2000. Oracle 8i's environment is consistent across platforms in terms of user and administrative interfaces, tools, and the database engine itself. Centralized administration of multiple databases is possible with Oracle Enterprise Manager (OEM). Advanced Replication is a sophisticated mechanism that allows bi-directional data replication in a distributed Oracle environment; multiple copies of one or more entire database or subsets of the database can be written to one or more target databases. Oracle 8i comes equipped with interMedia, a text and multimedia search engine. Oracle Lite is a version of the database with a small footprint for laptops and handheld devices. It can be synchronized with a larger database for partial replication. The Oracle Parallel Server configuration consists of multiple instances (Oracle memory and background processes) accessing the same database in parallel. Internet File Server (iFS), due to be released in production version during the second half of 2000, allows the storage of non-database files in any format to be dragged and dropped into the Internet File Server and stored inside the database. Oracle 8i supports Internet commerce applications with database support for Java in the form of JServer, a built-in Java Virtual Machine (JVM), and Web content management with an Extensible Markup Language (XML) parser for retrieval of XML documents stored within the database. Quality comes at a price, and Oracle has always being known for its steep prices. However, at the end of 1999, Oracle discounted its prices for small businesses. Oracle 8i Enterprise Edition costs $100 per power unit and Oracle 8i Standard Edition costs $15 per power unit. 24x7 support services packages are 7% of the software sale costs.

2. IBM

IBM's most recent DBMS is *DB2 Universal Database 6.1 (UDB v.6)*. UDB is an object-relational that can store text, documents, audio, video, images, and large binary objects (LOBs). UDB is available on many flavors of Unix (AIX, Solaris, HP-UX) and Windows 95/98/NT/2000. Centralized database administration tasks can be performed with Control Center. Similar to Oracle's iFs, DB2 offers external contents management with DataLinks. Mobile computing database support is provided to DB2 users through DB2 Satellite Edition; compared to its rivals Oracle Lite and Sybase Adaptive Server Anywhere, IBM DB2 Satellite Edition lacks solid administration tools. UDB also supports Web technologies through the use of extenders inside the database. These extenders can be used specifically for non-relational data such as XML, text, audio, video, and voice. IBM DB2 Enterprise Edition costs $12,500 per server CPU with an unlimited number of users and $199 per client.

3. Microsoft

SQL Server 7.0: Microsoft's high end DBMS is *SQL Server 7*. It has a suite of GUI management tools, the SQL Enterprise Manager, which provides centralized management. It is only available for Microsoft Windows NT, 95, 98, and 2000 and does not port to other platforms. It has provisions for data warehousing, OLAP, and replication.

It allows online backup, row-level locking, a maximum of 32 tables per join, and a maximum number of processors. It supports very large databases. Object Linking and Embedding (OLE) database support is available through Open Database Connectivity (ODBC) which allows connection to Oracle, DB2, Excel and Sybase data sources. XML is also supported. The parallel query execution option allows for faster queries. There is also distributed querying for accessing data from multiple sources. There is clustering and fail-over support for NT clusters. There are also provisions for connecting over the Internet with TCP/IP sockets or multiprotocol Net-Library. Microsoft sometimes offers promotional discounts of

more than 50% for current Access users. For non-Access users, SQL Server Standard Edition starts at $1,399 for five client licenses and goes to $3,999 for 25 client licenses. SQL Server Enterprise Edition starts with 25 client licenses at $7,999 and goes to $28,999 for 250 client licenses.

Microsoft Access: This DBMS is best suited for small, standalone applications. It comes with a prototyping tool, database, visual input screen generator, and report writer. A single file (MDB) stores both the application and data. It allows connection with other databases via ODBC (various data sources). It allows for data exports in various formats. Access has some drawbacks. It does not have provisions for transaction processing (commits, rollbacks), therefore the database could be out of sync. As a single file database, it does not allow multiple programmers to work concurrently on the same application. It does not allow online backup and recovery. It does not come with a procedural language like Oracle's PL/SQL. Access Basic is needed for complex data manipulation. It is interpreted and thus slows the application down because it is resource intensive.

Access is optimized for single-user desktop applications. Multi-user inquiry applications require that the MDB file be placed on a networked drive. Prototyping applications to be later ported to a more robust DBMS can also be developed using Access. Applications to avoid with Access include multi-user applications with sophisticated updates, large-scale applications developed by more than one programmer since there is no version control mechanism. High-volume applications cannot be implemented due to physical limitations. The maximum file size is 1 Gigabyte and the maximum number of files limited to 255. True client/server applications cannot be developed on Access due to its design for single machine storage and data handling

4. Sybase

The latest DBMS offering from Sybase Adaptive Server Enterprise (ASE) 12. Adaptive Server Enterprise 12 is available on several Unix platforms (AIX, HP-UX, Solaris) and Windows NT. Like competing DBMS, Sybase offers a GUI tool for database management, Sybase Central. Sybase ASE supports data warehousing; replication is offered through its Replication Server which is a separate database system for handling all aspects of replication. Sybase also offers Adaptive Server Anywhere, a small footprint database engine similar to Oracle Lite, for mobile computing. Lack of built-in database performance monitoring tools is a weak point of Sybase DBMS. To support electronic commerce applications, ASE version 12 has a built-in Java Virtual Machine (JVM) to allow database developers to write stored procedures using both Java and SQL. Like Oracle 8i, ASE version 12 allows XML documents to be stored, managed, indexed, and retrieved. Costs are as follows: $3,995 for the server, $1,430 per seat for concurrent users, and $795 per seat for network users.

IV. Database Management System Selection Criteria

The selection of a database management system is an important decision for any growing business. Although the preeminent consideration for all selection decisions is cost there are other important considerations centered around projected company growth, type of data, use of custom software versus manufacturer supplied products. Before choosing a DBMS the following issues need to be considered:

1. How much will the whole system cost? Cost is one of the major factors that ultimately determine which database solution will be implemented. The technical capabilities need to be balanced with the total costs of implementation and ownership. Cost factors to consider are as follows:
   - **Database License:** Some DBMS vendors use the cost per server, the cost per client, or cost per user pricing models or a combination thereof.
   - **Tools**: Some vendors ship their DBMS with integrated database administration tools, database design tools and database reporting tools. However, before buying a solution that does not include

some of these tools, the cost of in-house development of these facilities needs to be weighed against the costs of purchase and integration of third-party tools.

- **Implementation and Technical Support Teams**: These teams could include the following based on the size of the implementation: data architects, database administrators, business analysts, network administrators, system administrators, and database developers. Currently, hourly billing rates for members of these teams can range from $40 to about $100 depending on experience and role.
- **Training**: Training can take several forms: classroom training, computer-based training and internet/intranet-based training. Depending on the complexity of the DBMS and the role of the employee being trained (DBA vs. developer) up to five weeks of training could be needed.
- **Vendor Technical Support**: Complex issues related to the functioning of the DBMS sometimes require the kind of expertise that only the DBMS vendor can provide. Only businesses with a support contract can use this service.

2. What kind of data will be stored in the databases? For storage of alphanumeric information, a relational DBMS will suffice whereas multimedia information will require an object-relational DBMS.
3. How will the applications be developed? If applications will be developed in-house, the development team must determine if additional software will need to be purchased.
4. How mission-critical will the applications be? For applications requiring 24x7 availability or limited downtime a DBMS that supports high availability solutions like standby databases or replication will need to be considered.
5. Where will the users be located? For a company with several offices in geographically diverse locations, a distributed database implementation may be required.

V. Conclusion

The choice of a database management system is a complex decision based on technical and financial factors. This paper provided background information on the four most widely used DBMS and their main features. Before implementing a database solution, businesses need to define what information is going to be stored in the database, who is going to access that data, how long the data needs to be continuously accessed, where the of the users will be physically located, and what tools are necessary to access the information. Once these questions have been answered, the technical specifications and costs of the various DBMS can be used to rank their suitability for the project.

Bibliography
Ace The Technical Job, Database Edition, M. Rothstein, J. Glyck, D. Gilor, D. Matthews, D. Cohen, McGraw-Hill, 2000
Client/Server Survival Guide, 3rd Edition, R. Orfali, R, Harkey, J. Edwards, John Wiley and Sons, 1999
Management Information Systems, 6th Edition, Organization and Technologies in the Networked Enterprise. K. C. Landon, J. P. Landon, Prentice Hall, 2000
Oracle 24x7 Tips & Techniques, V. Devraj, Osborne/McGraw-Hill, 2000
Practical Microsoft SQL Server 7, B. McGehee, R. Kraft, QUE, 1999

About The Author
Mr. Osias is a database administrator who has been implementing databases for over four years. He currently works as a senior database administrator for a multinational chemical company in Central New Jersey where he is responsible for the administration of a large SAP installation. He is a graduate of Worcester Polytechnic Institute in Worcester, Massachusetts.