SEPTEMBER 1987

DISTRIBUTED DATA BASE MANAGEMENT:

AN EARLY LOOK

INPUT®

1280 Villa Street, Mountain View, California 94041-1194 (415) 961-3300

Published by
INPUT
1280 Villa Street
Mountain View, CA 94041-1194
U.S.A.

Information Systems Program (ISP)

Distributed Data Base Management: An Early Look

Abstract

As the general trend toward distributed data processing has evolved, the most significant factor inhibiting the development of distributed but integrated systems has been the availability of effective software for the management of distributed data base environments. Until just recently, the unavailability of such software has limited the design of distributed applications significantly.

In 1987 Data Base Management Systems (DBMS) vendors have begun to provide data base technology that supports the physical distribution and management of logically integrated data bases. Furthermore, the software systems being offered provide this capability across multiple hardware and operating systems environments with gateways to more traditional relational and hierarchical data base software packages.

This report provides an early look at this emerging technology, examines its current state of development, and gives INPUT's analysis of the implications and impacts it might have for IS management. The report also examines current user plans for application of Distributed Data Base Management Systems (DDBMS) and provides guidelines for successful implementation.

This report includes 65 pages and 26 exhibits.

Table of Contents

Table of Contents (Continued)

Exhibits

I

Introduction

I

Introduction

A

Purpose

Over the past few years INPUT has monitored and reported on the evolution of Data Base Management and Distributed Processing Technology. In 1985, 1986, and 1987 INPUT published reports analyzing the impact of distributed technology on the information services market and user applications.

- The 1986 report – **Departmental Computing & Software Direction –** and previous reports predicted and described the growth in distributed/ departmental computing.
- The 1987 report – **Future Data Base Markets, 1987-1982 –** announces the arrival of the first commercially available "distributed" data base management system (DDBMS) software.

This report looks at this new distributed data base technology from the information systems organization's point of view and addresses such issues as:

- Definition
- Maturity
- Relationship to existing DBMS technology
- Potential application areas
- Management issues

The goal of this report is to prepare the reader to understand and consider applying distributed data base technology.

B Scope

As the report will point out, distributed data base and the related software technology (referred to as DDBMS) is in the early phases of its product life cycle. There are commercially available products and some early application efforts, but for the most part there is just a curiosity of how this new software technology can be productively applied within an overall information systems strategy. That is, how does DDBMS fit in when viewed within the context of the principle information systems trends?

1. Related Information Systems Trends

There are a number of trends that are creating a need for and will soon be driving the use of distributed data base technology.

- There is a constantly growing and large installed base of distributed processing computers.
  - Distributed processing is a common and often major element of IS strategies.
  - The population and combined processing power of mini and personal computers is far outstripping that of the central mainframes in many organizations.
  - Significant amounts of data are being routinely transferred out of central data bases to computers that are beyond the desired "data administration control" of IS.
  - Significant amounts of processing power and storage are being consumed to transfer and store this duplicated data.
  - The so-called "islands of information" problem has become common place.
  - IS is beginning to look for alternative means to "re-integrate." DDBMS offers such a means.

- The explosion and maturing of end-user computing has caused this phenomenon to take on new characteristics.
  - Minicomputers and multiple user environments are supplementing PCs.
  - Departmental production systems are replacing, and being added to, adhoc/decision support applications, and they are being developed by the user.

- The user population is more experienced, mature, and confident, and is becoming more active in the IS decision process.

- Responding to the departmental systems challenge while striving for improved data management will cause IS to consider DDBMS.

• The coming of age of relational data base technology has offered an easier-to-use DBMS.

  - IBM has placed its stamp of approval on the relational model and expanded the market.

  - Relational DBMSs now exist with PC-like interfaces and portability across computing environments (mainframe to mini to PC).

  - The end user is beginning to understand data base through the relational concept.

  - An explosion in relational-based application development is now underway.

  - The relational model makes DDBMS possible, and thus it will become a logical step for many IS organizations.

These trends create an environment and the opportunity for distributed data base technology to be successfully applied.

2. Questions Addressed

This report addresses the following questions:

- What is a distributed data base?
- What are the components of a distributed DBMS system?
- What is the current status of DDBMS offerings?
- Where is this technology headed?
- What are the issues facing the user of DDBMS?
- What are the characteristics of the early applications?
- When and why will DDBMS capabilities be used?

This report does not address such issues as supporting required telecommunications networks or the economics of distributed data base systems. Nor does it predict when a fully capable DDBMS will be available from one or more vendors. However, it does provide the reader with the knowledge required to more easily investigate DDBMS technology and its application.

3. Report Organization

This report is organized as follows:

- Chapter II is an Executive Summary.
- Chapter III provides a definition of distributed data base, its components, and its variations.
- Chapter IV provides a look at the state of DDBMS technology and a summary of products and plans of key DBMS vendors.
- Chapter V looks at the issues, both technical and operational, the elements of success, and early efforts.
- Chapter VI provides INPUT's conclusions and recommendations.

C Research Methodology

While a great deal has been written on the subject of distributed data bases over the past few years, there has been little, if any, experience until very recently. In the preparation of this report INPUT has:

- Reviewed previous INPUT reports on data base and related subjects. INPUT believes that DDBMS must be viewed as yet another "step" in the evolution of computing and data management technology.
- Reviewed recent industry and vendor publications which along with previous INPUT research provided a basis for vendor and IS executive interviews.
- Interviewed DBMS vendors to learn what their customers are demanding and what the vendor is doing to provide DBMS capabilities. Both leading edge (e.g., Oracle and Relational Technology) and long-standing (e.g., ADR and Cullinet) companies were included.
- Interviewed a cross-section of information systems managers to learn who was investigating and/or implementing DDBMS technology and how it was being applied.

D

Related INPUT Reports

It is expected that INPUT will continue to monitor the DDBMS evolution.

The following recent INPUT reports are pertinent to this subject:

- **Future Data Base Markets, 1987-1992,** 1987.
- **Departmental Systems and Software Directions,** 1986.
- **Distributed Data Processing,** 1985.
- **Fourth Generation Languages Update - Potential Unrealized,** 1985.
- **DDS - Experience and Outlook,** 1985

II

Executive Summary

II

Executive Summary

This chapter summarizes the report that follows and provides an overview of a new software capability, distributed data base management.

A

Distributed Data & IS Trends

The concept of distributed data base needs to be viewed within the context of the key information systems trends. There are three current IS trends that directly relate to distributed data base and the use of distributed DBMS software. Those trends are (as shown in Exhibit II-1):

1. Distributed Processing:

The ever-growing use of distributed processing capabilities has put in place significant computing power, fostered a growing decentralization movement, and created new problems for IS (in particular the "islands of information problem"). IS is beginning to see a strong need to "re-integrate" its distributed information structure.

2. End-User Computing:

The computing phenomenon of the 1980s has become a mature and experienced force in IS decision making. The end user is demanding more function and power and is building his own production systems. The distributed strategy has become a departmental strategy creating additional data control pressures.

3. Relational DBMS Technology:

The move to relational DBMS systems is now well underway with use becoming common in distributed processing systems. Relational DBMS systems are being provided with easy-to-use application tools which are attracting the end user's attention. The relational model is the basis for DDBMS.

EXHIBIT II-1

DISTRIBUTED DATA BASE & IS TRENDS

Distributed Data Processing
- Commonplace
- Growing quickly
- Islands of information problem
- Re-integration of data requirement

End-User Computing
- Departmental systems explosion
- Mature, experienced end user
- Is losing application control

Relational DBMS Technology
- Commonplace
- Easy-to-use PC-like interface
- Production system use
- Basis of DDBMS technology

B Distributed Data Base & IS Strategy

Those IS organizations that are following a distributed processing strategy will want to look very carefully at distributed DBMS technology. It offers the chance to:

- Make data truly corporate and available to all without extensive redundancy.
- Separate the type of computing technology from data use.
- Balance end-user freedom and IS control.

Applying DDBMS technology to a distributed/departmental systems strategy can increase the benefits gained and lead to a more balanced environment.

Early efforts are best focused on distributed systems using mini and personal computers and a single DBMS technology (homogeneous versus heterogeneous environment). Early efforts should also emphasize learning to use relational technology. Developers have to learn to think relationally.

DDBMS will find a productive place in most IS strategies and play a primary role in those strategies that emphasize distributed systems.

EXHIBIT II-2

DISTRIBUTED DATA BASE & IS STRATEGY

Strategy Implications

- Make corporate data available to all
- Separate computing environment from use of data
- Balance end-user computing freedom & IS control responsibilities

Early Efforts

- Distributed/departmental systems
- Mini & PC versus mainframe
- Relational versus nonrelational

C

Distributed Data Base - A Definition

Distributed Data Base draws upon the capabilities of distributed processing and data base management systems technology. In simple terms, the data base for an application is spread across more than one computer but the application runs as if it were on a single computer.

The goals of the distributed data base concept are to spread processing loads of an application across smaller, less expensive computers; bring the application closer to dispersed users; improve control over an organization's data; and provide the local user with local autonomy over local processing.

Exhibit II-3 describes the distributed data base concept from both the IS and end-user points of view. They are quite different.

- The IS user looks at the subject from a control and technical point of view (what is the DBMS doing with the data).
- The end user looks at the subject from the point of view of ease of access to the data.

The best way to look at (define) distributed data base is as a collection of data bases on interconnected computers that are working together in a defined "partnership."

- The individual DBMSs on each computer manage the data relationship of the individual local partners.

- The distributed DBMS manages the data relationships that go across partners. Think of the distributed component as the "administrative" partner.

EXHIBIT II-3

DISTRIBUTED DATA BASE - A DEFINITION

INFORMATION SYSTEMS

A data element:

- Is stored a single time
- Is within a network of computers
- Is accessible for inquiry and update from any of the computers
- Is in conjunction with data elements on the other computers

END USER

The end user:

- Is unaware of where the data element is stored
- Is not required to do anything special to access the data in the desired relationships

DEFINITION

Partnership

- Collection of partners
- Distributed capability is the administrative partner

D

Distributed Date Base - An Example

Exhibit II-4 shows a distributed data base application that supports order entry, shipping, and invoicing.

- The data is distributed across three computers in three cities and linked by an on-line network.

- The order entry, shipping, and invoicing functions are performed by the users at local sites.

The application processing takes place at the local sites drawing on local and distributed data.

- When an order is entered in Los Angeles, the order entry transaction automatically queries the Chicago customer data base requesting the necessary customer information.

- When the order is accepted in Los Angeles a shipping instruction transaction is automatically sent to the Reno computer where the inventory and pick list data segments of the data base reside.

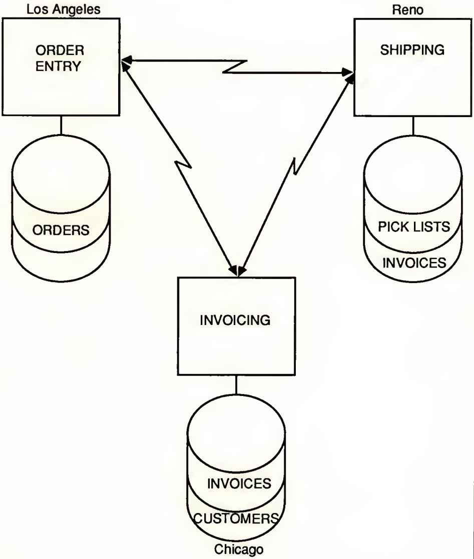
The data is distributed, but the application is integrated.

- The user is not aware of where the data is located.

- The data required to permit some, if not all, of the local site's processing is stored locally so processing can continue even if some part of the network is down.

EXHIBIT II-4

DISTRIBUTED DATA BASE - AN EXAMPLE



E

DDBMS Technology - Vendor Status

INPUT's research identified five DBMS vendors who have announced and are delivering distributed DBMS (DDBMS) software.

- Three are new, leading edge, relational DBMS vendors: Oracle, Relational Technology, Sybase.
- One is a longstanding DBMS vendor: Applied Data Research.
- One is a hardware/software vendor: Tandem.

The other vendors researched, primarily longstanding DBMS vendors, all indicated distributed capabilities were in development. Availability was 1988 or beyond.

The existing products support the partnership definition as do the strategies of the developing companies. This is particularly true for the three newer DBMS companies. Their products are portable, run under multiple environments (mainframe, mini, PC), provide local autonomy, and are integrated with their own development tools.

- They are based on the relational model
- They utilize SQL.
- They include active, integrated data dictionaries.
- They do, or will, include gateways to other DBMS technologies.

DDBMS systems are available today with adequate capabilities to support production applicatons. Exhibit II-5 provides a summary of key vendors and their current offerings.

EXHIBIT II-5

DBMS TECHNOLOGY - VENDOR STATUS

VENDOR	RELATIONAL	DISTRIBUTED
ADR	Datacom	D/NET
Cincom	Supra	
Cullinet	IDMS/R	
DEC	R/db	
IBM	DB2	
Informix	Informix-SQL	
Oracle	Oracle	SQL*Star
Relational Technology	INGRES	Ingres/Star
Software AG	Adabas	
Sybase	Sybase	Sybase
Tandem	NonStop SQL	NonStop SQL

F
DDBMS Components

Exhibit II-6 identifies the components of a distributed data base management system as implemented by Oracle and Relational Technology. The components are implemented at four levels:

Level 1 - The query processing and active data dictionary provide the foundation or first level.

Level 2 - The linkage between the local DBMSs makes up level 2. This level comes into play when different (heterogeneous) DBMSs are in use where a gateway provides the interface.

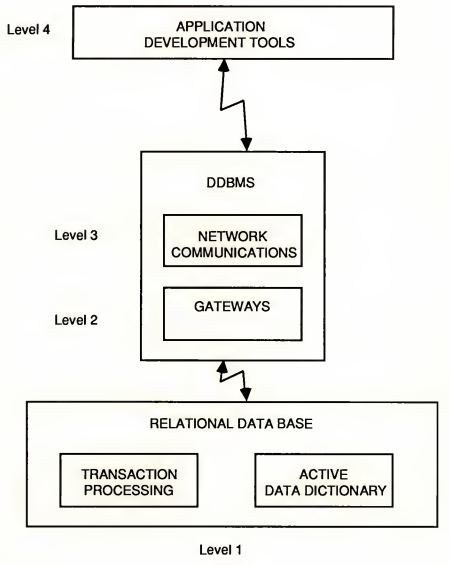
Level 3 - The third level is the network module that handles the communications between platforms. The network module runs on each of the computers in the network.

Level 4 - The top level contains the application development tools and the application itself.

An active dictionary exists at each node and manages the local and distributed data relationships for that node.

EXHIBIT II-6

DDBMS - THE COMPONENTS



G

User Issues - What to Look For

DDBMS, like any new computing technology, bring new and complex challenges. The issues identified cover technical challenges, IS management, and the end user.

Exhibit II-7 lists the issues identified by INPUT in this study.

- The technical issues include: data administration, data communications, processing performance, security and recovery, and the more recent issue of systems integration.

- The IS management issues center on being sure the organization is ready: Are the data control procedures in place, is relational DBMS technology understood and in use, and is the distributed/departmental systems program well structured?

- The end-user issues relate to this new computing professional's growing level of experience and desire for independence. DDBMS may be viewed as a roadblock to an end user's departmental objectives.

The issues are not insurmountable, and many are simply extensions of longstanding challenges. However, DDBMS does add a significant element to the level of complexity.

EXHIBIT II-7

DDBMS - IMPLEMENTATION ISSUES

Technical Issues

- Data administration exposures
- Data communications stability
- Processing performance impacts
- Security/recovery exposures
- Systems integration options

IS Management Issues

- Control over corporate data
- Status of relational DBMS technology
- Management of distributed/departmental systems programs

End-User Issues

- Maintaining previously gained independence
- Understanding and using the power of DBMS technology

H

DBMS - The Next Few Years

INPUT firmly believes that DDBMS will become a viable and commonly used data base technology over the next four to six years. In many ways DDBMS is the key to increased benefits from a distributed processing strategy as it will provide the balance required between local autonomy and corporate IS-desired integration.

Exhibit II-8 provides INPUT's projection of DDBMS's capabilities and likely application over the next few years.

- 1987/88: The next two years will be ones of experimentation, seeing the first real application, bringing additional vendors into the market, and seeing additional capabilities added to the DDBMS products.

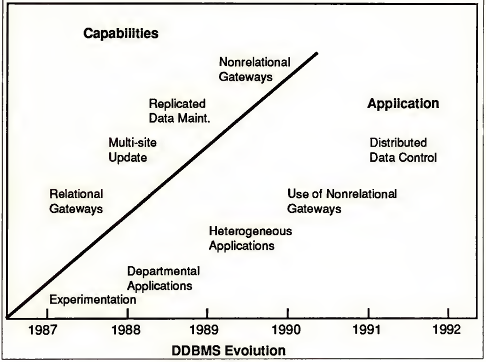
  By late 1988 DDBMS will be reasonably well understood and there will be enough applications to serve as a reference base.

- 1989/1990: During this period DDBMS will become common in the distributed/departmental systems area, and the gateway capabilities to relational and nonrelational corporate data bases will begin to be provided and used. Distributed processing systems will become more integrated within the overall information network of an organization.

- 1991/1992: By the early 1990's DDBMS will be commonplace and will have become a contributor to a new sense of corporate data integration.

EXHIBIT II-8

DDBMS - THE NEXT FEW YEARS



I

Recommendations - What To Do To Get Started

Recognize that DDBMS is a reality and set in place a plan to investigate, and experiment with it. Only an organization that plans no distributed/ departmental environment should ignore DDBMS (if they can).

INPUT recommends the following steps:

- Understand the existing DDBMS products and the baseline definition.
- Audit the data administration function and be sure the procedures are in place to move to a more dynamic and complex data management environment.
- Get on with the implementation of relational DBMS technology. All the DDBMS products will be relational; the end user finds the relational model understandable and it is the way of the future.
- When ready try a DDBMS experiment, keep it controlled, keep it out of the limelight, and use a homogeneous environment.
- Take a look at the existing technology from the leading edge vendors. Don't wait on your existing DBMS vendor. It could be too late.

DDBMS is new, still evolving, but real enough that it should be understood and considered. Exhibits II-8 summarizes INPUT's recommendations.

EXHIBIT II-8

DDBMS - RECOMMENDATIONS

- Recognize that DDBMS is real and learn about it
- Be sure your data administration function is ready
- Get on with the implementation of relational DBMS technology
- Try a learning experiment
- Do not wait for your current DBMS vendor; reach out to new ideas

III

Distributed Data Base - A Definition

III

Distributed Data Base - A Definition

A

A Baseline Definition

Prior to looking at the state of DDBMS technology and the issues related to its use, a definition is needed. In this chapter we will define distributed data base, take a look at the role of the DDBMS in a distributed data base application, and discuss variations that may develop as a result of increasing application of the technology.

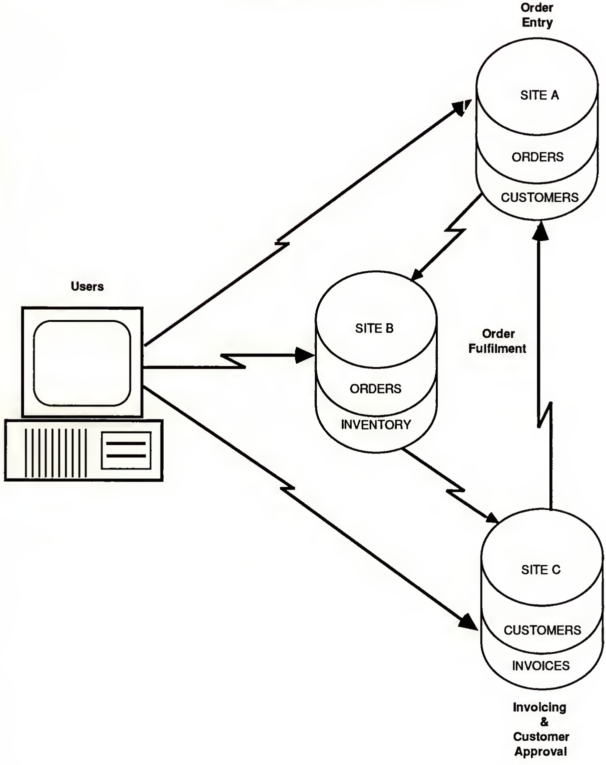
Where possible the definition will be by example. In most cases a simple product distribution application will be used as an example.

1. Framework

- To define distributed data base it is necessary to distinguish between it and distributed data access.
- In a distributed data access environment the user:
  - Is able to access data from more than one physical data base, and perhaps more than one computer;
  - Is aware this is happening, i.e., the seams are visible;
  - Finds the access is a structured, multi-step task with restrictions;
  - Can, for the most part, only inquire (retrieve) the data.

EXHIBIT III-1

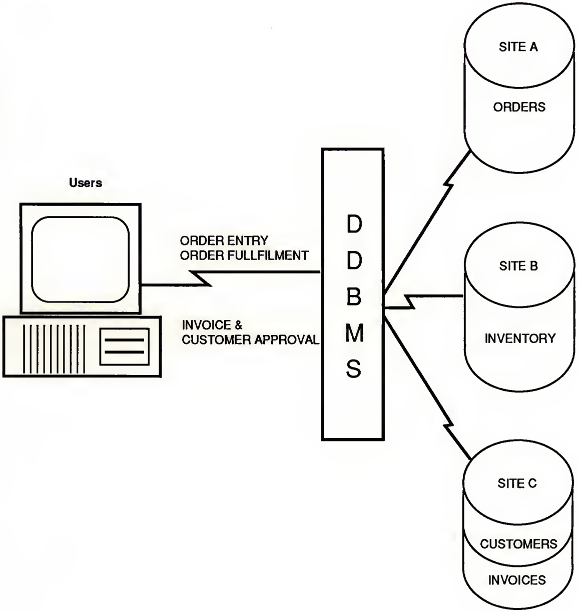
DISTRIBUTED DATA ACCESS



- Exhibit III-1 depicts a typical distributed access environment. The user (perhaps using a PC and a micro/mainframe link) can access both a customer's order and invoice information; however;
  - It is a multistep process - one data base at a time;
  - It generates redundant data;
  - To combine the data requires that new relationships be built after the data is retrieved.
- A distributed data base environment, on the other hand, "hides" the fact that multiple data bases are involved. The user is unaware that the data is physically distributed (the seams are transparent).
- Exhibit III-2 depicts the same application using a distributed data base. In this case:
  - The relationships to access both order and customer invoice status are maintained by the DDBMS;
  - No redundant data is required;
  - Each of the applications operates as if it was part of a single system.

EXHIBIT III-2

DISTRIBUTED DATA BASE



2. Definition

For purposes of this report we will define a distributed data base as a group or collection of "individual" data bases that act together in a "defined partnership." The partnership is defined by the structure of the application and the logical relationships between the data elements contained within the collection of individual data bases.

- The characteristics of this partnership include:
  - The local data bases would usually be on separate computers and capable of operating separately (independently) from each other for local processing.

    For example: When updating the status of an invoice at Site C in Exhibit III-2, the user would interface directly with the site C DBMS.

  - The DDBMS software acts as the "administrative" partner when access to multiple data bases is required, interfacing (directing) the data to the application.

    For example: If, when entering a new order, it is necessary to test a customer's status, the order entry application at Site A would ask the DDBMS to provide the customer status by accessing the customer data base at Site C.

- The distributed capability of the DDBMS would:
  - Manage the data relationships that span multiple DBMSs and sites.
  - Maintain in its data dictionary those relationships while the local data relationships would be managed by the local dictionaries.
  - Be responsible for maintaining any redundant data in the distributed data base in a transparent and timely fashion.

A distributed data base and its DDBMS acts in an identical fashion to a local DBMS except that it does so across multiple DBMSs and computers.

B

The End-User View

The end user of an application does not care, nor should he, where the data required to process an application resides. He only wants ease of access and reliable processing. Thus, to an end user a DDBMS must make the existence of multiple local DBMSs and computers transparent.

- The data can reside in any number of DBMSs and computers within a DDBMS application.
- The DBMSs can be the same or different.
- The environment must be easy to use; i.e., the user is not required to do anything extra when the application must access data from more than one DBMS.

Exhibit III-3 depicts an end user's view of the system in Exhibit III-2. It is as if the system resided within a single DBMS and computer.

C

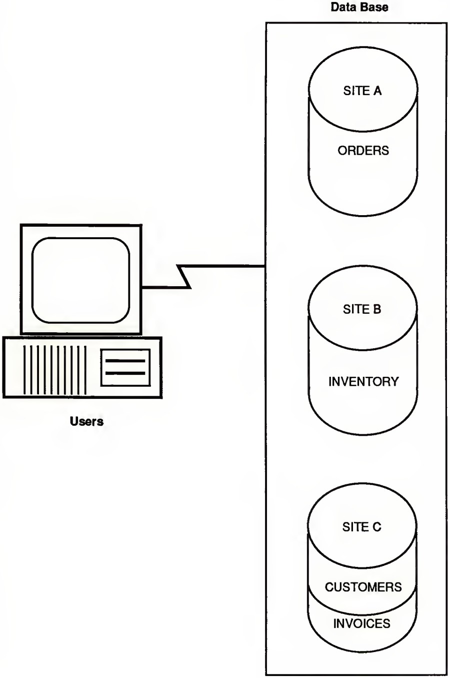
The Information Systems View

The Information Systems view focuses more heavily on the issue of data control and operating effectiveness (response versus cost versus control).

- In a pure sense a distributed data base supports "storing a data element a single time within a network of computers, while allowing it to be accessed and updated from any computer (site) in conjunction with data elements stored on other computers (sites)."
- The software that manages this data storage, a DDBMS, oversees the relationships between data elements within separated computers, manages security and access authorization, and helps position (model) that data within the DDBMS.
- The DDBMS maintains any redundant data on a timely (concurrent or deferred) basis.

EXHIBIT III-3

DISTRIBUTED DATA BASE - THE END-USER VIEW



To IS, a DDBMS is a "means to an end": maintenance of data control while supporting a distributed processing strategy. It provides:

- A means of controlling a system developed on a distributed basis which now requires integration.

- A means to distribute an already centralized application to gain the benefits of distributed processing without losing the existing data control.

D

Homogeneous and Heterogeneous DDBMS Environments

There is nothing in the definition that requires a distributed data base environment to consist of a single type of data base management technology. As long as the DDBMS (administrative partner) can interact with the other DBMSs (the other partners) in the context of the application, successful processing will result.

- If the DBMSs are all of one type (a single product such as INGRES) then the DDBMS environment is referred to as homogeneous.

- If more than one type of DBMS is used (for example, Oracle and DB2) then it is referred to as heterogeneous.

Exhibit III-4 depicts both a homogeneous and a heterogeneous DDBMS environment.

- In the heterogeneous environment DBMS type A is serving as the DDBMS. This means:

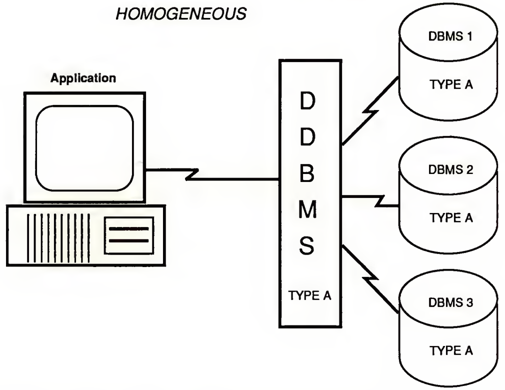
  - DBMS, Type A must have the capability (a gateway) to submit its data requests to DBMS Types B and C.

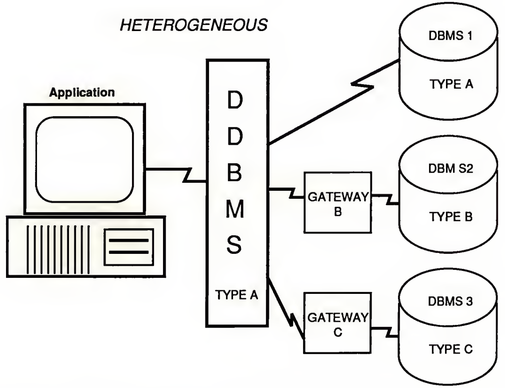
  - DBMS Type A must be able to translate its requests into Type B's and Type C's formats and then translate the response back to its format. (A step by step example is included in Chapter IV.)

  - Using the previous example, if DBMS 3 (Type C) contained the customer data base, then a request to provide customer status to DBMS 1 at order entry time would have to go through a "gateway" and be translated from Type A format to Type C format.

EXHIBIT III-4

DISTRIBUTED DATA BASE ENVIRONMENTS





E

The Twelve Rules Of A Distributed Data Base

The technical implications of the above definition are extensive. The tracking, translating, and processing that are required can imply a significant increase in overhead.

Much has already been written about the characteristics of distributed data bases. INPUT has chosen one of the more complete discussions and summarized it here. It provides a guidepost for evaluating both DDBMS technology and potential distributed data base applications.

The "Twelve Rules for a Distributed Data Base" were developed by The Relational Institute and its founders, C.J. Date and E.F. Codd. (For a more detailed discussion the reader is referred to the institute's report "What Is A Distributed Data Base.")

The "rules" listed in Exhibit III-5 are based on the concept that "to a user a distributed data base system looks exactly like a nondistributed system." The rules are:

Rule 1: Local Autonomy - Data within each local DBMS is locally owned and managed, and local operations are not affected by participation in the distributed environment.

Rule 2: No Reliance On A Central Site - All sites are treated equally; there is no master site required for processing.

Rule 3: Continuous Operations - There is never a need for a planned shutdown to accomplish normal processing.

Rule 4: Location Independence - Users do not need to know where the data is physically stored, and the data can be migrated to another physical location without impacting the application.

Rule 5: Fragmentation Independence - Data relationships can be split (fragmented) and stored on separate computers.

Rule 6: Replication Independence - Data can be represented at the physical level by many copies stored at separate sites, with the system maintaining the copies from the primary copy.

Rule 7: Distributed Query Processing - Query execution will be a distributed process consisting of local and multisite I/O as well as data communications.

Rule 8: Distributed Transaction Management - Both recovery control and concurrency control are managed locally and for the DDBMS as a whole. This is most commonly accomplished using a "two-step commit" process.

Rule 9: Hardware Independence - The same DBMS can run on differ ent hardware environments (homogeneous DBMS but hetero gene-ous computers).

Rule 10: Operating System Independence - The same DBMS can run under different operating systems (homogeneous DBMSs but heterogeneous operating systems).

Rule 11: Network Independence - The network component of the DDBMS is capable of communicating in multiple protocols.

Rule 12: DBMS Independence - The DDBMS software can support more than one type of DBMS.

These rules define a capability that is perhaps ideal, but serve as a standard for the DBMS vendor to work towards.

In Chapter IV we will look at the status of DDBMS technology and refine this chapter's definition.

EXHIBIT III-5

DISTRIBUTED DATA BASE - TWELVE RULES

1. Local autonomy
2. No central site reliance
3. Continous operations
4. Local independence
5. Fragmentation independence
6. Replication independence
7. Distributed query processing
8. Distributed transaction management
9. Hardware independence
10. Operating independence
11. Network independence
12. DBMS independence

IV

The State of Distributed Data Base Technology

IV

The State Of Distributed DBMS Technology

A Introduction

In this chapter we will look at the state of DDBMS technology, the products on the market, and how these products compare to the definition in Chapter III.

To provide a status report on DDBMS technology INPUT has surveyed the recent literature and talked with or reviewed the plans of a number of DBMS's vendors. Included were:

- The leading edge relational DBMS vendors: Informix, Oracle, Relational Technology, and Sybase.
- Longstanding IBM mainframe DBMS vendors: ADR, Cincom, Cullinet, and Software AG.
- Hardware/software vendors: DEC, IBM, and Tandem.

The following questions were addressed to each of the vendors interviewed:

- What are the primary market/customer demands that are driving the development of DDBMS technology?
- What is the status of DDBMS capabilities within the vendor's DBMS product line?
- What are the critical technology issues relative to successful DDBMS application?
- What application areas/characteristics are best suited for a DDBMS implementations?
- What are the nontechnical (administrative, political, operational) issues critical to successful DDBMS application?

The vendors' responses to these questions are included in this and the next chapter (Chapter V - User Issues And Experiences).

B

Status Of Relational DBMS

The relational model has come of age:

- IBM's DB2 is a success and the strategic basis for IBM's future DBMS plans. IBM's strong commitment to the relational DBMS market has caused the relational market to explode.

- Companies such as Oracle and Relational Technology now count their installations (not counting PC users) in the thousands. They offer their products on multiple platforms (mainframe, mini, and PC) and are leading the introduction of relational DBMS capabilities into the distributed/departmental processing arenas.

- IS organizations that have not "bitten the relational bullet" are hurrying to do so, and may be surprised to find relational DBMSs running on existing departmental processors.

- End users have found the "relational model" understandable and are using it. Oracle, Relational Technology, and others have brought ease of use to the relational DBMS environment.

Structured Query Language (SQL) has become the standard for relational DBMS systems and the ANSI standard has been approved. Those relational DBMS products that were not SQL-compatible will soon offer the capability.

The longstanding concern about performance of relational DBMS systems is being addressed by all vendors, and the first performance-oriented systems are being released.

For a more complete look at the DBMS market and its future, the reader is referred to INPUT's report **Future Data Base Markets, 1987-1992.**

C

What DDBMS Systems Are On The Market

Exhibit IV-1 lists the vendors surveyed along with their currently available relational and distributed DBMS products offerings. Five vendors have announced and are delivering DDBMS products.

- Oracle's DDBMS product is called SQL*Star and is an extension of the Oracle relational DBMS.

- Relational Technology's product is called INGRES/STAR and is an extension of INGRES.

- Sybase's product is called Sybase and was first released in May, 1987, with distributed capabilities included.

- ADR's product is called D/NET, has been on the market since 1986, and operates with DATACOM, ADR's DBMS which has partial relational capabilities. It only operates in an IBM mainframe environment. Version 2 is due in late 1987.

- Tandem's product is called NonStop SQL and is a relational alternative to the company's existing fault-tolerant, distributed data management capability. It only runs on Tandem computers and is tightly linked with the NonStop operating environment. Tandem's DDBMS is in the early release phase.

Each of the products (except D/NET) is based on relational concepts and SQL. No non-relational DDBMS products were identified or found to be under development. Much of the literature and most of the experts indicate that the relational model is a requirement for a DDBMS.

EXHIBIT IV-1

DBMS VENDORS AND THEIR CURRENT RELATIONAL PRODUCTS

VENDOR	RELATIONAL	DISTRIBUTED
ADR	Datacom	D/NET
Cincom	Supra	
Cullinet	IDMS/R	
DEC	R/db	
IBM	DB2	
Informix	Informix-SQL	
Oracle	Oracle	SQL*Star
Relational Technology	INGRES	Ingres/Star
Software AG	Adabas	
Sybase	Sybase	Sybase
Tandem	NonStop SQL	NonStop SQL

In response to the question "What are your product plans?", the other vendors were found to:

- Have something under development: IBM, Cincom, Software AG, Informix.
  - IBM's R/Star is reported to be in development but with no delivery announced.
  - Software AG's release, ADABAS 5, coming in 1987, may include some distributed capabilities as well as relational enhancements.
- Be concentrating on obtaining full relational/SQL capabilities: Cincom, Cullinet, DEC.
  - DEC is rumored to be readying a replacement for its R/db.
- Be porting their products to multiple platforms: Cincom, Cullinet, Informix.
  - Cullinet's IDMS/D for DEC/VAX is rumored for later this year and is expected to be relational.
  - Informix is porting its products to DEC/VAX and plans to port to IBM.
  - Portability is a main element of the product strategies of Oracle and Relational Technology. Their products already run under a variety of mainframe, mini, and PC environments.

DDBMS technology exists from a number of vendors. As we will see, the existing systems do not have all the functional capabilities to qualify as a pure DDBMS. However, the products that are available do work and provide extensive capabilities. Increasingly, a vendor's stated DDBMS direction is becoming one of the selection criteria in the customer's evaluation of relational technology.

When asked "What do you tell a customer who wants DDBMS capabilities?", a vendor that does not have the capability said: "We tell the client to build a distributed processing application using our relational DBMS, and plan to integrate it when the distributed DBMS capability is available."

D

A Closer Look At The Early DDBMS Technology

1. A Description

Using Oracle and INGRES (Relational Technology) "Star" products as the basic example, we can look more closely at this new technology.

- These relational and distributed DBMS products are part of a total development/data base product family.

- These products already or soon well run under MS-DOS, UNIX, VAX/VMS, VM, and MVS as well as other operating systems. Portability is key to the strategy of DDBMS and is a principle element of the strategy of each of these companies.

The term "star" has been adopted to refer to those capabilities that link the data base across multiple platforms as opposed to the relational DBMS itself.

- The term star was apparently first used by IBM (R* or R Star), at least on a common basis.

- Oracle and Relational Technology both market the "star" components as separate products.

Exhibit IV-2 diagrams the components of the distributed DBMS as offered by Oracle and Relational Technology.

- The product consists of four levels:

  - Level 1 (the bottom level) consists of the query processing logic, including optimizers and the local data dictionary.

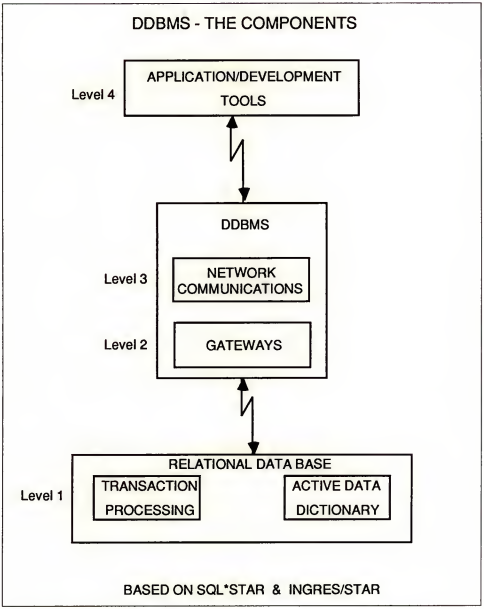
  - Level 2 consists of the linkage that interacts with each platform, including the gateway when the environment is heterogeneous.

  - Level 3 contains the network module that communicates with the other platforms using whatever protocols are required.

    - Levels 2 and 3 make up the "star" or distributed capability and permit an open architecture (multiple, computers, operating systems, and DBMSs).

  - Level 4 (the top level) contains the application development tools and/or the application itself.

EXHIBIT IV-2



- In the case of Oracle and INGRES, the development tools include a 4GL, form generators, and ease-of-use PC-like interfaces.

- Actual processing occurs in two steps using a two-step commit concept.

The Sybase implementation is somewhat different. There are only two levels and the programs within the open architecture are imbedded within the development tools and the local DBMSs, respectively.

Let's take a closer look at the data dictionary element:

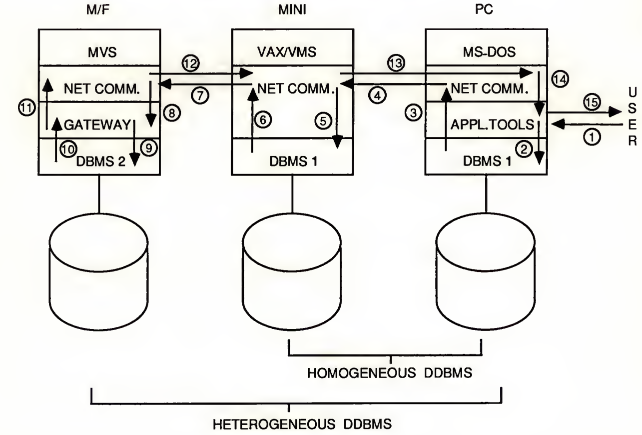
- In the Oracle and INGRES products the data dictionary in the local DBMSs includes the distributed relationships.
- Software AG's distributed environment, due in 1988, classifies its data dictionary at the top level and plans to have both local and global active dictionaries in a distributed environment. The global dictionary would contain the distributed relationships.
- ADR's D/NET uses a separate active dictionary within D/NET to oversee the distributed relationships. The local relationships are contained in local dictionaries.

2. A Processing Example

Exhibit IV-3 provides a picture of the elements of a heterogeneous DDBMS environment. It includes a DBMS Type 1 running on a PC and a minicomputer with a gateway to a mainframe running a DBMS Type 2.

EXHIBIT IV-3

DBBMS - A PROCESSING EXAMPLE



- The DBMS/Star program runs on all three computers with the "network" component on all three computers and the "gateway" component on the mainframe.

- Let's follow a transaction step by step starting at the PC:

  1. The PC initiates the transaction.

  2. The application (tool) queries the local DBMS 1.

  3. The PC's DBMS indicates it needs data from the mini and initiates a request sent through the PC's network module.

  4. The mini receives the request.

  5. The mini queries the local DBMS1 and learns it does not have all the data required.

  6. The mini's DBMS issues a request through its network module to the mainframe.

  7. The mainframe network module receives the request.

  8. The request is sent to the gateway and translated into DBMS 2's format.

  9. Once in DBMS 2 format the request is submitted for processing.

  10. When complete, the data is sent to the gateway.

  11. The data is translated back to DBMS 1 format and sent to the network module.

  12. The data is received by the mini, and if appropriate, combined with data from the mini DBMS 1 and sent on to the PC.

  13. This combined data (mini and mainframe queries) is further combined with data from the local DBMS 1.

  14. The data is returned by the application to the user.

- If the transaction caused an update to a data element on both the PC and the mini (e.g., a replicate exists) the star program would initiate and confirm the update through a series of comparable steps.

As can be seen, the star programs act as the "administrative partner" for the distributed application, knowing where the data is located, requesting the data, and being sure all necessary updates are processed. The star programs also oversee the recovery process when required.

E

Comparision To The Definition

Not unexpectedly, the current products lack some of the functional capabilities implied by the definition of Chapter III. But the existing products, as exemplified by SQL*Star and INGRES/STAR, do provide many of the capabilities.

Exhibit IV-4 contains INPUT's macro assessment of the current DDBMS technology versus the twelve rules. A number of the rules are directly supported while others are partially supported. As expected, the vendors have plans to release additional capabilities.

Although INPUT believes that the available products do provide the capability to build effective and secure DDBMS applications, it did not perform a detailed analysis of each product's capabilities nor examine the issues associated with processing overhead.

The following are examples of current limitations of most of these products:

- An update can be processed at only a single location (that is, data can be retrieved from multiple DBMSs but only one can be updated). Only D/NET can process multiple updates.

EXHIBIT IV-4

CURRENT DDBMS TECHNOLOGY VERSUS THE TWELVE RULES*

	RULE	STATUS
1.	LOCAL AUTONOMY	SUPPORTED
2.	NO RELIANCE ON CENTRAL SITE	SUPPORTED
3.	CONTINUOUS OPERATIONS	PARTIAL
4.	LOCAL INDEPENDENCE	SUPPORTED
5.	FRAGEMENTATION INDEPENDENCE	SUPPORTED
6.	REPLICATION INDEPENDENCE	PARTIAL
7.	DISTRIBUTED QUERY PROCESSING	PARTIAL
8.	DISTRIBUTED TRANSACTION MGMT.	PARTIAL
9.	HARDWARE INDEPENDENCE	PARTIAL
10.	OPERATING INDEPENDENCE	PARTIAL
11.	NETWORK INDEPENDENCE	PARTIAL
12.	DBMS INDEPENDENCE	PARTIAL

* SOURCE - THE RELATIONAL INSTITUTE

- The gateway components are for the most part still in development.
- The data dictionaries are not global in all of the products, which may impact processing efficiency.
- The ability to replicate data and perform the required timely updates is generally lacking, but under development.
- The star capability is not available on all of the platforms that the DBMS is currently available on, but again these "ports" are generally in process.

F

DDBMS And The Product Life Cycle

Experience has taught the IS profession that software capabilities often lead their general use by a significant time period. There is no reason to believe DDBMS will be different.

- DDBMS is in the early stages of its life cycle. Major technology strides have been made and commercially available products are available.
- Application (use) is beginning, but as will be seen in Chapter V, the efforts are mostly experimental.
- The next two to three years will bring major additions to DDBMS capabilities.

Chapter V will take a look at the current level of use.

V

User Issues and Experiences

V

User Issues and Experiences

A

Introduction

Chapter III provided a definition for distributed data base. Chapter IV provided a status report on the DDBMS technology. In this chapter the report will address the DDBMS subject from the user (both Information Systems and end user) point of view. The chapter includes:

- Discussion of the issues
- Current levels of application
- Potential uses
- Critical elements for success

INPUT interviewed twenty-three IS managers relative to their organizations' DDBMS activities. The group, while small, was very diversified.

- Company characteristics:
  - Fortune 100 companies in the banking, manufacturing, distribution, oil, utility, and pharmaceutical industries.
  - Smaller organizations in the education and medical supply industries.
- Characteristics of the IS environment:
  - Both centralized and decentralized structures.
  - Both centralized and distributed processing environments.
  - A few had progressive distributed processing environments.
  - Most had active end-user programs.

- Organizations with relational DBMS experience and others just in the investigative stage.

- Users of leading edge relational DBMS technology provided by Oracle and Relational Technology corporations.

The results of these interviews along with the vendor discussions form the basis for this chapter. First, a brief overview:

- Twelve of those interviewed (about one-half) could offer no significant input on the distributed data base management subject. In two instances their environment did not support the concept, but in general these organizations were not far enough into relational DBMS and/or distributed technology to contribute to the discussion.

- Four of those interviewed were in the process, or had recently completed, a study on relational DBMS and had included distributed capabilities as one of the selection criteria. The emphasis, however, was on getting ready to introduce relational concepts.

- Three were in the early stages of controlled experimentation with distributed capabilities as provided by Oracle or INGRES.

- While no true production applications were found, there are plans to have production systems within the next twelve months.

These results are not surprising as the existing DDBMS capabilities are less than one year old . Although the survey was unable to identify many users actually pursuing DDBMS applications, many respondents have given considerable thought to key issues and critical success factors.

B

What Are The Many Issues

INPUT has categorized the key issues as Technical, IS Management, and End User.

1. Technical Issues

As with each new stride in computing technology, DDBMS will bring with it new challenges. For the most part, DDBMS adds another dimension to the existing challenges of data management, data communications, performance, security, and the more recent challenge of systems integration.

Exhibit V-1 lists the technical issues that INPUT found to be most important.

EXHIBIT V-1

DDBMS - PRINCIPLE ISSUES

TECHNICAL

Data Administration/Modeling

- Active data dictionaries
- Cross-site versus local relationships
- Replicated data
- Multi-site administration

Data Communications/Networks

- Heterogeneous environments
- Site transparency

Processing Performance

- Computer versus network optimization
- Replicated data

Security/Recovery

- Local versus multi-site update
- Multi-node recovery

Systems Integration

- Heterogeneous environments
- Data networks

a. Data/Administration/Modeling

IS organizations that have implemented departmental and distributed processing systems:

- They are dealing with the issues of common and consistent definition of data routinely transferred between computers (the "islands of information " problem).
- If DBMS systems are in use with these applications, the issues of multiple dictionaries and data relationships have, at least, been recognized.

- The challenge of maintaining multiple copies of data bases weekly, daily, or when polled has been addressed.

All of these issues plus more are part of the DDBMS challenge.

- Having a data element maintained at a single site but accessable for use by multiple sites helps address the accuracy problem caused by redundant data, but introduces the problems of how the update is controlled, how and where the data relationship is administered, and security.

- Including a data element as part of both local and multi-site relationships makes the control and tracking of data relationships more complex. The local-site administrator must know what multi-site relationships exist before considering any local changes, and the central administrator (if one exists) must know all of the relationships.

- Using the PC as one of the sites in a DDBMS application and as the "sole" site for segments of a distributed data base raises security concerns. To one of the IS managers interviewed, the idea that corporate data would reside only on a PC and be part of a larger network application was abhorrent.

- Achieving performance in a DDBMS application will most likely introduce replicated data into the environment. The ability to decide what data should be replicated and then maintain all the copies is an extreme challenge for the technology and the data administration function. DDBMS will require an increased focus on data modeling; not just during the development phase but continuously as the application changes during its use.

- Permitting data elements to be updated from multiple sites adds another dimension to the data model and to security. Authorization controls must be complete and well administered.

The available and forthcoming DDBMS technology addresses these problems to some degree.

- The systems include truly active data dictionaries that are an integral part of the relational DBMS.

  - The dictionaries exist at each node, include the distributed relationships that are active at that node, and automatically maintain the distributed relationships.

  - Future versions may also have a global dictionary that oversees the entire data base and, with the local dictionaries, helps optimize per formance.

  - In at least one case the data dictionary can contain actual SQL procedures that are called by a transaction. By making the procedure part of a distributed relationship, consistency across platforms is assured.

- The early releases of DDBMS systems have some limitations relative to multi-site update and replicated data maintenance.
  - All but one of the available systems restrict "updates" to a single site. The transaction can retrieve data from any and all sites, but can update only one of those sites.
  - Exhibit V-2 shows an example of replicated data. The "customer" data base segments are replicated at each of the order entry points for processing efficiency. However, updates to customer data would occur at the central site and then be replicated (maintained) by the DDBMS at each order entry site.
  - Multi-site update and the automatic maintenance of replicated data (as required by the Exhibit V-2 example) is to be addressed by releases scheduled from the leading vendors in 1988.
- Data modeling is already a well recognized issue for the relational DBMS vendors. They all indicated plans to have data modeling tools to help achieve and maintain processing versus network efficiencies, and to help make the data distribution and replication decisions.
  - Some transaction optimizing tools already exist.

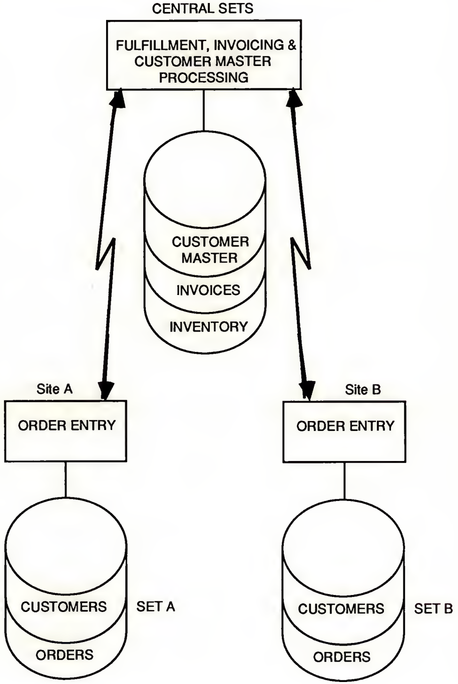
b. Data Communications/Networks

A DDBMS application requires a network operating with extreme reliability, in a heterogeneous environment, and with multiple protocols.

- Moving from a distributed processing network with its periodic data transfer and perhaps dial-up communications to one supporting a transaction based DDBMS environment is a major challenge.
- Moving from a terminal-oriented network to one linking multiple processors on a transaction basis is an equally challenging task.
- There are not a lot of IS communications functions currently overseeing a transaction-based multiple protocol environment.

EXHIBIT V-2

DDBMS - A REPLICATED DATA EXAMPLE

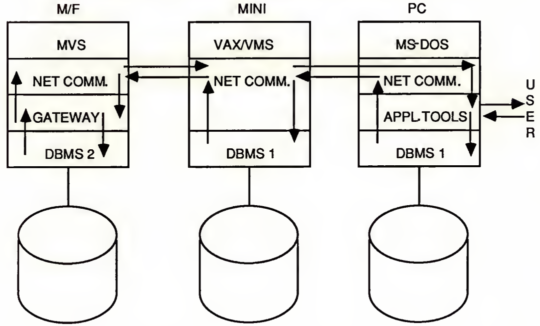


The DDBMS vendors have met this challenge by assuming the burden of managing the communication at each node in the network.

- Exhibit V-3 shows a "network" program operating in each node of a heterogeneous environment (different DBMSs as well as computers and operating systems). This network program, which is the same at all nodes, handles the transaction communication between nodes using the protocols specified at each node.
  - The network program handles the translation between the VAX/VMS and IBM SNA protocols.
  - As noted earlier, the "gateway" program handles the translation between the DBMS formats.
  - These two programs (network and gateway - the "star") allow the DDBMS transaction to get right to the door of the heterogeneous DBMS before translation takes place.
- For example, Oracle's SQL*Net carries the burden of communicating in numerous protocols permitting their relational DBMS to "fit into" an existing network whether in a DDBMS or just a distributed processing environment.

EXHIBIT V-3

DBBMS - A DATA COMMUNICATIONS EXAMPLE



- A major manufacturer is using SQL*Net in a distributed processing application to handle the heterogeneous communications environment.

c. Processing Performance

The success of the relational data base model can be attributed, at least in part, to its ease of comprehension and use. As such, much of the early use was in the decision support systems area where data base size and transaction volumes were nominal.

- That success has pushed the technology into the departmental systems area and higher volume applications. As a result, the performance restrictions of the early relational DBMSs have become an exposure and an issue.

- Now DDBMS adds additional processing overhead to the relational DBMS that could further aggravate performance. Included are:

  - The communications and processing load of the distributed data retrieval.

  - The processing load of the two step commit process.

INPUT's research found this issue to be a concern of IS managers. The performance reputation of relational DBMSs is a major cause of IS's feet dragging on relational DBMS technology. INPUT also found this issue to be well recognized by the vendors who have focused development efforts on the efficiency area.

- IBM has stated its intent to enhance DB2's throughput with each future release.

- Sybase has selected the performance area as its product niche. Its technology has been designed with throughput as the focus.

- Both Oracle and Relational Technology indicated to INPUT that improved transaction performance is a major area of investment in their current R&D programs.

- The products from Oracle, Relational Technology, and others include transaction optimization logic.

Whether relational and distributed DBMS performance ever reaches that of IMS or comparable products only time will tell. But, given the growing preference for relational DBMS technology in the using community and the almost total commitment to it by the DB vendors, it is reasonable to project ongoing performance improvements.

d. Security/Recovery

The longstanding issues of security and recovery come front and center in a DDBMS environment.

- Overseeing access and update authorization in a multi-site application adds new dimensions to the data security administration process.

- Having an application dependent on more than one node for routine processing increases the probability of failure and requires additional processing to assure all nodes can complete their part of the transaction processing.

- During a recovery process at one of the nodes those transactions that included distributed activity must be properly recovered.

Again, the existing technology addresses these issue to some degree.

- The journal logs of the local DBMSs apply to the distributed activity. During the recovery process the distributed and local transactions are considered.

- A distributed transaction does not complete processing (that is, "commit") at any site until all sites confirm they can process the transaction. If one site cannot commit, none of the sites execute. This is called the "two-step commit" process and is in general use by the DDBMS vendors.

- The security controls of the relational DBMS at each site also govern the distributed elements of the application. Access and update authorization can be controlled at the appropriate level, but not without additional administrative effort.

e. Systems Integration

Today's distributed processing environment is a mix of processing capabilities. Linking it together is a goal which can be accomplished, at least in part, by DDBMS technology.

- The issues of systems integration are primarily the challenges of heterogeneous technical environments and the related data communications and data format problems.

- The definition of distributed data base and the twelve rules set the requirement that a DDBMS support heterogeneous environments. This is a challenge that has been accepted by the leading edge DDBMS vendors. They have included the network component and have made portability a key element of their strategies.

- It seems reasonable to predict at this stage that one early use of DDBMS technology will be in the systems integration area.

2. IS Management Issues

Exhibit V-4 identifies the principle IS management issues and their elements. These issues center on the challenges of adopting new data base technology into an existing computing environment and maintaining a semblance of influence and control in a IS world strongly influenced by the end user.

a. Managing Corporate Data

Overseeing the definition and administration of corporate data is a challenge understood, but seldom conquered.

- Distributed systems, inadequate data dictionary technology, and data transfer volumes between computers of all types have made the data administration task far more complex.
- While IS has been able to enforce data administration standards at the point of transfer, local autonomy has, at times, limited the completeness of data administration at the distributed/departmental systems level.

EXHIBIT V-4

DDBMS - PRINCIPLE ISSUES

IS MANAGEMENT

Managing corporate data
- Corporate versus departmental data
- Maturity of data administration functions
- Data dictionary implementation

Adopting relational DBMS technology
- Corporate versus distributed use
- Mainframe versus mini use
- Single versus multiple DBMSs

Managing distributed/departmental systems
- Replacing file-based applications
- Data transfer between DBMSs
- IS influence in departmental systems

Furthermore, a major proportion of current distributed systems are not DBMS-based and not subject to automated management tools.

As part of the interview IS managers were questioned about their data administration functions and their use of data dictionaries. The general response was to lament their success in this area. The base job is being done well, but the total task needs renewed attention.

- They routinely complained about the lack of an active data dictionary for DB2 from IBM.

- They would like increased influence on data administration performed at local levels.

- They would like to make better use of the dictionaries they are currently using.

The more aggressive relational DBMS vendors have provided integrated active dictionaries and have extended their capabilities to the distributed application. The power of the data dictionary is a key factor in the power of the DDBMS.

- The growing use of these relational DBMSs in departmental systems offers IS the opportunity to work more closely with the end user and improve the data administration process.

- The first DDBMS application will provide the opportunity to link the data administration process across at least a portion of the organization.

Just as distributed processing stretched the data administration process, so will DDBMS. However, this time IS will have an integrated tool to administer data at multiple levels.

b. Adopting Relational Data Base Technology

Of the organizations interviewed, only four had measurable experience with relational DBMS technology (one started using Oracle in 1982). Some had DB2 installed but were only experimenting with it, while four were doing or had recently done the "obligatory" corporate study on relational DBMS.

- These studies tend to focus in how relational DBMS can be applied, which system(s) should be adopted, and can a single DBMS be used (homogeneous versus heterogeneous). Typically, these studies result in:

  - Inclusion of the need for distributed data base capabilities as one of the selection criteria.

  - Adoption of DB2 at the corporate level and one of the new relational DBMS (eg.Oracle, INGRES, etc.) for distributed/departmental use.

- Official sanction for what may already be happening, for example, the selection of a mini-based DBMS for a specific applications.

Sales of relational DBMSs are growing rapidly, forecasting a heavy growth in the development of relational production transactional system..

- In 1987, the combined sales of Oracle, Informix, and Relational Technology will exceed $225 million for relational BBMS products. The majority of these will be for use on minicomputers.

- DB2 is also selling well. In the next eighteen months DB2 application development will also grow rapidly.

The move to relational is well underway. Learning to develop production relational DBMS applications is a necessary precursor to considering DDBMS applications.

- IS has to learn to think "relationally," to use active data dictionaries, and to use the 4GL that is often part of the DBMS.

- Most IS users would like to see a DDBMS success story or two before venturing beyond individual relational applications.

- One IS manager indicated there were a number of commitments to be met by the vendors relative to performance and function on the DBMS, let alone proof that the distributed version is ready to go.

- Another manager who was well into DB2 development complained that IBM was nowhere near a distributed capability.

In summary, the move to relational DBMS is well under way, but there is much to be learned before trying DDBMS.

One interviewee who was starting a DDBMS application indicated the company was going to build it on a centralized basis first, and distribute it "later."

c. Managing Distributed/Departmental Systems

Integrated distributed processing has been a long sought after dream of some IS managers; however, they never dreamed they would be implementing under the strong end-user/departmental influences of the mid-1980s.

- In many organizations the differences in their distributed and departmental strategies are indecipherable.

- The end-user influence is now significant, continues to grow, and will influence future DBMS decisions.

Many of the distributed/departmental systems implemented to date have been flat file versus DBMS-based and have not always met the user's requirements for flexibility and ease of use.

- The attraction of the newer relational DBMSs with their integrated 4GL's and other user-oriented application - development tools can be attributed to ease of use.

- The active dictionaries provide improved control and a friendlier interface for the uninitiated in data base technology.

Now that corporate IS is sanctioning the use of a mini-based relational DBMS, focus will shift to converting existing distributed/departmental systems to relational tools. This will:

- Improve routine data interchange between corporate and departmental systems.

- Provide IS an opportunity to gain influence in the departmental systems arena through a proactive data administration program.

- Give IS an opportunity to train end users in the use of data base technology.

In summary, IS can use relational DBMS and later DDBMS capabilities to increase its effectiveness in the departmental systems arena.

3. End-User Issues

Although no end users were interviewed, it is possible to identify some key issues from the IS and vendor conversations regarding the "end user" point of view.

Exhibit V-5 lists the principle end-user issues and their elements. These focus on the maturing end user of computing technology.

a. Maintaining Influence

As the end user reaches out for more powerful computing technology, he or she will be concerned about local autonomy and with maintaining and increasing his or her influence on computing decisions that directly affect his or her business functions.

- IS managers may wonder why the end user wants to go beyond the freedom and safety of the PC to a multi-user environment. "Don't they know the problems they will encounter?"

EXHIBIT V-5

DDBMS - PRINCIPLE ISSUES

END USER

Maintaining influence

- Local autonomy
- PC orientation
- Vendor influence

Understanding and applying DBMS technology

- End-user DBMS understanding
- Ease of use
- IS responsibility and opportunity

- The answer to the question is, of course, "NO"; and consequently, end users are moving quickly in this direction. The PC turned out to be manageable, if not easy, but also limited. To get the real production job done the end user believes more power, function, and autonomy are required.

The end user will not give up the influence that has been gained in the early 1980s and wants to enhance his or her own computing power.

- He or she is or will be attracted to relational DBMS technology because of its relative ease of use.
- DDBMS, however, may receive a mixed reaction.
  - If suggested by IS it may appear threatening.
- "Why should I depend on the corporate data base on a transaction basis. Nightly updates are satisfactory, and others can wait for 24 hours for my department's updates."
- "What do you mean it will save computer and communications costs. How do I know the added constraints won't raise my department's costs."
  - If the vendor suggests DDBMS it may appear attractive, especially if the vendor places it in a departmental context.

The end user is trying to understand and apply DBMS capabilities from a local viewpoint and is likely to oppose anything that would appear to

require the planning and cooperative operational management approach that DDBMS will require.

b. Understanding And Applying DBMS Technology

The end user's experience with DBMS technology has generally come through exposure to Information Center and PC tools. The applications they have built have in many instances really been file-based in their design.

The end user does understand, at least in his terms, his needs and usually has enough computing knowledge to look for a solution. He is discovering the relational DBMS environment and beginning to adopt it, in particular as delivered by the leading edge vendors with their easy-to-use, PC-like interfaces.

This provides IS with both an opportunity and an obligation.

- As relational DBMS technology is introduced, in particular for a departmental system, IS must take on the obligation to train the end user in the DBMS application development process and data administration. IS must explain the underpinnings of production applications and why the application development process should be more deliberate.

- By providing training and proactive guidance, IS has the opportunity to influence the applications that result and provide a foundation to evolve to a DDBMS environment when appropriate.

C

The Early Experiences

1. What Is Being Tried

As noted in the introduction to this chapter, it is a little early in the DDBMS product life cycle to find much significant experience. However, some experiments and plans were identified through the survey.

- A small college whose administrative systems already use a DEC-based relational DBMS is considering the distributed capability between PCs and two DEC/VAXs as a means to offload processing and delay hardware upgrades. The DDBMS would also allow them to take advantage of the application porting capability, permitting development on the DECs and migration to the PCs.

- An oil company is considering the DDBMS capability as a means to share its exploration data between sites. Today, much of the data is duplicated at multiple sites and is costly and hard to maintain.

- A major university, which is run like a collection (partnership) of decentralized colleges, intends to use DDBMS technology to build its student information system.

- A major bank is using a leading edge relational DBMS to build a securities account management system to support worldwide trading. It will first be built with a centralized data base, but there are plans to distribute the data base across three sites (US, Europe, Far East).

Beyond this, very few specific DDBMS plans were identified by respondents, even at companies that had purchased the "star" components.

2. What Will Be Learned

By doing a first DDBMS application, the IS users interviewed can expect to learn that:

- Relational DBMS technology must be understood first. If one has not previously developed relational-based applications, this is a necessary first step. Any DDBMS application has local elements that must be considered, and the developer must learn to think relationally.

- The DDBMS technology currently available is powerful and reasonably easy to use. Using it in a mini/PC environment will quickly demonstrate the concept.

- The trade-offs between distributing a data base versus duplicating data in a distributed processing environment will be recognized as a compromise. Redundant data will not go away. It will just be easier to maintain under DDBMS technology.

- The IS data administration function must, in many instances, be more disciplined than it is today and become involved on an ongoing basis. The active data dictionaries must be successfully used to apply the DDBMS technology.

- The introduction of data base capabilities to the end user offers IS an opportunity to build a new and productive bridge to the end user, and a chance to regain some control over departmental use of corporate data.

- Those who begin to use the available DDBMS technology, versus waiting for IBM and other DBMS vendors, have a chance to jump ahead.

3. Where Will DDBMS Fit In

INPUT believes that DDBMS offers, over time, an opportunity for IS to broaden the impact of its "data management" process across an organization which is using a distributed processing strategy, and by doing so to provide an improved information management service to the organization.

- DDBMS offers the chance to make data truly corporate while distributed.

- DDBMS offers the chance to separate the of computing environment employed from the use of the data. The portability and computing environment independence of the early DDBMS products provide a major first step.

- DDBMS offers a balance between end user-freedom and IS corporate control responsibilities.

It will take a number of years (perhaps five or more) for DDBMS to become a major application development force. INPUT believes that early DDBMS efforts will concentrate in the distributed/departmental area. As Exhibit V-6 indicates, this is Level 2 in the INPUT Systems hierarchy. Some reasons follow:

- The leading relational DBMS vendors have been emphasizing this market segment and have a sizeable installed base. The applications are easier and localized permitting quicker development.

- The early DDBMS technology is available in mini and PC environments. While Oracle and INGRES are available under IBM's VM and MVS, the installed base is small and not all components of the products are available. The first DDBMS efforts will be in a homogeneous DBMS environment and mini to mini or PC to mini.

- The thought of actually spreading a corporate data base across multiple sites remains frightening to IS, while the user is less timid and has a simpler environment. The interaction and synchronization between corporate and departmental data bases can remain periodic.

- Many major corporate data centers are already using nonrelational DBMSs which implies a heterogeneous environment and the use of gateways.

- The gateways to nonrelational DBMSs are only in early development.

- IS is preoccupied with implementing relational technology, which must preceed distributed data base activity.

EXHIBIT V-6

LEVELS OF COMPUTING

LEVEL	NAME	COMPUTER	APPLICATIONS
1.	CORPORATE	MAINFRAMES	CORPORATE INTER-DEPARTMENTAL HIGH VOLUME TRANSACTIONS
2.	DEPARTMENTAL	MINI	DISTRIBUTED MULTI-USER INFORMATION CENTER
3.	PERSONAL	PC	DECISION SUPPORT PERSONAL PRODUCTIVITY

The exceptions to this will be Tandem's NonStop SQL and ADR's D/NET. These products are restricted to a single environment (Tandem and IBM, respectively) and will be used for strategic high-response, high- performance systems.

Some examples of possible early DDBMS applications in addition to those actually identified follow:

- An end user with a well developed PC-based application that is ready to go multi-user might end up in a mini/PC DDBMS environment. The front end of the application is to be used at multiple sites while the core of the application is to run at the central site on the mini.

- A group of warehouses each have a mini and currently share inventory on a periodic basis. By linking the data bases the redundant data is minimized, local autonomy retained, and improved service provided.

- A shop floor scheduling and tracking system uses two minis: one mini interfaces to data collection equipment to track order status and a second mini is used for scheduling. The two minis share a shop order and manufacturing routing data base. The shop floor tracking mini uses the data bases to direct an order through the shop and to report status, while the second mini develops and maintains the order schedules.

By its basic definition, DDBMS is an integration technology. As such it will be used to provide connectivity and control of distributed/departmental systems, both for new applications and those already in place.

D

The Critical Success Factors

Exhibit V-7 summarizes seven critical success factors for getting started with DDBMS technology. INPUT recognizes that this list will expand and change as experience is gained.

1. Know and understand relational DBMS technology. If your organization does not have relational DBMS experience, get it before tackling a DDBMS application.
2. Audit your data administration function to be sure the control procedures are in place to administer distributed/departmental processing data base applications.
3. Start with a controlled experiment. Do a prototype that is not likely to have to go immediately into production and that is not tied to a critical target date.
4. Use a homogeneous DBMS environment, just one vendor's DBMS.
5. Pick an interested, informed, and technologically mature end user. Educate him and keep him informed.
6. Pick a geographically dispersed application so the data communication elements of DDBMS can be fully tested.
7. Do not select a strategic company system for the first real application. Keep it within a department or business unit.

EXHIBIT V-7

DDBMS - CRITICAL SUCCESS FACTORS

1. Know relational DBMS technolongy
2. Audit the data administration function
3. Do a controlled experiment
4. Use a homogeneous DBMS environment
5. Involve a mature end user
6. Use a geographically dispersed application
7. Select a non-strategic application

VI

Conclusions and Recommendations

VI

Conclusions and Recommendations

A Conclusions

INPUT has found DDBMS to be a viable, if very new, technology, worthy of consideration by those IS organizations pursuing a distributed processing strategy. In the next year or two it will be a commonplace offering from a number of DBMS vendors, and it is available now from a small group of leading edge vendors.

1. Status Of DDBMS Technology

DDBMS technology exists today in a workable, if incomplete, form from a limited number of vendors.

- It is based on the relational data base model and SQL.
- It is available from three leading edge DBMS vendors (Oracle, Relational Technology, Sybase) and in a more narrow version from Tandem and ADR.
- The three software vendors are porting their technology to multiple operating environments, and are packaging their relational DBMS with 4GL and other easy-to-use application tools.

The DDBMS technology is in its very early implementation by a limited number of customers. There are some plans and a few controlled experiments, but no major production applications were identified by INPUT.

The next six to twelve months will bring significant change.

- Those who have purchased the DDBMS Star capability will complete their controlled experiments and will be building their first full production applications.

- Additional capabilities will by released by vendors bringing the technology closer to the experts' definition (the twelve rules).
  - Replicated data will be supported.
  - Multi-site update will be supported.
  - More operating environments will be supported.
  - Gateways to other DBMSs will be available.
- Additional longstanding DBMS vendors will be releasing fully relational DBMS products, their offerings of distributed data base capabilities thus increasing the richness of the available software set.

2. Issues Relating To DDBMS

The issues relating to the adoption of DDBMS technology are those of any new computing technology in today's complex computing environment. The technical issues are:

- Data Administration/Modeling - While providing the means to give ease of access to data, DDBMS complicates the data administration and modeling process: what data is to reside where, how to optimize the data model, and how to provide local autonomy as well as distributed integration are all potential problems.
- Data Communications/Networks - DDBMS success will depend on a very reliable communications environment. The DDBMS software may simplify the communications task to some degree, but it will add to the demand for capacity and reliability.
- Performance - Relational DBMS software is not known for its processing performance. DDBMS adds further overhead and reliance on other sites to complete a transaction, which in turn introduces the need for additional expertise in processing optimization, planning for and maintaining replicated data, and for the system administrator.
- Security/Recovery - Data base recovery is a well known challenge. With DDBMS the recovery process becomes spread across multiple sites with transactions that may have affected the data segments at other sites. The security (access authorization) process now becomes multi-site oriented. Thus, both security and recovery are more complex processes.
- Systems Integration - DDBMS provides a means to support the linking of heterogeneous operating environments. This creates a sought after capability and a motivation to forge ahead without adequate understanding.

The IS management issues focus on the successful application of DDBMS software. With IS management already challenged to give increasing freedom to, yet keep control of, the end user, DDBMS provides a new tool in the distributed/departmental systems war. This tool allows the IS management to regain some control, but not without addressing the management issues of adopting the new technology including:

- Understanding and successfully using relational DBMS technology.
- Strengthening the data administration function.
- Learning a new set of development tools.
- New development and design approaches.
- New data administration processes.
- New approaches to user-managed development.

The end-user issues are related to their growing maturity and influence relative to computing decisions and implementations. The end user is often leading the way in many aspects of systems strategy. Thus, end users will have an influence on how DDBMS is applied.

- IS management has an opportunity to train end users in the data base development process, the relational concept, and the eventual use of DDBMS.
- The leading vendors have been concentrating on the departmental systems area. There is a potential risk that departmental users will buy the concept without proper coordination with corporate IS.

3. The Future With DBMS

It can be said that:

- Distributed data processing is a reality but that it has not provided all of the projected benefits.
- Relational DBMS is now commonplace and the way of the future, but is just coming into common use for production systems.
- Integration is destined to be the key trend of the next five years as organizations strive for more connectivity and benefit from their distributed strategies.

Recognizing these factors, it is possible to project that DDBMS has a major role to play well into the 1990s.

- It is distributed by definition.
- It is based on the relational model.
- It is capable of operating in heterogeneous environments both by definition and current commercial availability.
- It is capable of and designed to accomplish systems and data integration by:
  - Linking existing distributed systems together.
  - Providing gateways to existing, nonrelational data bases (perhaps extending the life of these prior investments).
  - Supporting a new level of distributed/integrated processing that uses computing power more effectively, provides better data control and greater service to the user, and, therefore, provides a higher return on the investment.

The next five years relative to DDBMS technology and its application should develop as depicted in Exhibit VI-1.

1987/1988

- Broader awareness, understanding, and experimentation.
- First "real"applications.
- Expanded capabilities from the vendors.

1989/1990

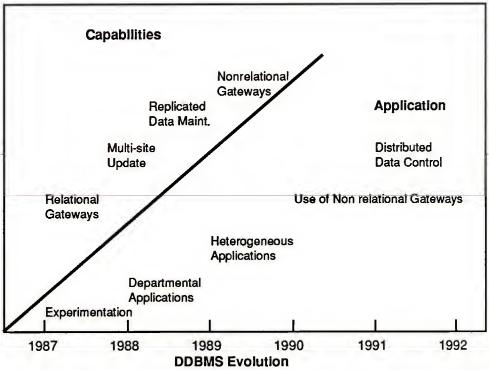
- Common use at the distributed/departmental level.
- First use of gateways to nonrelational DBMS's.
- Integration of existing distributed processing systems.

1991/1992

- Maturity of use
- Common use between corporate and departmental applications.
- A new sense of corporate data integration.

EXHIBIT VI-1

DDBMS - THE NEXT FEW YEARS



B Recommendations

Get on with relational DBMS implementation if you have not started. Do not necessarily use a strategic application, but choose one closer to a single user.

- Use the relational learning curve as an opportunity to train end users in data base design.
- Use the relational activity to strengthen the data administration function at the distributed/departmental systems level.
- Learn to use the active/integrated dictionaries.

Study the DDBMS subject, identify an experiment, and try it.

Use the available software if justifiable. Don't wait for IBM, etc., but do use it in a controlled, learning fashion.

Recognize distributed data base as a viable strategy and begin planning for its use. It may take many IS organizations two years to get to the first application, but the planning can start now.