AN ARCHITECTURAL COMPARISON OF CONTEMPORARY APPROACHES AND PRODUCTS FOR INTEGRATING HETEROGENEOUS INFORMATION SYSTEMS

Principal Investigator: Dr. Amar Gupta
Technical Advisor: Prof. Stuart E. Madnick
Researchers: Teresa Wingfield and Christopher Poulsen

Working Paper #3084-89
November 1989

Comments and suggestions should be directed to:

Dr. Amar Gupta
Principal Research Associate
Sloan School of Management
MIT Room E40-265
Cambridge, MA 02139
Telephone: (617) 253-8906

TABLE OF CONTENTS

ABSTRACT

Virtually all large organizations are faced with the problem of intelligently accessing and integrating information maintained in islands of disparate computer-based systems. Since a number of vendors claim that their respective products and services can overcome the problem of integration, a comprehensive study was made into the technical characteristics of all these products and services, as well as additional ones identified through other means. This report documents the results of this study in terms of a new taxonomy for classifying various approaches for integrating heterogeneous information systems. In addition, the principal concepts and issues pertaining to this field are also presented in this report.

SECTION ONE: INTRODUCTION

1.1. Background

Most organizations today depend on a portfolio of information processing machines, ranging from mainframes to microcomputers and from general purpose workstations to very specialized systems, to meet their computational requirements. While the individual hardware and software components in the growing array of computer systems may meet the objectives for which each was initially designed, their heterogeneity presents a major obstacle to ready access and assimilation of information. It requires tremendous amounts of time and effort to retrieve information from multiple systems today, and in many situations, it is simply not feasible to obtain integrated responses in desired timeframes.

As compared to the intra-organizational situation discussed above, the issues involved in integrating information on interorganizational basis are even more complex. First, there is greater heterogeneity in technical capabilities across organizations as compared to a single organization. Second, there is less likelihood of reaching consensus on common goals and standards. Third, non-technical issues, such as access rights and sharing of costs, acquire added importance when multiple organizations are involved. Integration of relevant information from various computer-based systems is indeed a challenging task.

1.2. Functional Deficiencies

The problem of inefficient, incomplete, and time-consuming access to information can be traced, from a technical viewpoint, to functional deficiencies at several levels as described below:

Structured and Unstructured Applications

Conventional and computer-based information systems have been designed with specific applications in view. These systems are efficient for performing the originally intended application, but they are inefficient in dealing with ad hoc queries or new unanticipated applications.

Information versus Knowledge

Traditional database systems focus on retrieval of data, and on performing elementary operations (e.g., sorting and merging). Such systems cannot use both data and programs to respond intelligently to queries like "How many vehicles can be made available to evacuate personnel from the vicinity of a disaster site within a specified period of time?" which require analysis of many factors.

Diverse types of Information

Different types of information (numerical, graphic, pictorial, speech, and video) are referenced in very different ways. During the sixties, computers were designed to manipulate, store, and retrieve numerical information. During the seventies, the focus was on textual information. Relatively less effort has so far been directed towards efficient storage and retrieval of pictorial information and its combination with numerical and textual information.

Communications

In spite of continuing effort by a number of national and international institutions, it is still difficult to transfer complex information across computers of dissimilar architectures.

Granularity

It is difficult to judge what volume of information should be made available, and how it should be arranged and tailored to meet the needs of the person requesting the information.

Security

It is necessary to determine who can access the information, and what each person is permitted to modify.

This aspect is especially important for applications involving more than one organization.

Semantics

It is difficult to specify exactly the subset of information desired without knowing details of the systems being accessed. Further, a particular piece of information may possess different connotations on various systems. Conversely, the same item (such as a particular spare part) may be specified by different numbers on various systems. The above list describes a few of the key technical problems.

In addition, there are major non-technical impediments that restrict integration. Because of the large number of issues and variables, a global solution to the problem of integrating distributed information systems has remained an elusive dream. Partial solutions are becoming available, however, and these solutions are identified and analyzed in later sections of this report.

1.3. Levels of Integration

Integration of information resources can be visualized at different levels. At the least complex end of the spectrum, information can be exchanged across a network using basic facilities such as remote procedure call [13] and message passing. At the next higher level, a new layer of software can be created to insulate users from idiosyncrasies of different host machines. This layer can provide uniformity at the level of logically centralized file servers as in the case of Andrew at CMU [2], or at the level of application programs as in the case of Project Athena at MIT [3].

The approaches described in the preceding paragraph require significant degree of user involvement. Also, they represent instances of integration of information. The focus of this report is, however, on integration of information systems. Further, the emphasis is on systems "dissimilar" from each other in terms of:

*Hardware, for example, one information system using an IBM mainframe and a second system functioning on Unisys hardware;

*Operating systems, for example, one information system using OS/MVS, and other using UNIX;

*Data models, for example, one system using network data models, and the other using relational data models;

*Database management systems, for example, one using INGRES, and the other using ORACLE database management systems, respectively.

Apart from the above differences arising from dissimilarities in hardware and software, different information systems use different sets of design decisions and underlying assumptions. For example, one system may be using annual salary of personnel while another may be using monthly salary; further, the salary in one case may be gross salary (typical of budgeting systems), whereas in the second case, it may be net salary (typical of systems used for printing checks). Similarly, a particular spare part may be specified by one part number in one system and by a different part number in a second system. All these differences must be reconciled when attempting to integrate dissimilar information systems.

There are two facets to integration--physical integration and logical integration. Physical integration, accomplished through physical connectivity, refers to the process of establishing actual communication links among disparate systems. Logical integration refers to the process of accessing disparate systems in concert for generating integrated responses. This involves reconciliation of different assumptions and perspectives embedded in the systems being integrated. At the logical level, the integration system must:

*Know where all the information is stored, along with the data formats and the query languages of the local systems;

*Decompose the query into subqueries that can be executed by local systems;

*Accumulate the results from all the subqueries;

*Reconcile differences among the results accumulated; and

*Formulate integrated answers.

All the above tasks must ideally be carried out with no modification to existing systems. However, many products discussed later in this report require significant changes to existing information systems.

Based on the level of integration achieved, and some additional characteristics, various types of integration approaches can be distinguished. These approaches, using the terminology of different researchers who have used dissimilar terms to refer to same or similar ideas, include the following:

*Cooperation System for Heterogeneous Data Base Management System [4].

*Federated Database [5,9].

*Composite Information System [6].

*Knowledge Based Integrated Information System [7,11].

*Multidatabase [8,10].

*Superdatabase [12].

Instead of attempting to cover all the views expressed by the individuals who coined these terms and phrases, the endeavor in this report is to encapsulate the main concepts and the emerging technologies and products that enable integration of dissimilar information systems.

1.4. Constraints

The design of an integrated information system is constrained by the following facts:

(a) The new system must integrate existing information systems, each of which was designed independently with little premonition, if any, that the system would form part of a larger system at a future date.

(b) Each existing information system must be fully understood in order to comprehend all underlying assumptions and semantics. Rarely are these systems fully and accurately documented.

(c) The design must allow flexibility to add or remove individual systems into the integrated configuration. As such, it requires detailed knowledge of both immediate and long-term constituents of the integrated system.

(d) The design must involve the least amount of changes to existing systems. Transactions, irrespective of whether they are queries or updates, involve retrieval or manipulation of information maintained in individual systems, each of which imposes its own set of requirements and standards which must be reconciled by the new system.

The goal is to mitigate all the above-mentioned problems in an efficient and effective manner.

1.5. Models and Schemata

A Distributed Heterogeneous Database Management System attempts to meet the above goal by providing the facilities to access, to aggregate, and to update information maintained in multiple, distributed, heterogeneous systems. A typical distributed heterogeneous approach is characterized by the following:

a) The use of a common data model, with appropriate mapping to data models of participating systems;

(b) The support of a standard user language for retrieving data on a global basis in all cases, and also for storing data in some cases;

(c) The ability to decompose a user request into an equivalent set of subqueries expressed in the format and manner required by participating systems;

(d) The capability to resolve incompatibilities in semantics, data structures, and other parameters;

(e) The mechanism to take the individual pieces of information provided by participating systems and to integrate them in order to provide a cohesive response to the original user request.

In addition to the above, some approaches facilitate implementation of sophisticated features such as authorization control and the ability to draw inferences in situations involving missing or conflicting data.

When data need to be retrieved from a single file, it is sufficient to simply know the file design including the keys for accessing records. One single view or schema of the file design meets the requirements of the systems analyst and the user. However, when there are several disparate information systems and there are many users from different disciplines, each user tends to look at the information from his or her perspective. By having more than one schema, it is feasible to handle complex problems, as well as to achieve data independence.

Most experts recognize the requirement for three types of schema. This notion of a three-schema architecture originated in 1977, when the Study Planning and Requirements Committee (SPARC) of the

American National Standards Institute (ANSI) published a report on this subject [14]. The three types of schema are as follows:

(a) Global conceptual schema: This schema is intended as a blueprint that shows the existence of all the information maintained in all the information systems. To enable intelligent access and integration, the conceptual schema must be designed to incorporate the semantics of all stored data. For this reason, it is sometimes referred to as logical schema. There is only one global schema for the entire enterprise.

(b) Local or internal schema: This schema covers the data in an individual information system, that may itself be hosted on one or more computers. Since there is one internal schema for each information system, a complex environment necessarily implies a large number of such local data models. The collection of such local models is referred to as a local schemata.

(c) External or application schema: This schema describes the user or the application view of the distributed data. Since each class of users or applications will view the data in a different manner, more than one external schema is required. This collection is referred to as external schemata.

The strategy used to map information between these three types of schemata determines the overall capability to mediate across dissimilar architectures, to maintain integrity of data, and to meet user expectations. Transformation of information between the three schemata is ideally managed via a three-schema data dictionary. Use of very powerful data definition language is critical in defining the global schema in terms of objects, events, and integrity constraints [15].

It is important to emphasize that many commercial products and approaches do not support the concept of a global schema. Instead, a relational table or matrix is established between the required external schemata and the available internal schemata. Such a two-schema approach, referred to as the interfacing approach, offers no technological control over data integrity [15]. Data are exchanged either through transaction passing or by controlled redundancy. In the former case, all transactions are transmitted from the source database to the destination database, where they are transformed to adhere to the requirements of the external schemata. In controlled redundancy, instantaneous copies of the source database are stored at multiple nodes of the network, and these copies are replaced by new versions at periodic intervals. Both alternatives of the interfacing approach attempt to meet new user and application needs in an ad hoc manner, without any consistent set of rules governing the overall computational environment.

To overcome the above problems, it is necessary to use a three-schema approach, especially in applications of a critical or a strategic nature. All the major research programs funded by the U.S. Government, such as Multibase, Integrated Information Support

System, and Integrated Manufacturing Distributed Database Administration System, adopted the three-schema approach. Comparative evaluation of eight major initiatives in this area appears in [16,17]. To maintain adequate integrity of shared data of critical and strategic importance, the need for utilizing a three-schema approach cannot be overemphasized. This aspect acquires added importance when an application needs to access data from various DBMS-based information systems, as well as from earlier file systems that contain no notion of DBMS.

1.6. Languages

The only database language that has gained endorsement from major computer manufacturers, and from standards establishing organizations such as ANSI and ISO, is SQL. SQL statements can be used to create tabular views that approximate the notion of an external schema. Further, it can be used to create a conceptual schema, although a semantically weak one [18]. Intended to be used for queries, SQL has no internal schema.

There are several flavors of SQL in existence. A number of vendors claim to offer support for multiple versions of SQL. Further, some SQL based integration approaches can deal with legacy systems that offer no support for SQL. In such cases user queries expressed in SQL are converted into the native languages of the participating information systems.

There is a recent trend towards using graphics techniques to select databases, files, and tables. From a user perspective, these icon-oriented approaches are far friendlier than traditional language oriented techniques. However, there is still no common graphics interface that has been endorsed by a critical mass of vendors. Even with a proprietary graphics interface, navigation with icons may be especially relevant for users at the top echelons of the organization.

1.7. General Direction

The basic concepts outlined above have been used in evaluating the principal products and approaches that are commercially sold. The list of vendors contacted is placed at Appendix A, and detailed specifications are tabulated in Appendix B. For the subset of products and approaches that hold some degree of relevance, a brief write-up was generated on each product. These write-ups are presented in Section 2 of this report. In Section 3, a new taxonomy is created for classifying different approaches to integration, and recommendations are made for selecting an appropriate set of technologies.

SECTION TWO: HIGHLIGHTS OF PRINCIPAL PRODUCTS AND APPROACHES

In Section 1, the fundamental principles of integrating dissimilar information systems were presented. In this section, principal products and approaches are discussed and analyzed using these fundamental principles.

The products and approaches described in this section were derived from numerous sources, including articles in computer journals and magazines, proceedings of relevant conferences, lists of conference participants, and personal discussions with representatives in various companies. From the initial list generated, the systems previously covered in References [1] and [16] were deleted, unless the complexion of the system had changed significantly during the intervening period. The material presented in this section is not an exhaustive survey of all issues, but rather a discussion that covers the whole field taking all significant design approaches in view. It should be emphasized that some companies offer customized services, rather than a product, making it difficult to evaluate them in detail. In this section, each product is described by an introduction, its functional architecture, and its salient features. Additional details are tabulated in Appendix B.

2.1. BBN Systems and Technologies: Cronus

Cronus is an environment designed to support coherent integration of heterogeneous computer systems. The objectives of Cronus are: to provide an effective, general-purpose, distributed computing environment composed of hosts with differing hardware and operating systems; to support an integrated computing environment consisting of both the native operating system environment and the Cronus environment; and to provide comprehensive support for development of large-scale distributed applications. Cronus began as a research project funded by the Rome Air Development Center and the Naval Ocean Systems Center in 1981. The project was a two-year joint effort between BBN Systems and Technologies Corporation and Xerox Advanced Information Technology, both of Cambridge, Mass. The first prototype from this project became operational in 1984.

Cronus supports both object-oriented and relational local data models. Global updates may be performed on an object by object basis. Update operators for insertion, deletion, and modification are in algebraic form and operate associatively on sets of tuples. Stored update procedures may include conditional statements and expressions.

FUNCTIONAL ARCHITECTURE

Cronus runs on top of existing operating systems on various hardware platforms. The Global Schema definition is supported for object-oriented data storage and retrieval. Every system resource is an object and is accessed through operations defined by the type of the object. Cronus supports heterogeneity by serving as a bypassable layer of abstraction between applications programs and native operating systems. Through this approach, both conventional access to native operating systems resources and services and access to a coherent, uniform (object-oriented) system interface, regardless of computer system base, are supported.

Distributed query processing and update capabilities require translation of high-level requests into sequences of algebraic operations on data stored at possibly different sites. The algebraic approach defines a collection of "physical" operators on tuples and takes into account the sort order, the presence of indexes, and the physical location. Transformation rules determine when two algebraic expressions are equivalent, and optimization rules specify when it is profitable to transform an expression into another equivalent one.

Salient Features

Cronus incorporates an evolutionary approach in its design. Cronus claims that its Open Architecture approach makes it possible to upgrade or replace components or technology with little or no impact on the rest of the system. There is uniformity across diverse operating systems, hardware platforms, and applications.

2.2. Booz Allen: IDIMS

IDIMS is an expert system oriented approach designed by Booz Allen to provide automated data access. The user obtains access throughout the distributed data environment with a single global manipulation language and a user-friendly, intelligent interface. Upon receipt of a query, IDIMS decomposes the query into a set of database calls and function applications; determines security violations on databases; formulates queries in appropriate languages and queries databases; applies functions to retrieve data and finally displays the data via a state of the art graphic interface, in a style appropriate to the type and the number of data elements to be displayed.

It is important to note that IDIMS is not a commercial "off the shelf" product, but instead a conceptual framework used by Booz Allen to achieve automated data access. Although a complete production system has not yet been implemented, Booz Allen claims that it has built systems which relate to each functional component of IDIMS.

FUNCTIONAL ARCHITECTURE

The IDIMS system consists of a Query Processor, a Knowledge base for database models and function models, a Query Language Generator, and a Natural Language Processor.

The Query Processor translates a global data manipulation language query into a series of data and function calls using the database and function models knowledge bases. In doing so, it plans the appropriate databases and functions to apply, chooses between plans using query optimization and time consideration techniques, and executes plans by calling the Query Language Generator to retrieve data. In addition, it determines whether there are potential security violations for the local databases queried.

The Data Base Models Knowledge Base is a structure which maintains data models for each of the distributed local databases. These models contain information and rules on security consideration, representative query language, data, data type, and host computer.

The Function Models Knowledge Base is a structure which maintains knowledge about the available procedures to alter or determine data attributes. Each function model provides for a specific logical data linking procedure such as correlate, sum, and join, and describes domain, range, and process definition. The Query Processor uses the Function Models Knowledge Base to determine which functions satisfy given constraints.

The Query Language Generator translates the global manipulation language into the appropriate local database query language. It then makes these queries and retrieves the data. Accordingly, the Query Language Generator maintains translations for each local database language and procedures to send queries to each computer on the network.

The Natural Language Processor provides an additional interface through which individuals can use limited natural language capabilities to query databases. This processor then translates the natural language into the IDIMS global manipulation language using conceptual dependency and case grammar.

SALIENT FEATURES

IDIMS is not a product, but an approach that eliminates the need for the user to know multiple query languages, locations of data, database paradigms, and manipulation schemata. A high-level, iconic, global query language is supported; however, updates are not supported.

2.3. Cambridge Technology Group: SURROUND

The Cambridge Technology Group (CTG) of Cambridge, Mass., specializes in providing customized distributed heterogeneous data integration within short time frames. A small group of users and managers from the client company interface with SURROUND developers at CTG to implement software interfaces as well as a rapid prototype of a functional application within a one week period. The goal of CTG is to construct a user-friendly system which focuses on one particular strategic application.

CTG applications support queries and updates. Local data models are supported via SURROUND interface and coordination tools. Complete data transparency is provided for geographically dispersed locations via dialup ASCII modems.

FUNCTIONAL ARCHITECTURE

The UNIX operating system is used as a major part of the SURROUND service to resolve incompatibilities of different systems. Software tools access various distributed environments, extract information from individual programs, and then transfer data to the UNIX environment where the information is joined together. Talk Tools are communication tools which imitate the code that the host computer understands. The host computer accepts commands from the Talk Tool and sends results back to the same Talk Tool in the UNIX environment. There are also a variety of tools for attaining connectivity, security, network management, and for gauging performance.

The construction of the global schema is tailored to the needs of the application. SURROUND does not define a data manipulation language; it supports the interface currently used by the user. SQL formatted files are accessed with modified versions of SQL Data Handler and communication tools on remote system perform global updates. Commander, a proprietary tool, handles transaction management functions.

SALIENT FEATURES

Rapid prototype systems have been developed for TRANSCOM, General Electric, and other organizations within one week spans of time.

CTG claims that it can provide production software in as little as three months. Record structures can be changed easily and new databases can be added quickly, making the SURROUND environment flexible.

2.4. Cincom: SUPRA

SUPRA, a product of Cincom, provides a commercial implementation of the ANSI/SPARC Three Schema Architecture. Such an architecture, described in Section 1.5, insulates applications from changes in the physical storage and conceptual database, reduces development and maintenance costs, and increases the useful life of applications.

FUNCTIONAL ARCHITECTURE

The SUPRA system integrates distributed heterogeneous databases and manages distributed processing by having a SUPRA kernel reside locally within each host machine or database. This has the effect of making the hosts aware of the integrated database network and in essence creates a "standard protocol" through which the databases can communicate.

The SUPRA kernel consists of two basic engine components, the Distributed Relational Data Manager (DRDM), and the Heterogeneous Data Management Processor (HDMP), along with a portion of the Global Directory. Using the kernel on each local host, the requesting node coordinates the entire unit of work, including the decomposition of the query into separate neutral protocol requests and the determination of where to send them. The local/receiving nodes, each with its own kernel, translate the request into the appropriate local data manipulation language, gather the data, perform calculations, and return the result to the requesting node.

In this decentralized structure, each node is intelligent and has the ability to coordinate with each other via knowledge of the network in the resident Global Directory. This approach to distributed database and processing management assumes that most of the data accessed by any given node reside locally and that an "independent" node, with the ability to process locally, has significant speed and reliability advantages. In those instances where the data are not local, the Global Directory contains the information to tell the DRDM where to get them. As such, the network is not dependent on the reliability of a central Global Directory or engine. If any one node goes down, only the data residing at that node become inaccessible, not the whole system. Additionally, by distributing the necessary processing information, every transaction need not access a central directory, thus eliminating a potential bottleneck. However, with more than one copy of the Global Directory, the SUPRA approach implies more maintenance and overhead compared to a single copy approach. Every time the Global Directory is updated at one node, the update must be reflected across the entire network creating a "cascade" of transactions throughout the network.

SALIENT FEATURES

SUPRA currently provides for the use of synonyms and the resolution of scale conflicts. The decentralized approach allows a synonym table oriented to each user's particular application and working language. The Global Directory contains the ability to retain rules for performing scale translation and calculation before data integration. A unique feature called "triggers", alerts the user to the possibility of a data conflict. For example, a trigger could be established to ensure that two numbers, before they are combined together, be within a certain percentage of each other. At execution time, when a user wishes to combine the two numbers, a violation of this requirement would lead to an interrupt, and the user to be prompted about the problem.

SUPRA utilizes a two phase commit procedure which provides good concurrency control. While assuring secure updates for SUPRA or other database management systems which support two phase commits, the product currently only allows updates from join views to distributed databases which use SQL.

2.5. Cognos Corporation: POWERHOUSE

PowerHouse consists of four components: PowerHouse 4GL, PowerHouse StarBase, StarNet and StarGate. PowerHouse 4GL is a fourth-generation language; PowerHouse StarBase is a relational database management system; StarNet is network routing software; and StarGate is a gateway provider to other supported local database management systems. The PowerHouse 4GL was developed by Cognos Corporation, Ottowa, Ontario while the remaining components were developed by InterBase Software Corporation, Bedford, MA. Cognos began to market the Relational DBMS of InterBase and its distributed technology in conjunction with the PowerHouse application development language in 1988.

FUNCTIONAL ARCHITECTURE

In this approach, there is no concept of a single global schema. Instead, multiple relational schemata may be referenced, even from within the same program. Relational operators are utilized to selectively access and merge pieces of data. Research on a distributed query engine is currently in progress.

A discussion of the capabilities of InterBase components is provided later under its own prototype overview. The main difference between PowerHouse and InterBase is global data manipulation language supported. Both products support SQL and Groton Data Manipulation Language, but the stand-alone PowerHouse 4GL is unique to PowerHouse.

One of the benefits of the PowerHouse 4GL is that it provides access to relative Record Management Service files and network databases in addition to sequential and indexed Record Management Service files and relational databases supported by InterBase. The Open Systems Relational Interface (OSRI) is a relational database protocol. PowerHouse 4GL elements for transaction processing, report generation and batch updating use a relational access approach which makes it possible to embed OSRI calls into the elements. This means that as more local data models are added to the configuration, it is possible to develop applications which access them using PowerHouse.

The PowerHouse 4GL product set includes PhD master data dictionary. Using PhD, it is possible to treat all meta-data as if they were resident in a single logical dictionary. Thus, one need not redefine meta-data in the PhD dictionary.

SALIENT FEATURES

The salient features listed in the description of the InterBase prototype also apply here. Cognos is attempting to incorporate expert systems and object-oriented techniques in the PowerHouse product.

2.6. Control Data Corporation: ASCENT

ASCENT is derived from the Technology Information System (TIS) developed by Lawrence Livermore National Labs (LLNL). Work on this system originally commenced in the early seventies. CDC has a technology transfer license with LLNL (actually with the University of California) to support this technology, and to take it into production.

ASCENT is a suite of software products that provides communications, networking, office automation, and database services within an interactive environment that is common to all users. The *gateway portion of ASCENT supports a powerful interpretive scripting language that can simulate an interactive user when connected to a heterogeneous remote computer. This gateway serves as a tool that can be programmed with the scripting language to log into databases and to extract data from a broad range of remote systems.

FUNCTIONAL ARCHITECTURE

A collection of software modules, known as the Intelligent Gateway Processor (IGP), use a UNIX system as the link between menu type front-end systems and geographically and technically disparate mainframes. The IGP front-end program, Integrated Information System (IIS), offers menus composed of descriptor files which might consist of classes of information resources, specific connections to remote hosts, local database options, or available system utilities. In addition to menus, IIS understands a command set that overlaps with the UNIX Shell command set. As IIS is implemented in a script-driven interpreter, UNIX commands can be recognized even without arguments, and prompts can be enabled for arguments or options.

The Network Access Machine (NAM), an extended implementation of the specifications published by the National Institute of Standards and Technology, utilizes a variety of communication methods to provide access to other computer systems. Currently NAM can interface with direct asynchronous connections between two hosts, most autodial modems, TCP/IP networks and X.24 networks. NAM uses an interpreted communication language to specify the logic necessary to establish connections and conduct dialogues with remote systems. This language includes structured constructs, extensive pattern matching capabilities and capabilities to insert "filters" in an existing connection. These filters are specialized programs written to modify the appearance of the data.

In a heterogeneous environment, ASCENT*gateway must be driven by the "host UNIX system" database management system to extract specific data elements from their remote computers under script language control. The DHDBMS places these data elements into the proper place in the "logical record" requested by the user. This approach requires no software to be added to any of the remote systems since each system feels that the scripting language interactions with other remote systems are identical in appearance to interactions with its own users.

SALIENT FEATURES

ASCENT*gateway allows automated access and connectivity to local or remote computing resources. ASCENT*mail provides complete electronic mail and message services within a computer network including addressing, forwarding, and mail file management functions. A menu-driven user interface is available to access integrated software applications, utilities, and system administration tools.

2.7. Control Data Corporation: CDM*PLUS

The concept of CDM*plus was developed and prototyped with direction and funding from the Integrated Information Support System (IISS, pronounced I^2S^2) initiative of the U.S. Air Force. CDM*plus defines and categorizes data using the ANSI/SPARC Three Schema Architecture discussed in Section 1.5.

FUNCTIONAL ARCHITECTURE

The CDM*plus dictionary/directory is loaded and maintained via the Neutral Data Definition Language (NDDL) which describes all meta-data to the ANSI/SPARC based three schema structure. NDDL is SQL based and describes objects with verbs such as create, define, alter, and drop. Each schema has unique commands. For example, in the internal schema, the NDDL can describe the three basic types of database models. NDDL provides inter-schema mappings using the conceptual schema or the data model as the base. NDDL is a human and machine readable language that can be used in either in interactive mode or in batch mode, as well as on forms.

The mapping strategy between the schema allows data to be accessed without regard to the location or the storage mechanism. For example, an application can join two relational user views from the external schema that each map to two entities. These entities in turn can be mapped to multiple databases and database types on multiple computers without burdening the user or developer with all of the details of the actual conversions, translations or mappings.

This independence is critical to provide a buffer for life cycle maintenance as applications and storage structures change, and to provide the capability for full relational operation (including outer joins and referential integrity) between heterogeneous databases.

Queries can be embedded into current applications written in Cobol, C, or Fortran. These queries are precompiled by CDM*plus using the external schema view and transformed into actual database queries. All code necessary to access the databases is generated including DBMS precompilations. Since the base schema is the conceptual schema or data model, domain constraint checking and referential integrity checking can be automatically enforced based on the conceptual schema. After the code is generated, it is transferred by the Code Manager to the appropriate computers where local compiles, DBMS precompiles, and links are accomplished.

All query processing is controlled by two CDM*plus components: the Network Transaction Manager (NTM), which controls unique process to process communication capabilities that are necessary in the distributed environment, and the CDM*plus Distributed Request Supervisor (DRS), which controls the actual queries on each host and formulates the messages sent to the NTM. This approach allows serial or parallel processing of queries. In the serial process, one query produces input for other queries.

Each local node contains an NTM and multiple CDM*plus runtime (CDMR) modules to facilitate distributed processing. The CDMR incorporates DRS, Aggregators, File Utilities and Code Manager. Query modules in applications that attempt to integrate information from multiple heterogeneous databases can be replaced by NDML statements. This allows for easier maintenance of legacy applications because changes to the systems only require re-precompilation to generate all the necessary code.

SALIENT FEATURES

CDM*plus requires the user to model his or her enterprise and to load that model into the CDM. The three schema architecture provides one global conceptual view of the complete data. In addition, it provides independence between the actual physical storage of the data, the users' view of the data, and the conceptual view of the data. Modeling and integration of applications and databases with CDM*plus is done in an evolutionary manner, allowing for the implementation of one application at a time rather than a full rewrite of the entire system.

2.8. Data Integration, Inc.: Mermaid

Mermaid, a trademark of Unisys, is an integrated data access product marketed by Data Integration, Inc. It allows users of multiple relational or network database management systems, running on different machines, to access and manipulate data using standard SQL, even when the participating systems don not support SQL. Integration may require translation of the data into a standard data type, translation of the units, combination or division of fields, and encoding values. Only updates to a single database are presently supported. Updates to replicated and fragmented relations may cross databases, but the updates to all databases are not made concurrently.

The Mermaid system is designed for an environment in which there are many databases that are independently administered and where it is not feasible to standardize on one DBMS. With Mermaid, new applications may be developed either at the global level when distributed access is required, or above an existing old database. Local administrators grant access permissions to users of Mermaid and then Mermaid adds an additional level of access control.

FUNCTIONAL ARCHITECTURE

The major processes are: User Interface Process, Distributor Process, DBMS Driver Process, and Data Dictionary/Directory (DD/D). The User Interface Process contains the SQL parser, the query validator, the query library, and the report generator. The Distributor Process contains the optimizer and the controller. The optimizer plans query execution and the controller starts processes and coordinates execution. There is one DBMS Driver Process for each database to be accessed. It contains the schema translation, the language translation, and the DBMS interface. The Data Dictionary/Directory (DD/D) is a database that contains information about the schemata, users, host computers, and network. It supports three layers of schema definition.

A user input is either a command to execute within the user interface or an SQL query. The SQL query is parsed, validated, translated into Distributed Intermediate Language (DIL), and sent to the distributor process which performs the Distributed Execution Monitor function. The optimizer plans the execution by consulting the DD/D for the location of the relations. It sends commands and subqueries to the controller for transmission to the DBMS drivers. The driver reads a message which may be a command to initialize, terminate, execute schema and data of the query from the global form into the local form and then translates it from DIL into the local DBMS language. The first time that a relation is retrieved, the data are translated into the global form.

SALIENT FEATURES

Mermaid treats files as large objects. Qualification is done on structured data. The user selects qualified objects to view. The system then starts a retrieval process and a display process that are customized for the file type.

Communication between processes is done using standard protocols. The main processes use TCP/IP on an Ethernet. If TCP/IP is not available on the computer with the DBMS, another protocol may be used to communicate between the controller and the driver or between the driver and the DBMS.

Mermaid is written as a set of processes with well defined interfaces so that it can be customized for different environments.

It can also be embedded in or called by other products and it can use other products to process the output.

2.9. Digital Equipment Corp.: Rdb/VMS,DBMS

Digital Equipment Corp. (DEC) offers two alternative methods for distributed data management in the VAX/VMS environment. One method is geared towards a decentralized environment with users having their own processing resources and connected through a Local Area Network. The other method is directed towards centralized distributed processing with a VAX cluster comprising two or more VAX processors, and the database located on a common disk. Each of these methods can be used with either of the two DEC data management systems: Rdb/VMS is a relational DBMS while Digital's Database Management System (DBMS) uses a hierarchical/network model. Further, VAXlink Client is a layered VMS software product that provides unidirectional IBM mainframe database access and extract capabilities; it works in conjunction with VAXlink Answer/DB Extractor to access IMS and VSAM data structures and to transfer extracted data to Rdb/VMS databases.

FUNCTIONAL ARCHITECTURE

VAX SQL is Digital's version of ANSI SQL and is used as an interactive and software development interface for relational databases that conform to Digital Standard Relational Interface (DSRI). Used to define, access, and update relational databases, it is characterized by an interactive data manipulation language, a data definition language utility, preprocessors and dynamic SQL for COBOL, FORTRAN, C, and PL/I. DEC's Database Management System uses the Network Manipulation Language (NML), a standard for network models.

A Database Control System (DBCS) provides the connections between the application program and the database locations for the Data Base Management System. Application programs transmit information to the DBCS. The DBCS transfers all information requested by a call for data to the User Work Area (UWA). Information in the UWA is then used to update the database.

The VAX Distributor distributes data from any Rdb/VMS, VIDA (an access tool for IBM databases), or any other database conforming to DSRI. The commands of the VAX Data Distributor allow the user to specify the data to be distributed, the type of copy to be made, the target node, and the transfer and update schedules. Three kinds of copies are permitted by this tool: extractions, extraction rollups, and replications. These copies effect only the target database.

Digital's distributed, active, open-architecture dictionary is the VAX Common Data Dictionary/Plus (CDD/Plus). VAX CDD/Plus provides one logical dictionary system that is physically dispersed, thus eliminating the need to develop separate dictionaries to support proprietary tools or applications.

SALIENT FEATURES

Digital claims to provide one of the best relational database management systems for the VAX/VMS environment. VMS is a widely used operating system for development and execution of a wide range of applications. Apart from the above products marketed directly by DEC, a number of other vendors have opted to select VAX hardware as the host system for reconciling the idiosyncracies of individual components in a distributed, heterogeneous environment.

2.10. Gupta Technologies, Inc.: SQL SYSTEM

Gupta Technologies, Inc. of Menlo Park, Calif., supports relational database networking. The SQL SYSTEM allows applications running on a PC to use data residing on minis, micros, and mainframes. The primary elements of the SQL SYSTEM are: SQLBase, SQLWindows, SQLTalk and SQLNetwork.

SQLBase is a distributed relational DBMS for PC networks. It can be used as a standalone DBMS, as a database server for personal computer based workstations on a network, and as a gateway to mainframe database management systems. The SQL language syntax of SQLBase is fully compatible with those of IBM's DB2 and SQL/DS. Gupta Technologies claims to be the first company to offer advanced concurrency, automatic consistency, and dynamic recovery for a distributed relational DBMS on a personal computer.

SQLWindows and SQLTalk allow the programmer or user to utilize SQL without having to develop and compile programs. SQLWindows is an advanced applications generator which supports graphics, menus, dialogue and help boxes and mouse pointing devices. It contains a complete procedural language and provides exits to other programs. SQLTalk is an interactive data management program for accessing SQLBase via the SQL language.

FUNCTIONAL ARCHITECTURE

The SQLNetwork components enable mini, micro and mainframe SQL database management systems to appear like distributed database servers on the same network. The three components comprising the SQLNetwork are: the SQLRouter, the SQLGateway, and the SQLHost.

Requests for data are handled differently based on the location of the databases. The SQLRouter is the software that directs SQL requests to databases within the PC network. SQLRouter must reside on each PC and requires NetBIOS on each workstation. The SQLGateway routes SQL requests to databases outside of a single PC network. It runs as a memory resident program on the PC/AT and accesses a file containing the databases to which it can establish a connection at load time. This file also contains the physical node location of these databases.

SQLHost software resides on the host computer and transforms the host DBMS into a database server that accepts SQL requests routed to it by SQLGateway. It translates SQL commands into a format acceptable to the host DBMS and forwards the information for processing. The results are then sent back to the network. A different version of SQLHost is necessary for each different SQL DBMS because of variations in SQL and application program interfaces.

SALIENT FEATURES

SQLNetwork was designed to allow applications to take advantage of the capabilities of a personal computer, to transparently access databases that reside on diverse servers, and to serve many users with a large amount of data. The user is provided with full cooperative processing capabilities in DOS, OS/2 and UNIX environments. SQLNetwork provides access to DB2 and development is underway for SQLHost Software to support SQL/DS, Oracle, Ingress, and Sybase.

2.11. Honeywell Federal Systems, Inc.: Distributed Database Testbed System

DDTS is a testbed emphasizing modularity and flexibility for distributed heterogeneous database management. The project began at Computer Sciences Center (CSC) of Honeywell with the aim of designing and building a system that could be utilized for evaluating alternate approaches and algorithms for managing data stored in several databases on different platforms. Although DDTS is not a product, Honeywell utilizes this technology in specific contracts.

FUNCTIONAL ARCHITECTURE

DDTS supports a five-schema architecture. The internal schema is divided into the global representation schema, the local representation schema and the local internal schema. In addition, there is an external schema and a conceptual schema which follow the terminology of Section 1.5. A bottom-up approach is used to design a global schema for multiple databases. First, a local representational schema is built from the description of the local database. Next, all local representational schemata are integrated to form the global representation schema. Finally, the conceptual schema describing semantic integrity constraints to be enforced by DDTS is constructed.

The DDTS system architecture is based on a model of distributed computing in which a set of abstract processors execute sets of concurrent processes. Local and remote processes communicate via message exchange. DDTS consists of a set of Application Processors (APs) which control user interfaces and manage transactions, and Data Processors (DPs) which manage data.

The GORDAS Command Interface (GCI) enables multiple retrieval and update requests expressed in GORDAS, a high-level query language. Before a GORDAS transaction is executed, the GORDAS requests are translated to an internal representation for processing by the Translation and Integrity Control (TIC) module. The internal representation possesses features of relational algebra and relational calculus, and is similar to an internal representation of SQL.

Materialization and access planning (MAP) result in a strategy for processing distributed queries, updates and transactions. The response time (the time elapsed between the start of the first transmission and the time at which the result arrives at the required computer) and the total time (the sum of the time of all transmissions) are minimized with MAP.

The Distributed Execution Monitor (DEM) performs distributed execution of a transaction as directed by MAP. Based on status information returned by the Local Execution Monitors (LEMs) during execution, the DEM decides to commit or abort a transaction. DDTS uses a two-phase commit protocol during transaction execution. To resolve concurrency control conflicts, the transaction is rolled back and then restarted.

The Local Operations Module (LOM) of each DP acts as an interface between the LEM and the local DBMS. LOM provides local translation and optimization of subtransactions that have been assigned to the DP by MAP. The LOM translates subtransaction requests in the internal representation to data manipulation requests in the language of the local DBMS.

SALIENT FEATURES

DDTS is a useful tool for evaluating approaches to the various problems associated with distributed, heterogeneous database management.

2.12. IBM: DB2, SQL/DS, SAA, and others

IBM faces a challenge in developing its integrating approach. It commands a major presence in the mainframe computer market and in the personal computer market. It can either come up with design approaches that enable integration of information hosted on its own systems, or it can develop approaches that encompass non-IBM architectures as well. So far, the emphasis has been on the former alternative.

The IBM System Application Architecture (SAA) serves as a framework for developing consistent applications across three IBM computing environments: System/370, AS/400 and PS/2. SAA is a collection of software interfaces, conventions, and protocols, and includes support for distributed files and distributed databases that are resident within one of these environments, and conform to certain design guidelines.

IBM is developing a new relational Unix oriented approach that would integrate their AIX line (System/370, RT, and PS/2) as well as communicate with their SAA database line (System/370, AS/400, and PS/2)). Details of this approach have not yet been released by IBM.

At a lower level, IBM provides support for linking information from distributed relational (SQL/DS) and hierarchical (IMS) local database management systems. In addition, its Data Extract product (DXT) can extract data from DB2 tables, SQL/DS tables, DL/I data sets, VSAM data sets, and physical sequential data sets, and convert them for loading into other DB2 tables, physical sequential data sets, and CMS files. DXT users can also write generic data interfaces for accessing non-IBM (OEM) databases on MVS.

In October 1988, IBM upgraded DB2 and SQL for distributed environments. The new DB2 and SQL (called SQL/DS) use relational concepts to integrate information. In either case, because a central DBMS is not assumed, each location has all the capabilities offered by a local relational DBMS and may participate in an overall distributed network as well. Local functions are not affected by the role of the system in distributed processing. Flexibility is provided for deciding how data will be distributed and processed. The SQL data manipulation language can access remote tables or a user may work with a local copy. Also, data may be stored in one large table or split into several parts to accommodate data which may be accessed and updated only at certain local locations. Extracts, snapshots, and data replication are various types of copies which enhance retrievals and updates.

A user may extract a copy of one database table and load it into another database. Extracts are useful for migrating data from one data format to another and for maintaining read-only copies of a subset of data that rarely changes. Snapshots are read-only copies of tables that are automatically made by the system on a periodic basis specified by the user. These are useful for locations that need an automated process for receiving updated information on an occasional basis. With the data replication support, any location may update any copy of a table, and these updates are sent to other locations automatically. This is effective for situations that require high reliability and quick data retrieval with few updates.

FUNCTIONAL ARCHITECTURE

Distributed relational data are accessed either on a remote unit of work or on a distributed unit of work. Remote unit of work permits access to a single, remote relational DBMS within a unit of work. A unit of work is a set of one or more related requests which must all complete successfully or not at all. An application program can query and update data from any number of tables residing in a single DBMS within a unit of work. In addition, application programs can access data in multiple database management systems one at a time, in separate units of work. SQL/DS and DB2 have remote unit-of-work capability as well as support for single SQL requests.

Distributed unit of work provides access to multiple relational database management systems within a single unit of work. An application program can read and update data from any number of database management systems within a single unit of work. However, a single SQL statement must refer to data from a single DBMS. Key features of distributed unit of work are: complete location transparency, coordinated system security mechanisms to determine whether the application is authorized to access the remote systems and data, and the two-phase commit protocol to ensure termination consistency. DB2 supports distributed unit-of-work.

SALIENT FEATURES

IBM states that its leadership in stand-alone database management products resulted from its advanced data security, data integrity, and data recovery capabilities. Its objective is to apply this expertise to the implementation of its distributed relational products. DB2 and SQL/DS have referential integrity which ensures the consistency of data values between related columns of two different tables.

2.13. Information Builders, Inc.: FOCNET

FOCNET, derived from FOCUS Network Products, is marketed by Information Builders, Inc., headquartered in New York City. Its nonprocedural language and DBMS is called FOCUS. Various FOCUS interfaces provide logical access to data in hierarchical, relational, and network models on remote machines.

FOCNET permits multiple individuals to update a database simultaneously. A change/verify protocol confirms that the record has not been updated by another user before it modifies the record. In cases where the record has been changed, the update is reapplied against the updated version. Also, this approach allows a user to define commit and rollback specifications for transactions.

FUNCTIONAL ARCHITECTURE

A client/server model layer is inserted on top of the operating system and the access-method architecture of each participating node. Its three components are client software, interfaces to network transport software and server software. This approach reduces network traffic and takes full advantage of available CPU resources.

Most local databases must have a Master File Description (MFD). All files must be described in the FOCUS Master Dictionary before FOCUS can access their data. The master dictionary provides much more than conventional file and data information. FOCUS file descriptions can also contain security restrictions, field aliases, data validation criteria, predefined relational join operations, and multiline column titles for use in reports.

FOCUS supports 4GL FOCUS, third generation languages, and ANSI SQL global data manipulation languages. Read/write interfaces generate the appropriate dialect of SQL for many well-known relational database engines which use variations from this standard data manipulation language.

Dynamic relational join operations permit linkage to up to 16 different tables and databases at one time. Processing is optimized by including record-by-record operations and set processing based on the functionality of the database. In addition, balance tree indexing accelerates the speed of data retrieval and online updates of large databases. Users can logically invert FOCUS databases to further optimize retrievals.

SALIENT FEATURES

FOCNET runs on many platforms under a variety of operating systems. The list of local database management systems supported is impressive as is the list of supported network services. FOCUS is the most widely used 4GL in the world. It is suited for developing software and provides for automated application development and maintenance.

2.14. Interbase Software Corporation: INTERBASE

Interbase is a product of Interbase Software Corporation, Bedford, Mass. Although Interbase operates in a variety of software and hardware environments, its Digital Standard Relational Interface (DSRI) compatibility makes it most suited for use with other DEC DSRI-compatible tools. Relational local data models, indexed and sequential Record Management Service files, and UNIX stream files are supported by Interbase.

Global updates to databases can be performed in the same interactive session, as Interbase allows multiple simultaneous transactions per process and provides for two-phase commits. These two features can be used to ensure that a transaction which updates more than one database will complete correctly or not at all. Partial commits and partial rollbacks are possible as well.

Interbase uses multigenerational data structures for data storage. Records are never updated, but only added. Old versions of the record are deleted by the next transaction which accesses the record. This means that update access to data being processed by another user is allowed. On actual update, Interbase will impose the record it originally retrieved with the record it is trying to update. If they are different, the transaction is rolled back and the user is informed of this fact.

Changes required to local databases are minimal. The Rdb database of DEC permits access to Interbase files through VMS. The ORACLE gateway requires an additional table to be defined on the database, access to tables in the data dictionary, and two servers to be started on the target node.

FUNCTIONAL ARCHITECTURE

Interbase supports two global data manipulation languages: SQL and the Groton Data Manipulation Language. These two languages may be used in precompiled third-generation language source programs or in the Query Language Interpreter (QLI). The QLI translates either language into BLR, a lower level language based on the DSRI, which may be generated and read by programs.

Although InterBase has no global conceptual schema, it provides a navigation path to distributed local database management systems through a database handler which defines node names for multiple databases. A program called Y-valve determines if requested data are on a local node or a remote node. In the case of remote data, the request is routed through the network and the remote data dictionary is accessed.

Interbase or RDB data requests are left in the BLR lower level language. With ORACLE, BLR is restricted to SQL through the gateway and passed to the ORACLE access method. Data format incompatibility between heterogeneous systems is transparently handled by a XDR layer.

SALIENT FEATURES

Interbase is designed to handle on-line complex applications (OLCP). It supports: multi-user, concurrent update; automatic transactions; serializable and high concurrency levels; two-phase commit; multiple database transactions; and multiple parallel transactions per process. These features provide excellent concurrency control capability.

2.15. Massachusetts Institute of Technology: Composite Information Systems

The concept of Composite Information Systems (CIS) has evolved at the Sloan School of Management of Massachusetts Institute of Technology. The first paper on this technique was published in 1978, and significant enhancements to the original concept have been made during the intervening years. While all other approaches discussed in this report tend to focus exclusively on technical issues, the CIS approach emphasizes that it is necessary to address a number of inter-related strategic, organizational and technical issues, in order to implement systems that can effectively span applications, functional areas, organizational boundaries, and geographic separations.

Strategic issues include motivating cooperation between multiple organizations, each with its own goals, priorities, and security needs. One critical factor for such cooperation is participant consensus on the issue of access to each others' technical and non-technical information. Issues such as domains of shared information, benefits to each participating group, and the role and the responsibility of each constituent must be resolved in advance, for the final system to succeed in a real world environment.

Organizational issues include the process of making controlled changes in complex organizational environments. The latest results from the realm of inter-organizational networks have been utilized to explain the multiple forces that modulate behavior of individuals, groups, and organizations. Also, the notion of focused standards has been conceived to serve as the foundation for neutral representation as well as for the delineation of more elaborated standards.

Technical issues relate to connectivity at both physical and logical levels. Physical connectivity refers to the process of actual communication among disparate systems, and covers aspects as bandwidths, security, availability, reliability, and internetwork protocol conversions. Logical connectivity refers to the process of accessing disparate systems in concert for composite answers. One of the major challenges in logical connectivity is to reconcile the different assumptions and perspectives in the systems being integrated.

FUNCTIONAL ARCHITECTURE

The operational prototype of CIS is called CIS/TK, which is abbreviated from CIS Tool Kit. This prototype is based on the three-schema approach, with one integrated global schema, multiple local schemata, and multiple application schemata. The CIS/TK architecture includes multiple Application Query Processors (AQPs), one Global Query Processor (GQP), and multiple Local Query Processors (LQPs). There is one AQP for each application, and this AQP converts an application model query, defined by an application developer, into a sequence of global schema queries, passes them to the GQP, and receives the results. Similarly, a Local Query Processor (LQP) interfaces between the global data processor and an individual query command processor (e.g., a DBMS).

The primary query processor is the GQP. It converts a global schema query into abstract local queries, transmits them to the appropriate LQPs, and joins the results before passing them back to the AQP. The GQP must know how to get the data, how to map global schema attribute names to the column names, and how to join results from different tables. This is done with the aid of concept agents. Conceptually, each concept agent is composed of three functional components, as follows:

(a) A concept definition which defines the rules and data, given the global schema and an application model;

(b) A concept processor which pursues the goal of the concept agent, given its rules and data; and

(c) A message handler which sends and receives messages on behalf of the concept agent.

The notion of concept agents is used to perform concept inferencing, in situations involving missing and conflicting information.

CIS/TK utilizes an object-oriented approach. Physical communication details and database idiosyncracies are encapsulated within the object. The object-oriented language utilized by CIS/TK is based on an enhanced version of Knowledge-Object Representation Language. CIS/TK is currently providing integrated access to several databases at MIT, and others outside MIT. External databases include Reuter's Textline and Dataline and I.P. Sharp's Disclosure.

SALIENT FEATURES

CIS is an integrated methodology that attempts to systematically deal with strategic problems, organizational problems, and technical problems that are inherent in any endeavor aimed at integrating islands of disparate information systems. CIS/TK is an operational prototype that utilizes innovative concepts to perform logical connectivity in an effective manner.

2.16. Metaphor Computer Systems, Inc.: DIS

DIS is a product of Metaphor Computer Systems, Inc. of Mountain View, California. Through the combination of a graphical interface and relational database technology, DIS lets users access data from multiple databases, perform analysis, and construct their own applications without programming. Designed for business professionals, DIS provides tools for strategic use of data from multiple sources; database query tools and report writers for direct and easy access to data; spreadsheets to analyze data; plot tools to view data graphically; and text processing tools to add commentary to the analysis.

Currently, DIS only provides for data access to SQL relational databases and to some flat files after they are converted to relational format used by DIS. There is no plan to provide direct access to non-SQL environments. However, to the extent local SQL database management systems can translate other databases into their own relational data model (e.g., Ingres), DIS can access them.

There is no logical schema which encompasses all database management systems in the network. However, a local schema of one or several databases is graphically depicted for the user. This provides the ability to "click" on each database and instantly see a list of its tables. DIS can be used to query multiple physical database management systems concurrently through multiple windows, and to combine and manipulate such data using the spreadsheet tool.

FUNCTIONAL ARCHITECTURE

DIS is a network of workstations (soon to include PS/2's and PC's) and servers that interact with relational databases. Databases can reside on Metaphor database servers as well as on non-Metaphor host machines. Metaphor files and communication servers can provide electronic mail and print services, plus access to a wide range of multi-vendor environments.

The main data access tool is the Query tool. When it is selected, the appropriate database is connected and each of its tables and its columns are displayed in a window. As the user chooses columns with the mouse, the Query tool draws a picture showing how the database tables will be joined to retrieve the data. Then it translates the graphic query into SQL and sends it to the database.

If data are not directly accessible to DIS via a gateway, they can be downloaded to a Metaphor database server. Host resident "formatters" are available for converting flat data files into the relational format used by DIS.

SALIENT FEATURES

This product claims to provide a true user friendly data access. Querying can be done without ever touching the keyboard. All tables and columns within a physical database are graphically provided to the user. The spreadsheet, the plotting facility, and the word processing tools are very practical features. The list of government and commercial customers attests to the ability of this product to provide management with easy access to information. IBM has invested in Metaphor Computer Systems, Inc. and is currently working with it to provide a similar interface for IBM mainframes.

2.17. On-Line Software International, Inc.: RAMIS Information System

RAMIS Information System from On-Line Software International, Inc., headquartered in Fort Lee, New Jersey, is used primarily for developing and running end-user applications. The system consists of many diverse components such as: a centralized database, a data dictionary, reporting and graphics facilities, links to external database management systems, data manipulation tools, a procedural language for application development, menu architecture development facilities, micro-to-mainframe communications, and data transfer facilities.

The RAMIS database may be comprised of both RAMIS and non-RAMIS files or databases. Most non-RAMIS files and databases may be accessed in a read-only mode. However, DB2 and SQL/DS relational databases may be queried and updated from within the RAMIS environment.

FUNCTIONAL ARCHITECTURE

RELATE is a facility to combine data from multiple files by operating on them as if they were flat relational tables. RELATE provides relational capabilities regardless of the organization of the source of the data. The user must specify the database or the file, the data to be retrieved, the record selection criteria and how the data are to be associated.

The RAMIS Database Manager reads and updates RAMIS, DB2, and SQL/DS files. It allows multiple physical files to be logically interconnected to from a single data view. A user is presented with one relational view of the data. In the same way, multiple physical databases may be treated as one logical database. Specialized interfaces are required for accessing other non-RAMIS files and databases.

RAMIS has a totally integrated, menu-based architecture front end to all components of the system. Application developers can utilize the menu interface, use a command language, and construct their own customized menu interfaces.

SALIENT FEATURES

The RAMIS Information System operates on various IBM platforms under diverse operating systems. It provides a flexible environment for managing, retrieving, manipulating and reporting data stored both in RAMIS format and in external databases and files. Direct interfaces within RAMIS allow users to read non-RAMIS files and databases using RAMIS language facilities. It is not necessary to convert external files into a RAMIS file to access the data.

2.18. Oracle Corporation: ORACLE

ORACLE was developed by Oracle Corporation, Belmont, California, to serve as a general purpose tool for providing a uniform, integrated interface for retrieving data from several existing database management systems. ORACLE connects with SQL/DS, DB2 and DBase III relational databases. There is also an ORACLE utility for 123.

ORACLE is constantly evolving to meet System R* specifications which IBM determined in the 1970s to define the "ideal" database management system. The influence of IBM is further seen in the global data manipulation language which is a superset of SQL, the IBM developed Standard Query Language. ORACLE is a widely-installed DBMS in the VMS environment.

Both distributed query and single-node transactions are supported by ORACLE. For updates at multiple nodes, there is no assurance that transactions will either be committed or rolled back together. Data loss can only be prevented by application programming techniques. Global updates and the two-phase commit procedure will be supported in the next product release.

FUNCTIONAL ARCHITECTURE

The Open Systems Architecture of ORACLE for distributed processing and distributed database management is SQL*Star. It is comprised of three ORACLE products: SQL*Net, Distributed ORACLE relational

DBMS, and SQL*Connect.

SqL*Net provides distributed processing capability, and the application software can run on a different machine from the ORACLE kernel. Users can interface SQL*Net to arbitrary network environments. SQL*Net is communications protocol independent. It consists of a generic layer which is common to all protocols and environments, and a custom layer which is specific to each protocol and operating system.

The Distributed ORACLE relational DBMS supports location transparency, single site transactions, multi-site updates in multiple transactions and site autonomy. To meet System R* specifications, ORACLE is planning multi-site updates in a single transaction, replication and table partitioning for future releases. Distributed data dictionaries contain information about local data and how to get access to global data.

SQL*Connect allows ORACLE to communicate with non-ORACLE database management systems. Plans for broadening access to other database management systems, as well as to non-database file systems (VSAM, Record Management Service) are continuing. Query optimization is achieved by using an array interface to transfer data in batches. This insures that the network is used efficiently.

SALIENT FEATURES

ORACLE products run on an impressive range of mainframes, minis, workstations and personal computers (over 90 hardware platforms), and support a large number of operating systems and a wide variety of network services. The designers intend to encrypt usernames, passwords, and SQL data. Approaches for meeting B1 and C2 National Computer Security Center (NCSC) security requirements are under development.

2.19. Relational Technology, Inc.: INGRES

INGRES is marketed by Relational Technology, Inc. of Alameda, California. Products of this company are used extensively around the world. Approaches originally designed for homogeneous environments have been extended to cater to heterogeneous situations.

FUNCTIONAL ARCHITECTURE

The INGRES approach consists of: INGRES/NET Network Protocol Support, INGRES/STAR Distributed Data Manager, and INGRES Database Gateways. INGRES/NET is a transparent software layer that provides INGRES protocol support for DECnet, TCP/IP, SNA and ASYNC communications. This software allows access to databases across a network and performs automatic data conversion. With the INGRES/STAR Distributed Data Manager, simultaneous access to distributed databases across a network is possible. INGRES/STAR provides complete location transparency and local autonomy. INGRES Database Gateways enable other types of databases and file systems to be included in a distributed database. The appropriate gateway module must reside on the local host. Relational, hierarchical, and network local data models are supported.

A two-phase commit protocol is used for distributed transaction processing. If a site involved in a transaction is inaccessible, other sites are automatically rolled back. When a system crashes, INGRES backs out all pending transactions and returns the database to a consistent state without operator intervention.

The global data manipulation language of INGRES is ANSI SQL. After receiving a SQL query from an application, INGRES/STAR breaks it into subqueries and directs them to the appropriate local data managers and gateways. Next, each local data manager executes its respective subquery and the data are transferred to INGRES. Data can be transparently joined and combined with INGRES SQL end-user interfaces or application development tools.

The INGRES/STAR Distributed Data Manager combines distributed query optimization with parallel processing. The distributed optimizer selects the best route for satisfying a multicomputer request. This approach reduces network traffic and improves performance. With this open, multi-server architecture, system administrators can customize their environment by assigning various servers with dissimilar priorities for different tasks, and by connecting transaction applications to a high priority server and batch report-oriented applications to a lower-priority server.

INGRES allows dynamic control of parameters to adjust the concurrency of the system to the needs of the application set. Transaction and decision support applications may be executed at the same time. Also, the multi-server architecture allows system-wide monitoring and control of database sessions for real-time system management. A virtually infinite number of simultaneous users is possible.

The INGRES data dictionary stores all performance related information-table structures and sizes, available indexes, and data distribution statistics. The more information is provided in the dictionary, the more intelligent is the Query Optimizer. The Query Optimizer produces query execution plans which can be examined to determine the effects of adding indexes, selecting different storage structures, and adding data distribution statistics.

SALIENT FEATURES

INGRES incorporates many of the features required for distributed transaction applications including two-phase commit, automatic recovery, and table fragmentation. Distributed query optimization strategies and full parallel query execution capabilities enhance its distributed on-line transaction processing capability.

2.20. Software AG: ADABAS

The ADABAS DBMS, developed by Software AG of West Germany, is a part of an open Integrated Software Architecture (ISA). Used in conjunction with NET-WORK (for distributed processing) and ADA-

NET (for distributed databases), ADABAS can support heterogeneous networks involving IBM, SIEMENS, DEC-VAX and WANG computers. The ADABAS DBMS supports relational, hierarchical, and network local data models.

ADABAS is designed for production, development, and information center environments. It claims to handle both high transaction volumes and complex ad hoc queries operating against large volumes of data. Distributed transactions can be in the form of a series of updates to different databases which might involve more than one network node.

FUNCTIONAL ARCHITECTURE

ADABAS uses either SQL embedded in COBOL, FORTRAN or PLI or NATURAL as its global data manipulation language. NATURAL is a fourth generation language based on a Communication Interface and standard protocols. NATURAL applications may be ported without modification across a wide range of system environments and hardware architectures.

The ADANET product of Software AG allows a user level logical ADABAS call to be translated into one or more physical calls in a transparent manner. A dynamically maintained ADANET translation table defines logical databases and their corresponding physical database representations. Three logical file structures are supported by ADANET: replicated, distributed and isolated files.

PREDICT is the centralized integrated active data dictionary. The structure of the metadata in PREDICT describes the logical and physical interconnections of all data and programs to be managed. PREDICT offers a user-friendly, menu-driven user interface for the management of data and application information.

SALIENT FEATURES

The Open Integrated Software Architecture (ISA) provides a high level of functionality in areas such as end user access, application solutions, application development and database management. It also provides software independence across multiple hardware and operating platforms without requiring any change in local databases. Global update processing is supported for relational, network and hierarchical local data models. The global data manipulation language of ADABAS reduces dependencies on third generation languages and is highly portable.

2.21. Sybase, Inc.: SYBASE

SYBASE is marketed by Sybase, Inc. of Emeryville, California. It can handle data in either hierarchical form or relational form. Data in a network DBMS are accessed via a relational interface.

FUNCTIONAL ARCHITECTURE

The Sybase Distributed System is comprised of three major components: the Sybase Data Server, the Sybase OpenServer, and the Sybase StarServer. The Sybase DataServer allows remote queries and distributed transactions across multiple DataServers while the Sybase OpenServer extends this capability to other applications or DBMS Servers, including non-SQL systems. The Star Server adds the functionality of site transparency, including cross-site joins and views, and the Distributed Data Dictionary. The SYBASE OpenServer is in the final stage of beta testing; the DataServer and the StarServer are commercially available.

The SYBASE DataServer allows multi-site update transactions, while ensuring the logical consistency of the data via a two-phase commit protocol. A 'commit service' running at a designated site logs the progress of a transaction. The commit service logs notes about when all sites will be prepared. If the transaction fails before that point, it will be backed out; if it fails after that point, it will be rolled forward.

With its multithreaded client/server architecture, SYBASE assumes many of the functions of an operating system. SYBASE avoids the overhead of process swapping by scheduling and activating users. Multiple users share a single copy of the SYBASE image, data and procedure cache, and transaction log. Because each additional user requires only 30 KBytes of memory, hundreds of users can simultaneously access and update data on any node in a distributed network, according to Sybase.

A Virtual Server Architecture (VSA) for Symmetric Multiprocessor Systems is scheduled for beta testing. The VSA treats processors as resources to be allocated on an as-needed basis. An application or a user is not restricted to a specific processor; instead applications may run on various processors at different points in time depending on the number of blocks for I/O requests or other reasons. Four key features of VSA are: synchronization control, automatic load balancing, optimal resource utilization, and dynamic configuration.

SYBASE uses a Distributed Data Dictionary, consisting of synonyms which are used to locate objects, and Remote Object Catalogs which contain information about the characteristics of those objects. With this approach, users and applications need not be concerned with changes in the physical location of an object.

SALIENT FEATURES

Sybase is a leader in secure database management systems. The SYBASE Secure SQL Server meets Division B security requirements defined by the National Computer Security Center. In compliance with Division B specifications, SYBASE offers the ability to store data of multiple security classifications in a single database.

Because networking was built as a key part of SYBASE, rather than as an add-on, network usage is optimized. The network is used more efficiently by sending a single 'execute procedure' command, instead of an individual SQL statement.

SYBASE allows server-to-server communications. Multiple users may share the communication link between servers, thus eliminating the overhead of opening and closing a linkage for each user.

2.22. Tandem Computers Incorporated: NonStop SQL

Tandem Computers Incorporated, headquartered in Cupertino, California, offers software and network compatible systems ranging from mainframes to low-cost network nodes. Because all Tandem processors use the same operating system and distributed database, as many as 255 systems with a maximum of 4,080 processors can be linked to form a geographically dispersed network. An individual system can support hundreds of terminals and a database with billions of bytes of data.

FUNCTIONAL ARCHITECTURE

GUARDIAN 90, a message based operating system, supports the parallel architecture of Tandem NonStop systems. Key features of GUARDIAN 90 are: fault tolerance, distributed data processing capabilities, high transaction throughput, linear expandability, security, data integrity, and the ability to link systems, resources, and devices.

GUARDIAN 90 supports fault tolerant process pairs. In the case where a system software error causes a process to fail, its backup process takes over in another processor. In addition, GUARDIAN 90 manages communication between processors. Database availability is always ensured. Mirrored volumes of data are stored on a pair of physically independent disk drives. When a failed drive is operational again, the system automatically brings it up to date. The systems resume normal mirrored operation after the restored disk is fully updated. Via its message system, GUARDIAN 90 coordinates all distributed activity in a fashion that makes all data accesses transparent to the user. An application can read, write, and update data regardless of the physical location of the data. A built-in password security system is used to control file, device and process availability. Tandem offers integrated security and encryption products to protect information from unauthorized access.

SALIENT FEATURES

The multiple parallel processing capability, the fault tolerance capability, and the expandability feature make NonStop SQL a good product for on-line transaction processing (OLTP).

Tandem Corp. has signed integration agreements with Oracle Corp, Sybase, Inc., Relational Technology, Inc. (Ingres), and Information Builders, Inc. (FOCUS) to port their front-end tools to Tandem Systems. Details of the approaches used by these companies are presented in their respective subsections.

2.23. TRW: TDIE

TDIE is a data integration engine developed by TRW, Redondo Beach, California. Unlike conventional efforts which use a gateway with query language mapping, TDIE has no global data manipulation language. A local DBMS is used and a connection to a remote DBMS is triggered when a user issues a demand requiring data integration.

TDIE is not oriented towards any particular environment. No change is required in existing applications. Additional local database management systems may be added to the system configuration and more than one global schema is permitted.

An integrated scheme performs uniform query and updates for hierarchical, relational and network data models. Automatic updating is performed by polling the potential source files at a rate suitable for a given data category. Types of data that are frequently accessed and must be current are polled at short intervals. This approach eliminates unnecessary activity and avoids unnecessary copying of bulk data by transferring only the changes in data.

FUNCTIONAL ARCHITECTURE

The TDIE is comprised of the Data Transform Manager and the Host Interface Manager. These may reside on one processor or they may be partitioned across multiple processors.

The Data Transform Manager contains the centralized dictionary which defines the global conceptual schema. This knowledge-based dictionary supplies scripts, translation software, and data locations for full data integration. Scripts are instructions for data transfer, which are sent to the Host Interface Manager, while the translation software contains the translation of one query language to another.

To add another local DBMS, translators and scripts must be added to the dictionary. TRW provides a front end tool called the Integration Advisor for interactive construction of a global integrating schema.

The Host Interface Manager is the bridge to the host applications. It provides for automatic updating and data retrieval. Time-critical applications in transaction processing or real-time control are not supported. Also, because TDIE is not a distributed heterogeneous DBMS, concurrency control is an important issue.

SALIENT FEATURES

TDIE supports global updates and enables the integration of hierarchical, relational, and network data models. It is layered on top of the existing systems and does not require any change to them. New systems can be added as requirements change. Also, the TDIE can be implemented in increments as desired.

2.24. Xerox Advanced Information Technology: Multibase

Multibase was developed by Research and Systems Division of Computer Corporation of America. This division was acquired by

Xerox and is now called Xerox Advanced Information Technology (Xerox AIT). Although Multibase is a prototype and not a product, Xerox AIT utilizes this technology for providing customized distributed heterogeneous database management systems for its clients.

Multibase provides a uniform, integrated interface for retrieving data from several existing database management systems. It allows a user to reference data in heterogeneous databases, through a common query language, using a single conceptual schema. Multibase is designed to serve as a general tool, without specific orientation towards any particular application area. It allows existing applications to operate without change and also permits new local systems to be included in an existing multibase system configuration.

The integrated access available through Multibase does not provide either the capability to update data in local databases or the ability to synchronize read operations across several sites. In order to process user queries, the system must request and control specific services offered by the local systems (e.g., locking local items).

FUNCTIONAL ARCHITECTURE

Multibase uses the language DAPLEX as its global data manipulation language. DAPLEX provides constructs that allow users to model real world situations in an efficient manner. A three level schema of definitions is employed as discussed in Section 1.5. The process of global information retrieval involves two main components. These are: Global Data Manager (GDM), and Local Database Interface (LDI). The GDM performs tasks of Command Processor, Decomposer, Merger and Distributed Execution Monitor. The LDI design of Multibase is based on the needs of a local DBMS, for example, a more sophisticated LDI is designed to support a file system as compared to the one that supports a DBMS.

SALIENT FEATURES

Multibase provides an integrated scheme for uniform query access to dissimilar systems including hierarchical, relational, and network local data models, and File Systems. Global query optimization is performed in Multibase.

SECTION THREE: DISCUSSION AND CONCLUSION

The first section of this report explained the basic principles of integrating disparate information systems. In the next section, a broad spectrum of commercially-marketed products and approaches were discussed. Based on the material covered so far, this section develops a set of recommendations for selecting the optimal product or products. In particular, the information presented in Section 2 and Appendix B is utilized to develop a taxonomy for classifying all the alternative approaches.

In Section 1.4, it was emphasized that changes to existing information systems should be as little as possible; in the ideal case, no change should be required. Subsequently, in Section 1.5, the importance of a single global conceptual schema was emphasized. Many products and approaches tabulated in Appendix B indicate that multiple global schemata can be created to suit the needs of various users. From a technical viewpoint, these are external schemata, not global schemata. These are some of the considerations that were examined while developing the broad categories described in the succeeding subsections.

3.1. Unintrusive Integration

This is the most elegant form of integrating information systems. The word "unintrusive" connotes that the integration methodology does not demand that specific modules be resident in each participating information system (see Figure 1). There is a definite assumption in products belonging to this category that existing information systems cannot, and should not, be tampered with. On top of these systems, a single global conceptual schema is introduced.

Use of this approach in a multiorganization or a multidivision environment implies that the existing information systems of various divisions and organizations will not require any modification. These systems will be accessed by a central system that holds a global schema with names and locations of all data items in participating systems. A user will access the central system, which will retrieve information from appropriate sources. The use of a central system and a single global schema offers appropriate controls on data integrity, access rights, and the feasibility of developing a new information system over time.

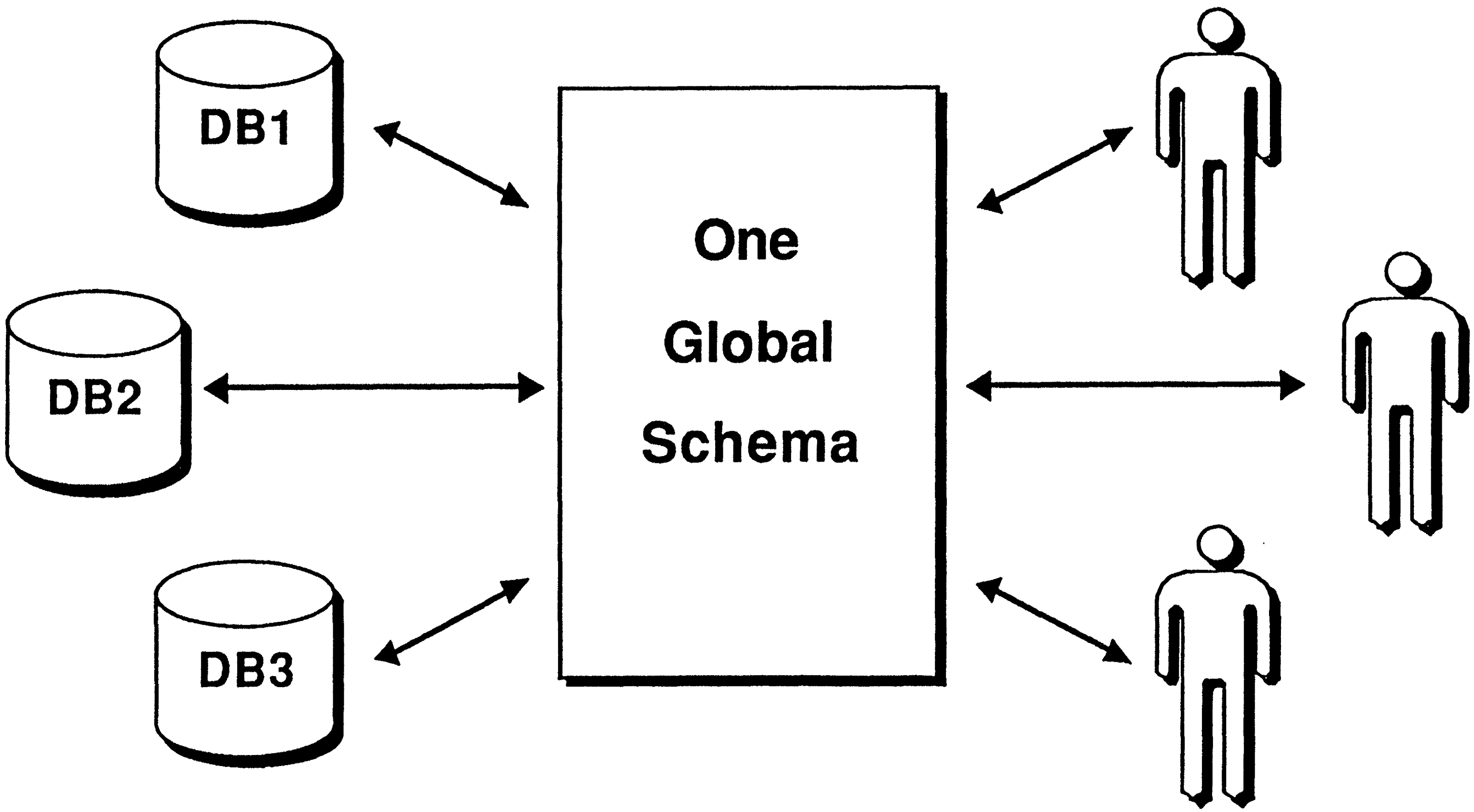
The main disadvantage of this approach is that is difficult to make global updates to multiple systems as there is no control over the capabilities of individual information systems. The application program must, therefore, be designed to safeguard the integrity of the updates.

Notable products in this category are: Multibase (developed by Computer Corporation of America and now marketed by Xerox); Mermaid (developed by Unisys and now marketed by Data Integration, Inc.); and CDM*Plus (developed under the IISS initiative of U.S.

Figure 1:

Unintrusive Integration

(No change to local information systems)



Air Force and now marketed by Control Data Corporation). The development of all these three products involved huge sums of money, and the acceptance of these products is still not commensurate with the investment. Defense organizations have been in the forefront in evaluating and field-testing these products.

3.2. Intrusive Integration

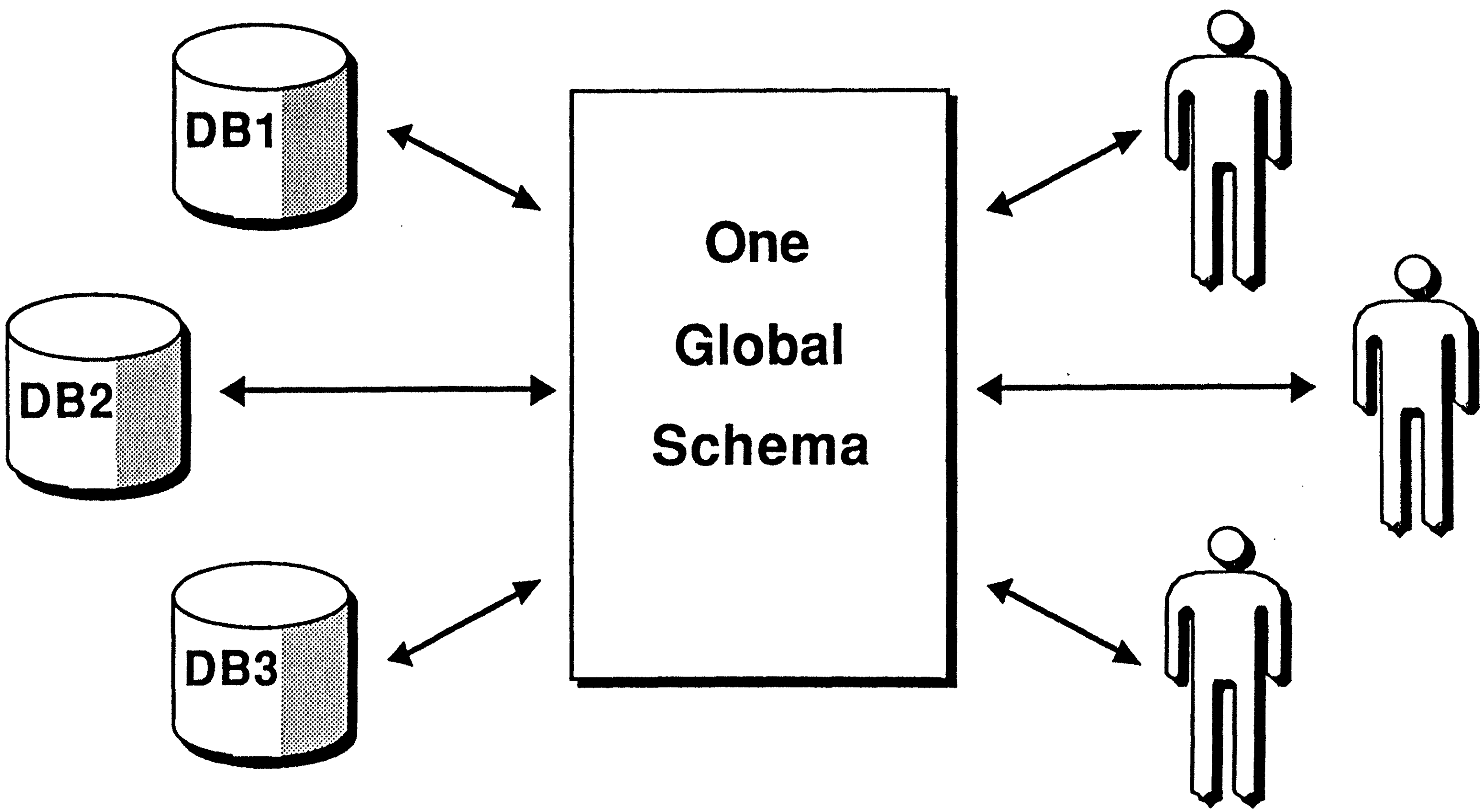
Approaches in this category require some modification of existing information systems or the location of specific modules within these systems. These modules are typically utilized to perform network services, to store information on location of specific files, or to initiate secondary processes in case some information resources are not operational or unusually slow in responding to requests (see Figure 2). If these modules are designed and implemented properly, and if the existing information systems offer the flexibility to accept these modules from technical and organizational considerations, then this approach can offer significant performance improvements over the approach outlined in Section 3.1. The main drawback, however, lies in the fact that existing information systems frequently use older technologies, employ outdated programming practices, offer little good documentation, and can only be modified after prolonged effort.

Notable products in this category include SUPRA (original technology developed in West Germany and now marketed by CINCOM), INGRES (from Relational Technology Inc.), and Sybase (from Sybase).

Figure 2:

Intrusive Integration

(Some changes to local information systems)



In the case of Sybase, it appears that the present family of products support SQL databases only, and support for non-SQL based databases and file systems is still being tested. There are some restrictions on the types of data models and data formats supported by these three products. For example, Supra requires data to be in first normalized form and keys to be in ascending or descending order. CINCOM is interested in developing a common protocol for integrating heterogeneous databases, that could be utilized by multiple companies. Because of these reasons, it is important to analyze potential applications carefully to determine which of these products, if any, is appropriate in the short-term and in the long-term.

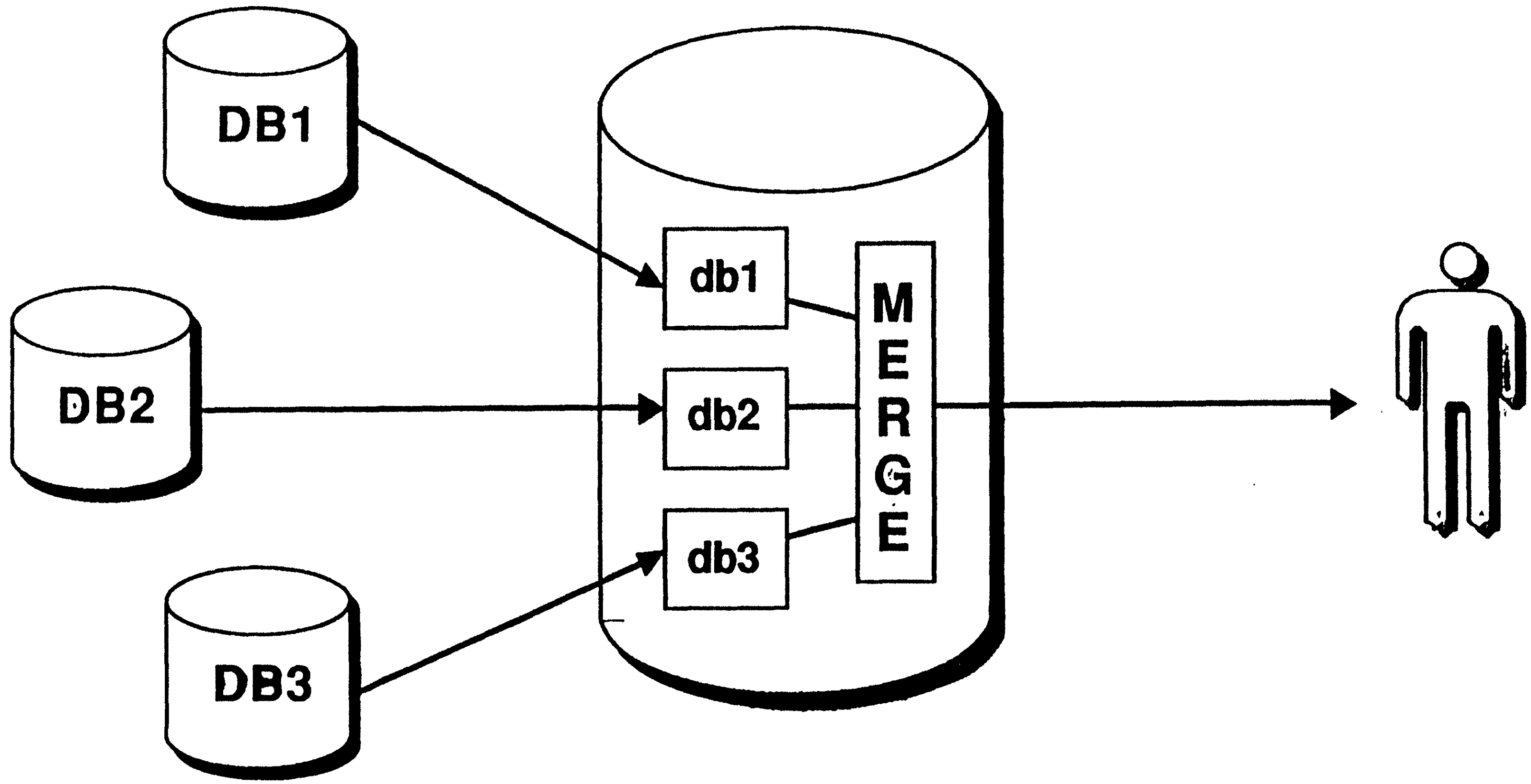
3.3. Interfacing with Controlled Redundancy

In this approach, no effort is made to create a global conceptual schema. Instead, based on the perceived needs of the user, strategies are designed to get the necessary pieces of information on a "just-in-time" basis (see Figure 3). As an example, consider a hypothetical application which requires that information on 100,000 items (which could be key customers, spare parts, or inventory levels) be available every morning. Assume that the existing information systems that are used to track the position of these items are updated with different periodicity: very few on a yearly basis, some on a monthly basis, a large percentage on a weekly basis, another large percentage on a daily basis, and the rest on a six-hourly basis. In such a situation, information can

Figure 3:

Interfacing with Controlled Redundancy

(Selective "Download" and Integrate)



be obtained in advance; further, to fulfil anticipated requests one can get information only for the items that have changed since the previous status report.

One product that exemplifies this approach is Data Integration Engine from TRW. In this patented approach, when a change in the status of a source information system occurs, this change is transmitted to the destination application with a priority that would allow it to reach the destination before the next scheduled time of retrieval. The software to accomplish this task and other related tasks is comprised of two types of material: a basic data engine, which is site independent, and additional material which must be customized to each site. The basic data engine is written in C and can operate in a UNIX or a VAX environment. The additional material can be generated with an expert system called Integration Advisor, which was originally written in LISP and has recently been converted to C. According to a technical representative of TRW, the ratio of the size of the basic data engine to the size of the additional material is 5:1. Incidentally, TRW has recently been awarded a multiyear contract by the U.S. Department of Defense.

3.4. Interfacing with Transaction Transfer

This is the simplest technique for managing transactions in a heterogeneous information system environment. Instead of creating a global schema or prestaging data to meet expected needs, this approach implies that action begins after a user enters a command. Unless the user is willing to tolerate long delays, this approach is good only in situations where only a small number of information systems need to be accessed in response to the user query (see Figure 4).

Among the notable products in this category, the one that deserves mention here is DIS from Metaphor Computer Systems, Inc. DIS possesses two unique characteristics. First, it is completely icon-driven and as such it is very easy to learn and use. Second, IBM has invested in Metaphor Computer Systems Inc. and they are working together to provide a similar interface for IBM mainframes. As such, the DIS interface has the potential of becoming a commonly used standard for dealing with multiple information systems.

3.5. Additional Evaluation Criteria

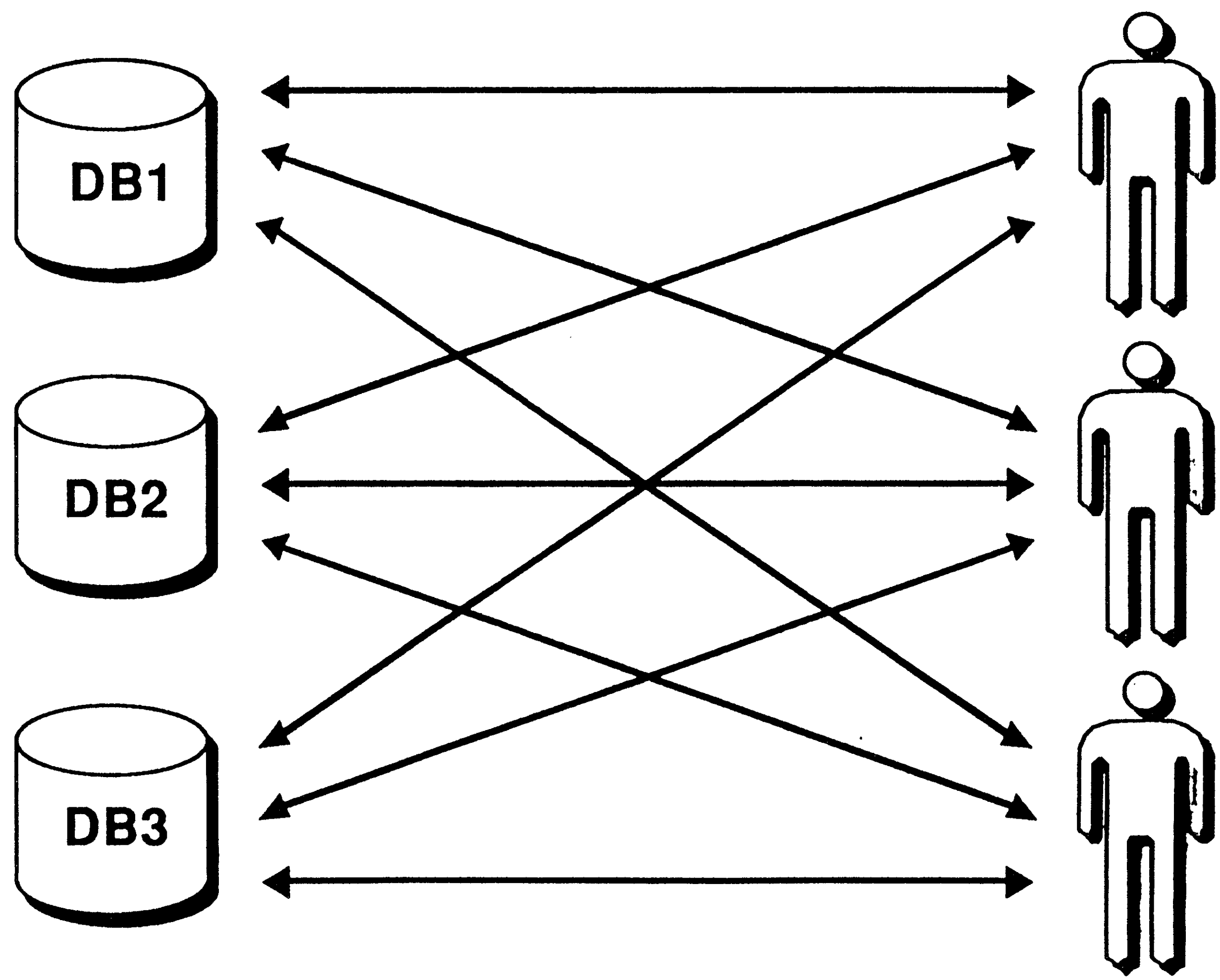
In addition to determining which of the four distinct approaches described in Sections 3.1-3.4 above may be most appropriate for a particular application, it is necessary to examine a number of other technical and non-technical issues before embarking on a specific implementation strategy. These issues are delineated in the following paragraphs.

3.5.1. Composition of Transactions

Transactions are of two types: queries and updates. In our

Figure 4:

Interfacing with Transaction Transfer



context, a query involves reading of data in multiple computer-based systems, manipulating the data retrieved to make them consistent, and then presenting the result to the user. The original data in source computer-based systems remain unaltered by a query. An update, however, results in modification of data in various information systems. It is much more difficult to deal with updates than with queries. If an application involves updates, then a significant number of the alternative approaches discussed in Section 2 should be deleted from consideration. The ones to be deleted can be readily identified from Appendix B, which contains a specific question on how update transactions are managed.

3.5.2. Security Considerations

When one makes it easier to automatically access relevant pieces of information in disparate systems, the same capabilities become available to any unauthorized user of the system. Any integrated set of systems is more vulnerable to security violations than a single information system.

None of the vendors was specifically asked to comment on this issue, although a few of them have provided information on their compliance level with security requirements defined by the National Computer Security Center. This issue may hold significant weightage in the case of large defense, banking, and insurance applications.

3.5.3. Corporate Expertise

Most large computer corporations have tended to concentrate on integrating systems of their respective makes only. For example, in Section 2, there was a discussion of IBM and DEC suggesting solutions pertaining to their respective families of products. Frequently, smaller companies come up with non-traditional concepts, and larger companies tend to endorse them after these concepts have attained some degree of maturity and acceptance by the market.

The fact that the companies who developed good prototypes such as Multibase and Mermaid later spun off their development groups must be viewed with caution. This is an evolving field, and user organizations must be careful about selecting only the set of products and services that are likely to be supported by more than one company.

3.5.4. Applications and Development Strategy

In the current business environment of increasing globalization and corporate reorganizations, applications and functions are apt to change in relatively short periods of time. The present technical report has been written without any specific application in view. As an organization determines which applications are important in the short-term and the long-term, it would become feasible to make more specific comments about the relevance of different products and approaches.

It is important to note that any endeavor towards integration has ramifications for future integration initiatives. The selected approach must offer the flexibility to be enhanced in a modular fashion. Further, public-domain approaches should be preferred over proprietary ones. Based on these considerations, a development strategy can be delineated.

3.6. Conclusion

The integration of existing information systems involves a careful analysis of many technical, organizational, and strategic issues, and the surmounting of many technical and nontechnical barriers. This technical report has focused on products and their capabilities related to logical access to relevant pieces of information. For discussion of other approaches and issues, the reader may wish to refer to Ref. [2] and the set of eight books on Knowledge Based Integrated Information Systems Engineering published by M.I.T.

The problem of heterogeneity can be overcome either by integration or by interfacing. To decide which option to select, it is worthwhile to look back to the stage when the concept of database management system originated. If one wanted to deal with a single clearly-defined application, it was better to write a customized program to do so; however, if there were many users with diverse needs, it became increasingly attractive to use the concept of a database management system, even though it involved a larger investment at the initial stage. The same considerations apply at the metalevel today. If access to multiple information systems is needed exclusively for a small set of clearly-defined applications, the interfacing alternative should suffice. However, if there is likely to be an array of broad applications, it appears preferable to take the integration alternative. Remember that "just as the evolution of integrated database management technology displaced conventional piecemeal programming approaches in the 1960s, similarly the new information integration approaches will lead to a significantly different computational environment by the year 2000 A.D.!" [1].

REFERENCES

[1] Gupta, Amar, Integration of Information Systems: Bridging Heterogeneous Databases, IEEE Press, New York, 1989.

[2] Morris, J.H., Satyanarayanan, M., Conner, M.H., Howard, J.H., Rosental, D.S.H., and Smith, F.D., "Andrew: A Distributed Personal Computer Environment," Commun. ACM, Vol. 29, 3, March 1986, pp. 184-201.

[3] Balkovich, E., Lerman, S., and Parmelo, R.P., "Computing in Higher Education: The Athena Experience," Commun. ACM, Vol. 28, 11, Nov. 1985, pp. 209-215.

[4] Adiba, Michel, and Portal, Dominique, "A Cooperation System for Heterogeneous Data Base Management Systems," Inform. Systems, Vol. 3, 1978, pp. 209-215.

[5] McLeod, Dennis, and Heimbinger, Dennis, "A Federated Architecture for Database Systems, " in AFIPS Conf. Proc., Vol. 49 (National Computer Conf., 1980), pp. 283-289.

[6] Lam, Chat Y. and Madnick, Stuart E., Composite Information Systems - A New Concept in Information System, Working Paper #35, Center for Information Systems Research, Sloan School of Management, Massachusetts Institute of Technology, May 1978.

[7] Gupta, Amar, and Madnick, Stuart, Knowledge-Based Integrated Information Systems Engineering: Highlights and Bibliography, Massachusetts Institute of Technology, 1987. (NTIS/DTIC Accession # A 195850).

[8] Litwin, Witold et al, "SIRIUS Systems for Distributed Data Management," in Distributed Databases, edited by: H. Schneider, North-Holland Publishing Co., 1982, pp. 311-366.

[9] Heimbinger, Dennis, and McLeod, Dennis, "A Federated Architecture for Information Management," ACM Transactions on Office Information Systems, Vol.3, No. 3, July 1985, pp. 253-278.

[10] Thompson, Glenn Ray, Multi-database Concurrency Control, Doctoral Dissertation, Oklahoma State University, Stillwater, Oklahoma, 1987.

[11] Madnick, Stuart E., and Wang, Y. Richard, "Integrating Disparate Databases for Composite Answers," in Proc. of the Twenty-first Annual Hawaii International Conference on System Sciences, Vol.II, January 1988, pp. 583-592.

[12] Pu, Carlton, Superdatabases for Composition of Heterogeneous Databases, Technical Paper No. CUCS-243-86, Columbia

University, New York, 1986.

[13] Gibbons, Phillip B., "A Stub Generator Multilanguage RPC in Heterogeneous Environments, IEEE Transactions on Software Engineering, Vol. SE-13, No. 1, January 1987, pp. 77-87.

[14] The ANSI/XI SPARC DBMS Framework Report of the Study on Database Management Systems, AFIPS. 1977.

[15] Appleton, Daniel S., "The Technology of Data Integration," Datamation, November 1, 1985, pp. 106-116.

[16] Bhalla, S., Prasad, B.E., Gupta, A., and Madnick, S.E., "A Technical Comparison of Distributed Heterogeneous Database Management systems," in Integrating Distributed Homogeneous and Heterogeneous Databases: Prototypes, edited by Amar Gupta and Stuart Madnick, M.I.T., Cambidge, MA, 1987, pp. 161-218.

[17] Reddy, M.P., Prasad, B.E., and Reddy, P.G., "Query Processing In Heterogeneous Distributed Database Management Systems" in Reference [1] above, pp. 264-277.

[18] Thompson, Paul, "Reference Models for Information Engineering," in Reference [1] above, pp. 291-291.

BIBLIOGRAPHY

[1] Gupta, Amar, and Madnick, Stuart, eds., KBIISE--Vol. 2: Knowledge-Based Integrated Information Systems Development Methodologies Plan, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195851).

[2] Gupta, Amar, and Madnick, Stuart, eds., KBIISE--Vol. 3: Integrating Distributed Homogeneous and Heterogeneous Databases: Prototypes, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195853).

[3] Gupta, Amar, and Madnick, Stuart, eds., KBIISE--Vol. 4: Object-Oriented Approach to Integrating Database Semantics, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195854).

[4] Gupta, Amar, and Madnick, Stuart, eds., KBIISE--Vol. 5: Integrating Images, Applications and Communications Networks, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195854).

[5] Gupta, Amar, and Madnick, Stuart, eds., KBIISE--Vol. 6: Strategic, Organizational, and Standardization Aspects of Integrated Informations Systems, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195855).

[6] Gupta, Amar, and Madnick, Stuart, eds., KBIISE Vol. 7: Integrating Information Stystems in a Major Decentralized International Organization, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195856).

[7] Gupta, Amar, and Madnick, Stuart, eds., KBIISE--Vol. 8: Technical Opinions Regarding Knowledge-Based Integrated Information Systems Engineering, M.I.T., Cambridge, MA, 1987 (NTIS/DTIC Accession # A 195857).

[8] Wang, Richard, and Madnick, Stuart, eds., Connectivity Among Information Systems, CIS Project, M.I.T., 1988.

[9] Special Issue on Distributed Heterogeneous Systems, Proceedings of the IEEE, Volume 75, No. 5, May, 1987.

[10] Bachman, Charles W., and Ross, Ronald G., "Toward a More Complete Reference Model of Computer-Based Information Systems," Computers and Standards, Vol. 1, North-Holland Publishing Company, 1982, pp. 35-48.

[11] Cole, Robert, "A Method for Interconnecting Heterogeneous Computer Networks," Software-Practice and Experience, Vol. 17, No. 6, June 1987, pp. 387-397.

[12] Gavish, Bezalel, and Pirkul, Hason, "Computer and Database

Location in Distributed Computer Systems," IEEE Transactions on Computers, Vol. C-35, No. 7, July 1986, pp. 583-590.

Appendix A

COMPANIES CONTACTED AND APPLICABILITY

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
AT&T	Doug Justice Marketing Interstate 85 at Mount Hope Church Road McLeansville, NC 27301 Phone: (919) 279-3491 FAX: (919) 279-3980	Rapid	Relevant. Upon completion of contract negotiations, AT&T will provide more information.
Advanced Computing Environments	Beverly Phone: (415) 941-3399	N/A	Not relevant. ACE is an educational institution with no product or R&D.
Amperif Corporation	Nardis Martinez RDB Product Manager 9232 Eton Avenue Chatsworth, CA 91311 Phone: (818) 998-7666	RDM/2	Not relevant. RDM/2 is a relational DBMS. Integration of other distributed databases is not currently within its capabilities.
BBN Systems and Technologies Corporation	Steve Vinter Cronus Product Manager 10 Moulton St. Cambridge, MA 02138 Phone: (617) 873-2876 FAX: (617) 873-3776	Cronus Distributed Computing Environment	Relevant. Details included in document.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Booz Allen & Hamilton, Inc.	David Subar Senior Consultant Artificial Intelligence Practice 4330 East West Highway Bethesda, MD 20814 Phone: (301) 951-4580	IDIMS	Relevant. Details includ-ed in document.
CRI, INC.	No Contact Phone: (408) 980-7499	Relate/DB	Unknown. Answering Ser-vice takes calls but none returned.
Cambridge Technology Group	Steven Schenefeld 219 Vassar Street Cambridge, MA 02139 Phone: (617) 876-2338 FAX: (617) 499-1777	Surround	Relevant. Details includ-ed in document.
Cincom	Ken Pagnotta Regional Technical Manager 35 Braintree Hill Park Suite 100 Braintree, MA 02184 Phone: (617) 849-1020 FAX: (617) 849-1030	Supra	Relevant. Details includ-ed in document.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Cognos Corporation	Judd Lowe Senior Sales Representative 2 Corporate Place I-95 Peabody, MA 01960 Phone: (508) 535-7350 FAX: (508) 535-8241	PowerHouse	Relevant. Details included in document.
Computer Associates International Inc.	Jack Hadder District Manager, Defense 8227 Old Court House Road Vienna, VA 27181 Phone: (703) 356-7700 x635 FAX: (703) 847-0290	Datacom/DB	Not relevant. Datacom/DB does not support heterogeneous database environments.
Computer Corporation of America	Pete Goldman Senior Technical Representative 1800 Diagonal Road Alexandria, VA 22314 Phone: (703) 836-5200 FAX: (703) 836-6590	Multibase	The Research and Systems Division which developed Multibase has been sold to Xerox. See appropriate entry below.
Consultants for Management Decisions, Inc.	Todd Collins One Broadway Cambridge, MA 02142 Phone: (617) 497-4434 FAX: (617) 497-2305	N/A	Not relevant. Company builds custom software systems to serve complex information needs of large organizations.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Control Data Corporation	William Bradley CDM[t] plus Business Development Integration Technology Services 2970 Presidential Dr., Suite 200 Fairborn, Ohio 45324 Phone: (513) 427-6323 FAX: (513) 427-6301	CDM[t] plus	Relevant. Details included in document.
Control Data Corporation	Roger West Seminary Rd. Suite 1000 Alexandria, VA 22311 Phone: (703) 998-3575 FAX: (703) 998-3571	Ascent (Intelligent Gateways)	Relevant. Details included in document.
Cullinet	Alan Miller 400 Blue Hill Drive Westwood, MA 02090 Phone: (800) 551-4555	IDMSR & Enterprise Database	Not relevant. IDMSR runs only on IBM 370 Architecture and Enterprise Database runs only on the VAX.
DB/Access, Inc.	Mike Bergeron Director OEM Sales 2900 Bordon Ave. Suite 101 Santa Clara, CA 95051 Phone: (408) 735-7545 FAX: (408) 735-0328	Access Star	Not relevant. Access Star does not support data integration.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Data Access Corporation	Katherine L. Cronin Manager Government Accounts 1400 S.W. 119th Ave. Miami, FL 33186 Phone (305) 238-0012	Dataflex	Not relevant. The product is a relational DBMS, but does not provide the ability to integrate distributed databases.
Data Integration, Inc.	Marjorie Templeton President 3233 Federal Avenue Los Angeles, CA 90066 Phone: (213) 390-4507	Mermaid	Relevant. Details included in document.
Digital Equipment Corporation	Glen Light Government Software Services 721 Emerson Road St. Louis, MO 63141 Phone: (314) 991-6552 FAX: (301) 731-3018	Rdb/VMS, DBMS	Relevant. Details included in document.
Electronic Data Systems	Craig Rogers 13600 EDS Drive Herndon, VA 22071 Phone: (703) 742-1572 Phone: (703) 742-1555 FAX: (703) 742-1325	Custom-designed applications	Initial review suggested further investigation. The contact was requested to fill in a Standard Technical Specification Sheet on 8/29/89.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
General Motors Research Laboratories	Chin-Wan Chung GM Research Laboratories Warren, MI 48090 Phone: (313) 986-1386	Dataplex	Unknown. GM is working with EDS and several DBMS vendors on a commercial product. This product is still in the prototype stage.
Gupta Technologies	John Ambler 12200 Sunrise Valley Drive Reston, VA 22091 Phone: (703) 264-9002 FAX: (703) 264-0568	The SQL System	Relevant. Details included in document.
Hewlett Packard	John Blue 1550 Wilson Blvd. Arlington, VA 22209 Phone: (703) 284-2635	Custom-designed applications	Initial review suggested further investigation. The contact was requested to fill in a Standard Technical Specification Sheet on 8/17/89.
Honeywell Federal Systems, Inc.	Hal Hannickel 7900 West Park Drive Mail Stop 603 McLean, VA 22102 Phone: (703) 827-3776 FAX: (703) 827-3729	DDTS	Relevant. Details included in document.

7

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
IBM	Barbara Lallas IBM / 5C1 1 Copley Place Boston, MA 02116 Phone: (617) 638-2770 FAX: (617) 638-1735	SAA, DB2, and SQL/DS.	Relevant. Details included in document.
Information Builders, Inc.	Michael Howard FOCNET Product Manager 1250 Broadway New York, NY 10001 Phone: (212) 736-4433 FAX: (212) 967-6406	FOCNET	Relevant. Details included in document.
Information Dimensions, Inc.	Debbie Fitzgibbons 50 Stanford St. Suite 800 Boston, MA 02114 Phone: (617) 227-7155	DM	Not relevant. The product is a relational DBMS, which integrates homogeneous, relational database management systems.
Informix Software, Inc.	Donna Brewer Regional Sales Representative 4100 Bohannon Drive Menlo, Park, CA 94025 Phone: (415) 322-5166	Informix SQL	Not relevant. The product is a relational DBMS. Distributed capabilities will not be available until September 1990.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Interbase Software Corporation	Brian Smith Senior Sales Representative 209 Burlington Rd. Bedford, MA 01730 Phone: (617) 275-3222	Interbase	Relevant. Details included in document.
Intercim	Jim Padden Product Manager 4030-2 West Braker Lane Suite 200 Austin, Texas 78759 Phone: (512) 345-8872 FAX: (512) 345-0305	FDL/FDM	Not relevant. Product integrates manufacturing applications to one relational DBMS to provide factory decision control systems.
Intersystems Corporation	Steve Reagon Sales Representative 1 Memorial Drive Cambridge, MA 02139 Phone: (617) 621-0600	M/SQL	Not relevant. The product is a relational DBMS but integrates homogeneous relational database management systems only.
Massachusetts Institute of Technology	Richard Wang Assistant Professor Sloan School of Management Phone: (617) 253-0442	CIS	Relevant. Details included in document.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Metaphor Computer Systems	Ron Goldstein 2010 Corporate Ridge Suite 510 McClean, VA 22102 Phone: (703) 749-3620 FAX: (703) 749-3684	DIS	Relevant. Details included in document.
Must Software International	David Chester 1030 Massachusetts Ave. Cambridge, MA 02138 Phone: (617) 868-2950 FAX: (617) 876-7918	Nomad	Initial review suggested further investigation. The contact was requested to fill in a Standard Technical Specification Sheet on 8/14/89.
On-Line Software International	Peter MacPherson 75 Second Avenue Needham, MA 02194 Phone: (617) 449-8860	Ram's Information System	Relevant. Details included in document.
Oracle Corporation	Christie Ellsworth Technical Account Manager 3 Bethesda Metro Center Suite 1400 Bethesda, MD 20814 Phone: (301) 907-2739	ORACLE	Relevant. Details included in document.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Plus 5 Computer Services	Carter Crandal 1968 Innerbelt Business Center Drive St. Louis, MO 63114 Phone: (314) 426-3900	Unix Mumps	Not relevant. The product is not a DBMS, does not integrate or translate data, and provides no global schema.
RIM Technology, Inc.	Bill Holter Marketing Manager 1775 12th Avenue NW Suite 115 Issaquah, WA 98027 Phone: (206) 392-4776	RTRIM	Not relevant. The product is a DBMS limited to integration of homogeneous, relational databases.
Relational Technology, Inc.	Lori Dreyfus 6 New England Executive Park Burlington, MA 01803 Phone: (617) 727-5060 FAX: (617) 272-3135	Ingres Star	Relevant. Details included in document.
Sharebase (Formerly Britton Lee)	Paul Bass Director, Technical Marketing 14600 Winchester Blvd. Los Gatos, CA 95030 Phone: (408) 378-7575 FAX: (408) 866-0774	Sharebase	Not relevant. The product integrates homogeneous database management systems only.

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Smith, Abbott & Co.	Ed Geier 1725 K Street NW Suite 501 Washington D.C. 20006 Phone: (202) 223-4990	Autopro	Not relevant. The product performs distributed management only for its own databases.
Software AG of North America, Inc.	Patti Malanka Federal Systems, Inc. Suite 210 11490 Commerce Park Dr. Reston, VA 22091 Phone: (703) 620-0100	ADABAS	Relevant. Details included in document.
Sybase	Peter Loux Federal Programs Manager 4600 South Ulster St. Suite 700 Denver, CO 80237 Phone: (303) 721-3308	Sybase	Relevant. Details included in document.
TRW	Tony Materna Systems Engineering & Devel. Div. One Space Park Bldg. DH2/Room 2328 Redondo Beach, CA 90278 Phone (213) 812-6021	TDIE	Relevant. Details included in document.

12

13

Appendix B

DETAILED CHARACTERISTICS OF

PRODUCTS AND APPROACHES

APPENDIX A
COMPANIES CONTACTED & APPLICABILITY

COMPANY	CONTACT PHONE & ADDRESS	PRODUCT/ SYSTEM	APPLICABILITY
Tandem Computers, Inc.	Dave Magram Mail Stop 219-44 5300 Stevens Creek Blvd. San Jose, CA 95129 Phone: (408) 553-6757 FAX: (408) 553-6960	NonStop SQL	Relevant. Details included in document.
Unify Corporation	Frank Verardi 3870 Rosin Court Suite 100 Sacremento, CA 95834 Phone: (916) 920-9092	Accell/SQL	Not relevant. The product is a DBMS with a 4GL language, but does not integrate heterogeneous database management systems.
Unisys	Anita Holmgren Deputy Dir. of Advanced Projects Network and Information Securities Division 5151 Camino Ruiz Camarillo, CA 93010 Phone: (805) 987-9300	Mermaid	Rights to Mermaid have been licensed to Data Integration, Inc. See appropriate entry above.
Xerox Corporation	Gene Kakalec Adv. Information Technology Dept. 1800 Diagonal Rd. Suite 300 Alexandria, VA 22314 Phone: (703) 836-1823	Multibase	Relevant. Details included in document.

15

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Acronym Derived From	None	Intelligent Distributed Information Management System	None	None
Environment (Software and Hardware)	General. SUN 2, 3, 4, 386i: SUNOS, and Mach 2.0 (on SUN 3). VAX Family: Ultrix, VMS, Masscomp RTU, Symbolics Genera, C, Fortran and CommonLisp are supported.	General.	General. The Surround tools have been ported to virtually every UNIX platform including NCR, HP, SUN, IBM, AT&T, Unisys, and Digital for interconnectivity and added functionality.	General. DEC, IBM, NIXDORF
Local Data Models Supported	Object-oriented and Relational	Hierarchical, Relational, Network, and Object Oriented models can be supported through the Database Models Knowledge Base.	All, including flat file, hierarchical, relational and network, via Surround^{SM} interface and coordination tools.	Relational and Network.
Local DBMS Supported	Proprietary object database; possible to integrate other databases; Informix and Oracle are currently supported; vendor has experience with Ingres and Vbase.	See remarks.	IMS, IDMS, Oracle, DB2, and all others on remote systems. Ingres, Oracle, Informix, Sybase and all others on local Surround box.	SUPRA DPDM and VSAM.
Global Data Manipulation Language	SQL Subset	The Query Processor supports a high level, iconic, global query language. Facilities can be added to support natural language, if appropriate.	Any user interface, icons, voice, SQL, graphics, forms, etc.	SQL

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Acronym Derived From	None	None	Common Data Model * Plus	None
Environment (Software and Hardware)	General. PowerHouse 4GL on DEC VAX, HP3000, HP9000, DG MV (AOS) and IBM PC and PS/2 (DOS and OS/2). PowerHouse StarBase, StarNet and StarGate on VAX/VMS.	General. A UNIX-based software product ported to a number of hardware platforms including Pyramid, Gould, Cubix, DEC, NCR Tower, ARIX, ELxsi, CDC910, SUN, and Celerity.	General. CDM*plus operates on VAX/VMS, HP/UX, Pyramid/OSx. CDM*plus accesses applications and databases on the above systems plus IBM/MVs and Amdahl/MVS.	General. Portable - written in 'C'. Server with optimizer/controller/network on a UNIX processor such as SUN. User interface on UNIX system available. DBMS on diverse hardware and OS.
Local Data Models Supported	Sequential, Relative, and Indexed files; Network and Relational databases.	This depends on the UNIX database supported by the UNIX platform being run on.	Relational, Network/Codasyl, and Hierarchical.	Relational and Network. Also, supports objects in files. Qualification to locate files is done on a structured relation that contains the file name. Retrieve process and display process are specific to object type.
Local DBMS Supported	Image3000, TurboImage, KSAM, MPE on HP. C-ISAM on HP UX. RMS, Rdb/VMS, PowerHouse StarBase on VAX/VMS. INFOS, ISAM, INFOS DBAM, DG/SQL on DG AOS. PHISAM C-tree Deriv. on IBM PC.	See above.	Oracle, Ingres, DB2, IDMS on IBM/MVS. Oracle, Ingres, VAX DBMS on VAX/VMS. Oracle on Pyramid. Oracle on HP. Others on request.	IDM (Sharebase) on VAX. Ingres on SUN 120, SUN 170 or other UNIX.
Global Data Manipulation Language	PowerHouse 4GL; high-level operations such as join (inner and outer), restriction (selection) and sorting. Powerhouse StarBase: fully relational DBMS supporting SQL and GROTON DML.	Would have to be programmed using C, and Shell Scripts in combination with *gateway scripts.	Neutral Data Manipulation Language (NDML), a SQL based neutral query language. NDML can be embedded in Cobol, Fortran, and C.	SQL

16

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Acronym Derived From	None	None	Distributed Database Testbed System	DB2 - DATABASE 2. SQL/DS - Structured Query Language/Data System
Environment (Software and Hardware)	General. VAX/VMS.	General. DOS 3.1 through 4.1, OS/2, and UNIX on a SUN platform.	General. Bull-HN's DPS/G and DPS/88; IBM-PC; Charles River's Workstation MOD-400 and GCOS8; DOS; UNIX	General.
Local Data Models Supported	Hierarchical/Network and Relational	Relational (extended). SQL Level I + enhancements are supported.	Network (CODASYL) and Relational.	Relational and Hierarchical.
Local DBMS Supported	DBMS and Rdb/VMS.	SQLBase is both a local and a client-server database. SQLBase is compatible with IBM's DB2 and SQL/DS. Connections for Oracle, DB2, and SQL/DS are performed by Gupta tools.	IDS-II, MDBS-III, and RAM (Relational Access Manager).	SQL/DS on IBM/VM. DB2 on IBM/MVS.
Global Data Manipulation Language	SQL, RDO, DSRI-Rdb, DBMS-NML.	SQL with enhancements (SQL-Talk), pre-compiler for ANSI COBOL, and application program interfaces for C.	GORDAS and SQL	SQL

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Acronym Derived From	FOCUS Network Products.	Interconnective, heterogeneous hardware, operating system and database.	Composite Information Systems.	Data Interpretation System.
Environment (Software and Hardware)	General. IBM PC/DOS, IBM/VM, IBM/MVS, VAX/VMS, WANG/VS, MPE/X, MPE XL (HP), Guardian 90, UNIX, various other platforms and operating systems.	General. Distributed minicomputer and workstations. Networked applications in UNIX and VMS environments. Compatible with VAX/VMS, SUN, Apollo, HP, and OS/2.	General. Written in LISP, C, and UNIX shell. Runs on UNIX System V. DBMS on diverse hardware and software.	General. The DIS workstation can be a combination of PC/XT, AT, or 386s running DOS 3.0 or higher, most PS/2 models running DOS, and PS/2 M70 & 80 running DOS or OS2.
Local Data Models Supported	Hierarchical, Relational, and Network.	Relational.	Relational, Menu-driven, and Hierarchical. Conversion between a local database and CIS is done through a Local Query Processor (LQP) operating on the CIS processor.	Relational and relational interface. Interface to other non-relational DBMS facilitated through TTY, 3270, and 3780 emulators that are invoked through simple mouse point-and-click actions by the user.
Local DBMS Supported	ADABAS,C-ISAM,DATACOM/DB,DBMS, DB2,Dbase,DMS,DMS/TX,MCBA,PACE, SPEED2,FOCUS,IDMS,IDMS/R,Informix,INGRES,ISAM,Model 204, Oracle,QSAM,System 2000,Rdb, RMS,Sharebase, and others.	Interbase on VAX, Apollo, SUN, HP, OS/2 Rdb/VMS.	Informix-SQL on AT&T 3B2. Oracle-SQL on IBM PC/RT. SQL/DS on IBM 4381. Finsbury Dataline (menu-driven Hierarchical) I.D. Sharp Disclosure (proprietary command-driven language).	VAX (VMS) - Oracle, Ingres, Sybase, and Rdb, IBM - SQL/DS and DB2, Teradata DBC, Share-Base 700 and Metaphor DBS. Iconic tools to merge data from PC Dif, ASCII and Lotus files.
Global Data Manipulation Language	3GL, 4GL FOCUS language, and ANSI SQL.	BLR, a lower level language based on the Digital Standard Relational Interface is a message oriented language designed to be generated and read by programs.	Entity-Relation (ER) - like query language.	User points-and-clicks at same iconic interface across all supported relational DBMS. SQL automatically produced. Metaphor gateway servers convert SQL as required for platform.

19

M.I.T. Sloan School of Management
Working Paper Series

Papers by Stuart Madnick

Paper #	Date	Title/Author(s)
1839	11/86	"A Microcomputer-Based Image Database Management System," Prasad, B., Gupta, A., Toong, H., and Madnick, S.,
1776	4/86	"INFOPLEX: Research in a High-Performance Database Computer," Madnick, S., and Hsu, M.
1685	6/85	"Impact of Schedule Estimation on Software Project Behavior," Abdel-Hamid, T., and Madnick, S.
1366	10/82	"An Integrative Approach to Modeling the Software Management Process: A Basis for Identifying Problems and Evaluating Tools and Techniques," Adbel-Hamid, T.K., and Madnick, S.
1334	5/82	"The Dynamics of Software Project Scheduling: A System Synamics Perspective," Adbel-Hamid, T.K., and Madnick, S.
1333	5/82	"A Model of Software Project Management Dynamics," Adbel-Hamid, T.K., and Madnick, S.
1047	4/79	"Properties or Storage Hierarchy Systems with Multiple Page Sizes and Redundant Data," Lan, C.Y., and Madnick, S.
1046	8/78	"Inforlex Data Base Computer Architecture - Concept and Directions," Lam, C.Y., and Madnick, S.
993	5/78	"Compostie Information Systems - A New Concept in Information Systems," Lam, C.Y., and Madnick, S.
920	3/77	"A Systematic Approach to the Design of Complex Systems: Application to DBMS Design and Evaluation," Andreu, R.C., and Madnick, S.
910	3/77	"Trends in Computers amd Computing: The Information Utility," Madnick, S.
883	1/77	"Institutional and AD-HOC Decision Support Systems and Their Effective Use," Donovan, J.J., and Madnick, S.
846	4/76	"Virtual Machines A Concept That has Comparative Advantages in Security, Integrity, and in Decision Support Systems," Donovan, J.J., and Madnick, S.D.

Please Use Attached Form When Ordering Papers

30

M.I.T. Sloan School of Management
Working Paper Series

Papers by Stuart Madnick

Paper #	Date	Title/Author(s)
776	3/75	"An Application of a Generalized Management Information System to Energy Policy and Decision Making: The User's View," Donovan, J.J., Gutentag, L.M., Madnick, S., and Smith, G. N.
771	3/75	"Data Security and Data Processing," Madnick, S.
770	3/75	"Infoplex - Hierarchical Decomposition of a Large Information Management System Using a Microprocessor Complex," Madnick, S.
769	3/75	"Design of a General Hierarchical Storage System," Madnick, S.
723	7/74	"Virtual Information in Data-Base Systems," Folinus, J.J., Madnick, S., and Schutzman, H.B.
722	5/74	"Application and Analysis of the Virtual Machine Approch to Information System Security and Reliability," Madnick, S., and Donovan, J.J.
648	3/73	"An Approach to Information System Isolation and Security in a Shared Facility," Madnick, S., and Donovan, J.
645	2/73	"Recent Technical Advances in the Computer Industry and Their Future Impact," Madnick, S.

Please Use Attached Form When Ordering Papers

M.I.T. Sloan School of Management
Working Paper Series

Papers by Stuart Madnick
John Norris McGuire Professor of Information Technology

Paper #	Date	Title/Author(s)
3157	6/90	"The Composite Information Systems Laboratory (CSL) Project at MIT," Madnick, S., Siegel, M., and Wang, R.
3180	5/90	"The Elusive Silver Lining: How We Fail to Learn from Failure in Software Development," Adbel-Hamid, T.K., and Madnick, S.
3119	2/90	"A Polygen Model for Heterogeneous Database Systems: The Source Tagging Perspective," Wang, Y.R., and Madnick, S.
3120	12/89	"Motivating Strategic Alliance for Composite Information Systems: The Case of a Major Regional Hospital," Osborn, C.S., Madnick, S., and Wang, Y.R.
3100	12/89	"A Polygem Data Model for Data Source Tagging in Composite Information Systems," Wang, R., and Madnick, S.
3084	11/89	"An Architectural Comparison of Contemporary Approaches and Products for Integrating Heterogeneous Information Systems," Gupta, A., and Madnick, S.
3092	10/89	"Schema Integration Using Metadata," Siegel, M., and Madnick, S.
3091	10/89	"CISL: Composing Answers from Disparate Information Systems," Madnick, S., Wang, Y. *et al.*
3061	8/89	"Information Technology Platform for the 1990's," Madnick, S.
3015	5/89	"A Global Query Processing in Composite Information Systems," Wang, Y.R., and Madnick, S.
2075	9/88	"Software Productivity: Potential, Actual, and Perceived," Abdel-Hamid, T., and Madnick, S.
2061	8/88	"Logical Connectivity: Applications, Requirments, and an Architecture," Madnick, S., and Wang, Y.R.
2031	6/88	"Motivating Strategic Alliances for Composite Information Systems: The Case of a Major Regional Hospital," Osborn, C., Madnick, S., and Wang, Y.R.
2029	6/88	"Inter-Database Instance Identification in Composite Information Systems," Wang, Y.R., Madnick, S., and Horton, D.
2025	6/88	"Connectivity Among Information Systems," Wang, Y.R., and Madnick, S.

22

M.I.T. Sloan School of Management
Working Paper Series

Papers by Stuart Madnick

Paper #	Date	Title/Author(s)
2022	5/88	"Implementing Message Oriented Transaction Processing for Distributed Database Management Systems," Bhalla, S., and Madnick, S.
1862	5/88	"Evolution Towards Strategic Applications of Databases Through Composite Information Systems," Madnick, S., and Wang, Y.R.
1981	2/88	"A Framework and Comparative Study of Distributed Heterogeneous Database Management Systems," Bhalla, S., Prasad, B., Gupta, A., and Madnick, S.
1980	2/88	"Modeling the Dynamics of Software Project Management," Abdel-Hamid, T., and Madnick, S.
1977	1/88	"On the Portability of Quantitative Software Estimation Models," Abdel-Hamid, T., and Madnick, S.
1975	1/88	"Facilitating Connectivity in Composite Information Systems," Wang, Y.R., and Madnick, S.
1866	12/87	"A Conceptual Model for Integrated Autonomous Processing: An International Bank's Experience with Large Datebases," Frank, W., Madnick, S., and Wang, Y.R.
1876	4/87	"Hierarchical Timestamping Algorithm," Hsu, M., and Madnick, S.
1875	4/87	"Time Slice Manipulation in Information Heirarchy," Hsu, M., and Madnick, S.
1871	3/87	"An Integrative System Dynamics Perspective of Software Project Management: Arguments for an Alternative Research Pardigm," Abdel-Hamid, T., and Madnick, S.
1868	3/87	"Perspectives on the Effective Use, Planning, and Impact of Information Technology," Madnick, T.
1867	2/87	"The Economics of Software Quality Assurance: A System Dynamics Based Simulation Approach," Abdel-Hamid, T., and Madnick, S.
1864	2/87	"A Temporal and Spatial Locality Theory for Characterizing Very Large Data Bases," Madnick, S., and Moulton, A.
1861	2/87	"Key Concerns of Executives Making I/S Decisions," Madnick, S., and Wang, Y.R.

Please Use Attached Form When Ordering Papers

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Remarks	The RAMIS Information System Product includes a DBMS, a Data Maintenance Facility, a Report Writer, a compiled application development language, and a developers toolkit. The developers toolkit provides dictionary utilities, application creation, and relational capabilities. Also, a full function 4GL/DBMS for the personal computer with Micro-Mainframe communications is available.	Oracle provides true heterogeneous, distributed processing. Product is designed to give developers flexibility to build and customize database applications easily and quickly. Additional tools are available for the end-user who is not equipped or interested in learning the SQL*Language.	Relational Technology is committed to deliver the best performing and highest quality relational database system on the market. This combined with a unique distributed open-architecture solve the most difficult data management needs.	The Software AG Product line allows for creation of truly distributed data processing systems with full user transparency.
References	Furnished upon request.	Furnished upon request.	Furnished upon request.	Federal Aviation Admin. Mr. Rod Phillips (405) 680-3422 U.S. House of Representatives Mr. Jim Daley (202) 226-6426 Center for Disease Control Ms. Carolyn Seagraves (404) 639-0475.

23

	SYBASE: SYBASE	TANDEM: NonStop SQL	TRW: TDIE	XEROX AIT: MULTIBASE
Remarks	Open server designed to provide users with toolset for building custom servers to heterogeneous databases.	As many as 255 systems (4,080 processors maximum) can be linked to form a geographically dispersed network.	Designed to provide the complete spectrum of integration needs from automatic updates to query-only extracts in a heterogeneous data environment.	Continuing work on internal funding, looking for strategic partners to commercialize product.
References	Air Force Military Airlift Command Air Force Strategic Air Command Air Force Intelligence Agency	Tandem Reference Material - Introduction to NonStop SQL - NonStop SQL Installation and Management - NonStop SQL Programming Reference Manual	U.S. Patent Number 4,714,995 awarded to TRW on Dec 22, 1987.	Furnished upon request.

24

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Remarks	Digital claims Rdb/VMS is the best relational database on the VAX.	Claims to excel in personal computer Relational database management, personal computer application development, and host connectivity for access to DB2.	Work redirected to data dictionary and engineering tools area due to lack of corporate interest in productizing DDTS.	Referential Integrity ensures the consistency of data values between related columns of two different tables. DB2 and SQL/DS participate in Systems Application Architecture (IBM exclusive) permitting data access across processor platforms. Data from DB2 and SQL/DS may be accessed/updated from IBM's knowledge based system products.
References	Furnished upon request.	Furnished upon request.	Pat Dwyer and Jim Larson "Some experiences with a Distributed Database Testbed System", Proc. of IEEE, Vol. 75, No. 5, May 1987. Elmasri et al., "Notes on DDTS", ACM SIGMOD Record, July 1981.	Furnished upon request.

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Remarks	4GL FOCUS language is the most widely used in the world. IBI claims to be a leader in Hierarchical, Relational, and Network data integration.	Designed to handle On-Line-Complex-Applications (OLCP). "Complex" is described in terms of the transaction types and intervals, the datatypes and the transparent network processing features of the database. Advanced as compared with traditional On-Line Transaction-Oriented products.	CIS was designed to be non-intrusive and support integration amony very disparate databases. It is part of a larger effort that addresses both advances to theory as well as deeper understanding of the uses and application needs.	Each database possesses no knowledge of the existence of another database within the Metaphor DIS. The object oriented application development environment provides rapid prototype/development capabilities to integrate the results of any number of database queries with each other and with data acquired through client developed personal computer and VAX based applications.
References	Furnished upon request.	Furnished upon request.	Wang, R. and Madnick, S., "Facilitating Connectivity in Composite Information Systems" to appear in ACM Database. Wang, R. and Madnick, S., "The Inter-Database Instance Identification Problem in Integrating Autonomous Systems", Proceedings of the Fifth International Conference on Data Engineering, Feb. 6-10, 1989.	Furnished upon request.

26

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Remarks	Funded by the Rome Air Development Center (RADC) and the Naval Ocean Systems Center (NOSC).	IDIMS eliminates the need for the user to know query languages, locations of data, database paradigms, or manipulation schemata. IDIMS is not tied to any hardware/software platform.	Surround^{SM} allows end users to drive the design process, developers to implement applications rapidly, existing systems to be preserved, and systems integration to be done on an open systems platform.	An open protocol will be published to allow other vendors to develop interfaces to database management systems not currently planned to be supported by Cincom.
References	Furnished upon request.	Furnished upon request.	Furnished upon request.	Furnished upon request.

27

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Remarks	The 4GL provides transaction rollback on flat or index file systems, but makes use of database rollback facilities on relational systems. An optimistic concurrency policy is used to maximize on-line throughput. The 4GL includes a report writer, form/screen transaction manager, volume (batch) processing facility, data dictionary and other components.	According to CDC, it is hard to describe the power of ASCENT because it is versatile.	The CDM*plus enables integration through centralized modeling, rigorous mapping between existing and conceptual schema, and robust process-to-process communications capabilities. In addition, it provides a suite of tools for analyzing the impact of change, and bridging to popular modeling tools. It also has extensive reporting facilities.	Designed to be modular, portable, and easy to customize. Written as processes with well-defined interfaces to support replacement of components such as different user interface, query languages other than SQL, output into an application or a database, data dictionaries in different database management systems and diverse local database management systems.
References	Furnished upon request.	Furnished upon request.	The CDM*plus product resulted from Integrated Information Support System (IISS). The first document, in a set of 120, that describes the components of IISS is: DApro Project IISS, Document Number: FTR620300001. Obtained at MANTECH DIR, Integration Tech. Div., Attn: D. Judson, WRDC, WPAFB, Ohio 45433.	M. Templeton, P. Ward, E. Lund, "Pragmatics of Access Control in Mermaid", the IEEE Technical Committee on Data Engineering, Sept. 1987. M. Templeton, "Schema Translation in Mermaid", submitted to conference, Aug. 1989.

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Current Installations And Applications	The RAMIS Information System has approximately 1600 installations in the United States.	Installations exist within the Department of Defense, Federal Civilian Agencies, and a wide range of commercial entitities for numerous applications.	Sample of customers using INGRES/Star and/or Gateways: Corning Incorporated General Motors Morgan Guaranty 3M Corp. Carnegie Mellon University ICI Americas Scotland Yard Sierra Geophysics	True distributed processing is currently being implemented at Software AG Darmstadt, Germany. Software AG products are currently installed at over 3,000 customer sites. Specifically, in the Federal Government, Software AG products are represented in nearly 100 agencies and departments.
Current State Limitations And Future Directions	Future direction lies mostly in the DB2 and SQL/DS arena, and in complete SAA compatibility and compliance.	Not all application packages are ported to all 90+ hardware platforms; ongoing effort to resolve this. Specific company directions are to port Oracle across the IBM product offering.	Future enhancements include: - Automated replication of global model to other nodes. - Deferred "snapshot" copies of database tables copied to other nodes. - Additional support of unusual native databse data types. - Concurrent writes to multiple local databases under transaction control (i.e., 2-phase commit). The current INGRES/Star supports multiple-database reads and writes to a single local database under transaction control.	Software AG provides state of the art tools, implemented using 4th generation technology, which are portable across multiple system environments through its Open Integrated Architecture (ISA).

	SYBASE: SYBASE	**TANDEM: NonStop SQL**	**TRW: TDIE**	**XEROX AIT: MULTIBASE**
Current Installations And Applications	Military Intelligence Organizations. Military Command and Control. Environments similar to the task of integrating TRANSCOM's subordinate commands.	Currently installed in a number of Fortune 500 companies. Applications in various industries including finance, manufacturing, government, health services, insurance and marketing.	The TDIE is in production use within the Manufacturing Div. of TRW. It integrates a Process Planning System on a VAX with a Job Shop Control system on an IBM.	Multibase is installed as a testbed at Cecom at Fort Monmouth.
Current State Limitations And Future Directions	Open server is in beta testing and is single-threaded. Connection from local database to an open server allows access to all front-end tools available for the Sybase Relational DBMS. Future directions will greatly expand network support, performance via multi-threaded processing, and commercial-off-the-shelf gateways to Model 204, DB2, and ANSI-SQL databases.	Currently implemented only on Tandem NonStop Systems. Integration agreements have been signed with Oracle Corp., Sybase, R.T.I. (INGRES), and I.B.I. (Focus) to have their toolsets ported to Tandem Systems. This will give users access to NonStop SQL through one of these front-ends.	The TDIE is a TRW commercial product. Currently implementing a front-end tool to the TDIE called the Integration Advisor to allow easier construction of the integration interfaces.	Multibase does not currently support updates. Xerox AIT is currently accessing market needs for this capability.

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Current Installations And Applications	Worldwide general purpose, multi-user, centralized or distributed applications.	Over 1800 to date; many in major corporations and government agencies.	Currently installed as a prototype or testbed. A personal data query and a shop-floor control application have been implemented.	Worldwide. There are 6000 licenses for DB2. It is the leading DBMS worldwide. There are 2000 SQL/DS licenses. Sample applications include forecasting, tracking, banking, human resources, sales reporting, etc.
Current State Limitations And Future Directions	Currently, there is only location transparency available for a single physical site. Location transparency for multiple physical sites is desirable.	SQLNetwork provides transparent access to IBM's mainframe Relational DBMS (DB2) that allows users to interact with DB2 and SQLBase from their personal computers without any coding changes. Future releases of SQLNetwork will provide cooperative processing access to VMS based Relational database management systems (Oracle, Ingress, and Sybase) and IBM's SQL/DS in the VM environment.	Testbed only; distributed updates are not generally implemented on a variety of data management systems; future work will focus on distributed data dictionaries, schema integration tools and real-time performance.	DB2 only operates in the MVS environment and SQL/DS only operates in the VM environment. Future directions include: - Consistent SQL across platforms (PS/2, AS/400, S/270). - Additional distributed data/processing functions. - Improved optimization, index management, and data management. - Higher availability (raised limits for table and database sizes, and more concurrent utility and application access.

31

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Current Installations And Applications	FOCUS and FOCNET are used in financial, academic, and government areas, including manufacturing, budget tracking, and General Ledgers. There are approximately 4,000 copies of FOCUS used on mainframes and micros; approximately 125,000 copies of PC FOCUS have been sold.	InterBase has roughly 1,500 licenses in various industries including aerospace, electronics, financial trading, and manufacturing.	CIS has been used internally at MIT since 8/89 to support the Placement Assistant System (PAS) using the remote databases described above.	Approximately 10,000 workstations in use across 200 government and commercial clients in the U.S. and Western Europe such as Army Deputy Chief of Staff, Operations, Bureau of Labor Statistics, Postal Service, Transportation Analysis, and Coast Guard for Helicopter Readiness. Also, many Fortune 500 companies such as Proctor & Gamble, Chevron Chemical, and British Telcom.
Current State Limitations And Future Directions	Connectivity, front end tool development, and artificial intelligence are ongoing efforts.	It is not possible to have views which span more than one database, or triggers which modify tables in a different database. Work has begun to define a global schema which would allow these features.	CIS version 3.0 was completed in August 1989. Certain components have been converted from LISP to C and shell for efficiency. Current CIS supports access to all the databases described earlier. Studies of strategic, organizational, physical, and logical connectivity are continuing.	Metaphor plans to support every popular relational DBMS that offers an SQL Interface. The same or similar iconic access to non-relational databases is under consideration by one or more firms currently providing gateway technology and software for access to Metaphor's DIS.

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Current Installations And Applications	Currently installed at 15 sites on over 150 machines. Applications include office automation, distributed simulations, expert system decision aides, and an intelligent document retrieval system. C, Fortran, and Commonlisp languages are supported.	Components of IDIMS have been developed by Booz Allen for the FAA, Air Force, TSC, TRANSCOM, Army, etc.	Over 200 different client applications including: Cigna Insurance's Investment Management Reporting System, Royal Caribbean Cruise Line's CruiseFax, and Sheraton's in-hotel reservation system.	Furnished upon request.
Current State Limitations And Future Directions	Not currently a commercial product. System includes documentation, tutorials, installation manuals, and telephone hot-line and electronic mail bug reporting and diagnosis. Future directions: increased software development toolkit support, additional replication algorithms, more extensive WAN support.	See remarks.	Future directions include releasing tools to handle complex image as input.	Base software is in general release. Local update supported now. First network implementation in next maintenance release.

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Current Installations And Applications	Over 12,000 installations on HP VAX and DG employed in a wide range of commercial applications including order processing, inventory control, health care, payroll, financial management, etc.	ASCENT and LLNL's TIS software are installed in over 60 places within the government and commercial world. ASCENT was recently included in the AFCAC 251 contract awarded to AT&T by the Air Force recently to be run on the AT&T 3B2-600 Gs. TSC currently has the TIS version running on several of DLA's DMINS Gould computers.	In use at: Northrop Aircraft Division, G.E. Aircraft Engine, US Air Force Materials Lab., US Navy Air Depot Operation Ctr. Used on the following Air Force Projects: Eng. Change Analysis System, Integrated Sheet Metal Center, Precursor Computer Aided Acquisition and Logistics System.	Company incorporated 7/89. Mermaid software is owned by Unisys and licensed to Data Integration for product development and extensions.
Current State Limitations And Future Directions	PowerHouse StarBase, StarNet, and StarGate were first released in Jan/89 on VAX/VMS and are not yet available on other platforms. VMS-VMS and UNIX-VMS connectivity is available. Interconnection of many other machines is working in the lab, but is not available for release. PowerHouse 4GL cannot be invoked from 3GLs, but can call 3GL subroutines. PowerHouse StarBase provides embedded SQL and GDML support for C, COBOL, and FORTRAN programs (Basic, Pascal and Ada are working in the lab). 3GL applications can access PowerHouse StarBase, Rdb/VMS and RMS.	There has been some thought given to combining the strength of ASCENT with the work done on the IISS system at Wright-Patterson AFB. This combination of technologies would replace the current communications methods used by IISS with the *gateway. It would also replace the precompiled Cobol programs in the remote hosts with scripts thus removing the requirement for software in the remote host.	CDM*plus does not support IMS access except through the use of Oracle's IMS/Connect facility which is under development. The IMS access issue is of major concern within the DOD/Aerospace community. The modeling and the integration of applications and databases is done in an evolutionary manner. Additional DBMS, applications, and computers may be added by user request. Customized implementations are available.	Mermaid has been recoded to move from prototype to product. Data Integration is now completing support utilities and manuals. Specific DBMS interfaces may be developed under contract with one user and then licensed to others. Other customization may be done under contract.

35

	SYBASE: SYBASE	TANDEM: NonStop SQL	TRW: TDIE	XEROX AIT: MULTIBASE
Extent of Changes Required to Local Databases (Modification Of Data, Programs or Modules Required To Be Resident, etc.)	None. (Answer applies to the general case of a heterogeneous data source, whether it is any kind of database management system or not).	No change is required.	No change is required to any database. An extractor or update capture program may be needed on the source, and an updater program may be needed on the target side.	No change is required.
Specific Components Of Product Or Service Currently Under Development	The Sybase OPEN-SERVER is in the final stage of beta testing. The Sybase multi-level Secure Server is in the final stage of beta testing. This server is the database in DOD's national MLS SYSTEM TESTBED. Support for Oracle is planned, DB2, M204, and IDMSR are under development.	Release 2 of NonStop SQL is currently in test and improves parallel processing capabilities for performance. Support of nulls will also be provided in that release.	The Integration Advisor is currently under development with an initial prototype expected in December 1989.	A reanalysis of the architecture for commercialization is under development.
Approximate Price Range Of Product And Services	Competitive with major relational DBMS products.	Prices depend on the number of processors: they range from $4000 to $64,000 for 1 to 16 processors.	Price includes installation and is dependent on the complexity of the site specific installation.	Custom software development has varying prices.

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Extent of Changes Required to Local Databases (Modification Of Data, Programs or Modules Required To Be Resident, etc.)	No change is required.	SQL*Net for network communications; SQL Loader for data import.	None, except to resolve any unusual local DBMS data types that are not supported in INGRES. For non-INGRES local database management systems, the appropriate DBMS Gateway module must reside on the local host (eg, the Rdb Gateway). These Gateway modules are available with INGRES.	The database administration or system administration defines all files and views to PREDICT, with unique database identifiers. With NETWORK and ADANET, no change needs to be made to the applications or the data. To the user, all data access is completely transparent.
Specific Components Of Product Or Service Currently Under Development	All components listed are fully functional and have been released. Continuous enhancements are added yearly.	Basic software offering is complete. Ports to additional hardware platforms is an ongoing process. Enhancements to distributed transaction processing, two phase commit, and global updates for additional platforms.	INGRES/Star-global distributed database creation and maintenance module. INGRES/Net-communications layer supporting communication between heterogeneous hardware via DECNet, TCP/IP, LU 0. LU6.2 support is planned for the near future.	ADANET currently under development. Planned to be released 1/90. VIEW PROCESSOR currently under development. A release date planned in 1990. VIEW PROCESSOR is only necessary for multiple user views.
Approximate Price Range Of Product And Services	Price ranges depend on the operating environment. A Reporter ranges for $19,000 to $45,000. The full RAMIS Information System (including DBMS) ranges from $49,000 to $126,000.	Pricing varies greatly based on hardware platform.	RMS Gateway $ 500-$ 35,000 Rdb Gateway $1,200-$ 83,000 INGRES/Star $1,000-$ 71,000 INGRES/NET $ 500-$ 35,000 Protocol Spt (each) $ 150-$ 12,000 DB2 Gateway $1,500-$106,000 SQL/DS Gateway $1,500-$106,000	Depending on product, size of machine, and operating system utilized, prices vary from $700 - $139,200.

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Extent of Changes Required to Local Databases (Modification Of Data, Programs or Modules Required To Be Resident, etc.)	No change is required.	No change is required. SQLBase has local and client-versions that are identical.	No change is required except for distributed updates which require compatible commit protocols.	Inventory dependent.
Specific Components Of Product Or Service Currently Under Development	Proprietary.	SQLNetwork connections to IBM's SQL/DS on the mainframe and connections to Relational database management systems on minicomputers, a UNIX version of SQLBase, a pre-compiler for C, and Lotus 1-2-3 and Excel add-ins for direct access to SQLBase from spreadsheets are under development.	Data dictionary tools based on ANSI-standard Information Resource Dictionary System.	In addition to the pure IBM, non-UNIX solution, IBM is developing a new relational UNIX database. This database will integrate the AIX line (PS/2, RT, S/370), as well as communicate with the SAA database line (PS/2-OS/2, AS400, S/370).
Approximate Price Range Of Product And Services	Run time only on remote nodes are included in the VMS license. A development license ranges from $5,000 to $120,000.	SQLHost - $20,000. SQLGateway - $1,995 for each SQLGateway. SQLBase single user software development kit - $1,295. SQLBase multi user software development kit - $2,995. SQLRouter - No additional charge.	DDTS is still a testbed; not a product. A variety of services in R&D in this area at a variety of price ranges are available, as long as they are in Honeywell's strategic areas of interest.	Prices depend on model. DB2 - One time charges range from $113,400 to $244,850. The monthly lease charges range from $3,210 to $4,530. SQL/DS - One time charges range from $15,980 to $47,140. The monthly lease charge is $761. Educational and volume discounts are available.

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Extent of Changes Required to Local Databases (Modification Of Data, Programs or Modules Required To Be Resident, etc.)	No modification to application or data is necessary to enable remote access and cooperative processing. However, if operating system calls are embedded in application, these calls must either be replaced or removed.	The Oracle gateway requires an additional table to be defined in the database, access to some tables in the data dictionary granted, and two servers to be started on the target node. Digital Equipment Corp.'s Rdb requires access to the files through VMS.	None. Local autonomy is a key design criterion. CIS is based on two key design principles: (1) system non-intrusiveness (2) data non-intrusiveness	Host Interface language.
Specific Components Of Product Or Service Currently Under Development	FOCNET consists of two types of servers: the data server, and the programmable application server. The data server is available today in all environments. The programmable application server is available between VAX and IBM, PC and IBM. New implementations will be released in 1990.	The Oracle gateay is currently in beta test and is scheduled for the release in December 1989 (Oracle on VAX, Apollo, and SUN is planned). All other parts are commercially available.	Many new components are under development, such as: graphical user interface, automated data semantics synchronization, semi-automated interdatabase instance identification, LQP generation, interfaces to new databases (LQPs).	DB2, and AT&T 3B2/Oracle gateway servers (BetaTest). Microchannel workstation compatibility.
Approximate Price Range Of Product And Services	FOCNET is approximately $100,000 for a mainframe, $50,000 for a minicomputer, and $1295 for OS/2.	Price per license ranges from $5,000-$95,000 (U.S. list) depending on configuration.	Depends on complexity of the environment.	$65,000 - $400,000.

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Extent of Changes Required to Local Databases (Modification Of Data, Programs or Modules Required To Be Resident, etc.)	No change is required.	No change is required to local databases. IDIMS queries each of the databases in their native language, based on their structure, semantic and syntactic data stored in the database models knowledge base.	No change is required.	Local DBMS must handle indices. Drop and recreate entity. Local platform must host SUPRA Kernel.
Specific Components Of Product Or Service Currently Under Development	Extension of network monitoring service, programming support for ADA, security enhancements, performance optimizations, new replication algorithms, porting to new hardware platforms, resource management algorithms.	Each group of databases and functions must be combined in a different manner, as such Booz Allen has developed systems, such as databases and expert systems, that demonstrate their capability in each area.	Extensions to existing application network management capabilities. Incorporating ISDN interface to PBX to enhance existing telephony capabilities of the environment. Enhance capability to deal with image data.	IMS/DB in detailed design phase. Supra PDM in implementation phase. VSAM completed. Planned local database support: Supra PDM, DB2, RDB, RMS, IMS, DL/1, SQL/DS.
Approximate Price Range Of Product And Services	No license fee for the government or its contractors. Maintenance service is approximately $35,000/year; optional. Application development support provided at cost.	Dependent on application.	For guaranteed fixed price delivery of strategic applications prices range from $150,000 - $8,000,000 for development. On going maintenance is 15% of development cost.	Licensed on a per CPU basis, varies by CPU size. $30,000 to $200.000+.

39

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Extent of Changes Required to Local Databases (Modification Of Data, Programs or Modules Required To Be Resident, etc.)	Usually none. Some complex data structures may not be readily accessible using PowerHouse (e.g. graphs, trees, records with variable-length variants, etc.). Once the software is installed, most supported database systems will be readily accessible.	No change is required.	The CDM*plus 3 schema architecture requires only the loading of the definition and structure of the local database, thus no change is required.	None. Local autonomy is a key design criterion.
Specific Components Of Product Or Service Currently Under Development	StarBase and StarNet on HP3000 and HP9000. Gateways to various database systems (e.g. Oracle). New end-user query tools for decision support. A CASE front-end tool for database design.	The DECNet interface currently only runs on Pyramid computers and operates in a "stand-alone" product mode. This capability is being integrated within the standard *gateway.	Release 2.0 available in late 1989, provides for full DEC-DEC, and DEC-IBM processing. Features available in 1990 include TCP/IP support, additional bridges to CASE tools, expanded dictionary capability, and additional DBMS access.	Data dictionary builder and utilities for user installation, recovery, and monitoring are being rewritten. Additional DBMS interfaces are being developed. M204 started on government contract, but not completed. DB2 and Teradata interfaces are under development.
Approximate Price Range Of Product And Services	First copy is $5,000 to $500,000 for Digital VAX. Discounts for multiple copies, run-time versions, etc.	Price is determined by the relative size and power of the UNIX platform running ASCENT. The one-time-charge for a PC-sized platform is $1,000. The one-time-charge for Super-mini-sized platform is $30,000.	The CDM*plus basic product consists of the CDM*plus dictionary, NDDL, CDM*plus Tools (reports, DDL translators, Impact Analysis), NDML, 7 days of training for up to 15 individuals, and complete installation and support. The cost for the basic product is $150,000 for a VAX/VMS 8600 version.	Server for 8 users with a user interface, controller, and data dictionary is $30,000. Each DBMS interface with a translator is separately priced from $5,000 on a mini to $20,000 or more on a mainframe.

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Global Update And Transaction Management	The RAMIS Data Maintenance Facility (DMF) provides a way to load files and to modify, browse, and delete data in RAMIS databases and DB2 and SQL/DS tables. A non-procedural language is used to accomplish this. In one request, data may be added, deleted, modified, verified and logged. Transactions may be accepted from various sources such as the screen, sequential transaction files, and from the personal computer in fixed or comma delimited format.	Offers data and referential integrity for data concurrency, before and after image files and data segment rollback for recovery. Global updates and two-phase commit supported on certain platforms.	Current verion of INGRES/STAR supports multi-database reads and updates to a single local database under transaction control. Coming version (beta in Sept., 1989) supports multi-database updates (i.e. 2-phase commit), including full multi-size recovery in the event a local node or the coordinator node should fail.	ADANET supports distributed transactions which can be a series of updates to different databases. They may involve more than one node within a network. When a "Commit" call is received, ADANET synchronizes all updated databases, and checks them for the eventuality of the restart/recovery procedure.
Network Services	Personal computer Networking. LAN support includes: BANYAN VINES, IBM personal computer LAN, Novell NETWARE, and any LAN that is 100% compatible with DOS 3.1.	Async, Ethernet, TCP/IP, 3270, STARLAN, LU6.2, DECNet, CMX, DCAM, NetBIOS, APPC, FTP, EXCELAN, INTERLAN GATEWAY, TOWERNET (Depends upon the hardware platform being used).	INGRES/STAR via INGRES/NET. Supports thin and thick wire ethernet using DECNet or TCP/IP protocols, and SNA-LU0.	VTAM, CTCA, DECNet, Hyperchannel, Interlink, TCP/IP and combinations of SNA Gateway.

42

	SYBASE: SYBASE	TANDEM: NonStop SQL	TRW: TDIE	XEROX AIT: MULTIBASE
Global Update And Transaction Management	Distributed updates to other Sybase servers available via "two-phase" commit protocol. Does not currently support transactions at the global level; however vendor has committed to do so in the future.	Full global update and transaction management is supported. Any column/row of a distributed database can be updated regardless of physical location. Transaction Management is provided by the Transaction Monitoring Facility (TMF). Each node within the network monitors transactions on the local system, as the local database transaction completes it sends a message to the originating system which logs a complete transaction after hearing from all nodes.	Automatic updates are permitted and a cooperative access framework based on "ownership" is utilized.	Not supported.
Network Services	Supports the standard network protocol for the host environment (e.g. TCP/IP on SUN hardware, and DECNet on VAX hardware). Committed to expand network support. Currently provides byte-swapping between protocols supported.	A distributed database is implemented using an EXPAND network which can operate over dedicated lines, leased lines, microwave links, satellite links, X.25 PDNs and Ethernet.	SNA, Ethernet, TCP/IP, MAP/TOP, Async.	Global Data Manager (GDM) and Local Data Interface (LDI) incorporate network interface for establishing communication.

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Global Update And Transaction Management	Provided by Rdb/VMS and DBMS.	Currently supports multiple databases and multiple cursors per transaction. Two phase commit to be supported in the first quater of 1990 .	Transaction management is done using a distributed execution monitor and a local execution monitor. Updates to certain versions of certain databases are supported.	Supports: Remote request - Support for single SQL requests. Remote unit of work - Support for several SQL requests within an application program. All requests within a commit scope address a single remote DBMS. Distributed unit of work - Support for several SQL requests within an application program. Each request addresses a single DBMS. The requests within a commit scope can address more than one remote DBMS. Distributed request - Request may address more than one DBMS. Two Phase Commit.
Network Services	DECnet.	NetBIOS, TCP/IP, LU6.2/APPC, SNA/SDLC, all major LANS via NetBIOS, Lan Manager, and LAN Server.	Communication interfaces based on TCP/IP and MAP protocols are supported. DSA protocols are supported for long-haul communications.	SNA includes: - Advanced peer to peer communications (APPC/LU6.2) - Low entry networking (LEN) - Token Ring/Local Area Network (TR/LAN) - Synchronous Data Link Control (SDLC) - X.25.

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Global Update And Transaction Management	FOCUS utilizes the check/verify model, enabling multiple users to view the same record simultaneously. The check/verify model tracks the changes of the record by keeping the image of the record in a scratch pad area, and then informs users if the record has been changed.	Multiple databases can be updated in the same interactive session. From a program, two additional features, multiple transactions and automatic two-phase commit, can be used to make multiple updates conditional on the state of individual nodes. Transactions can be tailored to commit all, commit some, rollback some, or rollback all data depending on the need of the individual application.	Present version only supports data retrieval. Future development will provide various levels of global update and transaction management facilities.	Not supported directly. Metaphor's object oriented application development environment can facilitate rapid development of applications involving small numbers of updates. The DIS includes iconic data entry and modification tools. Direct host based processing is more efficient for high volume transactions.
Network Services	FOCNET supports LU 6.2, LU 0, LU 2 (3270 data stream), TCP/IP, and NetBios. It runs on EtherNet, TokenRing, IBM channels, and VTAM compliant communication layers. It is, in the case of LU6.2, compliant with SAA, using APPC and CPI-C facilities. It converts protocols when needed; routes messages, controls queuing and task management; authenticates users on networks; works under security subsystems; compresses and encrypts data; and handles error logs.	InterBase has a standard message oriented protocol for communications between distributed databases. Compatible with TCP/IP, DECnet, and Domain.	Uses Ethernet and TCP/IP for Local Area Network access and asynchronous ASCII communications to remote databases. There is a flexible "communications server" module that makes it relatively easy to add new network services.	The DIS is LAN-based. It uses Ethernet and XNS for Local Area Network access which can coexist with TCP/IP, DECnet, and other 802.3 protocols. There are bridges to Token-Ring LAN. WAN communications can be via SNA, APPC internetwork routing, and X.25.

44

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Global Update And Transaction Management	Updates my be performed on an object by object basis.	IDIMS does not update or alter data in any of the databases. Transaction Management is not currently addressed by the system.	Global updates performed through Data Handler and communication tools on remote systems via database and database server in a local database. Transaction management function performed by proprietary tool, Commander, used for process monitoring, audit trail, and transaction processing.	Secure only when Supra database management systems are involved. May be secure for other database management systems that support two phase commit. Insecure for database management systems that do not support two phase commit.
Network Services	Synchronous and asynchronous remote procedure calls, data representation translation, network monitoring service, LAN and WAN, and dynamic resource location. Underlying transport mechanism is TCP/IP.	Database queries over a network are made transparently to the user. Actual networks supported are dependent on implementation.	All network protocols supported including: IBM/SNA (3270, RJE, LU6.2), BSC, TCP/IP, async, X.25, ISDN. At the application level, Surround is independent of network protocols.	LU6.2, TCP/IP, and DecNet

45

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Global Update And Transaction Management	Updates across multiple databases are permitted. PowerHouse StarBase supports the two-phase commit protocol to guarantee the automaticity of such updates. Automaticity cannot be guaranteed for other participants. In practice, most applications can be distributed in such a manner as to avoid the need for consistency across two or more databases.	This would have to be handled by the DHDMS. *gateway would just be a tool if required to facilitate the connection to remote systems.	CDM*plus was designed to allow and support distributed global updates. Referential integrity is enforced, however, multi-node updates can not be guaranteed.	Transactions that can be passed to one database are permitted. Replicated copy updates go to primary copy and then to secondary. Fragmented relations are updated without guaranteed concurrency. Read concurrency is not guaranteed, but it does make selects and projects of all relations in parallel as a first step to minimize time between read of local relations.
Network Services	PowerHouse StarNet supports DECNet and TCP/IP. OS/2 LAN Manager has been developed but is not yet commercially available. MAP support has been prototyped in the lab.	*gateway supports three types of connections: asynchronous, TCP/IP 802.3 ethernet, and DECNet. The actual network services are provided by UNIX.	CDM*plus uses LAN and WAN communications within the SNA and DECNet arena on currently supported hardware. TCP/IP is under development and will be available in the first half of 1990.	Uses Ethernet and TCP/IP between user interface and Mermaid controller. Controller uses TCP/IP or another process to process protocol (LU6.2 planned) between it and the local database host. All communication between local database translators goes through the controller to avoid many to many protocol problems. Translator may also reside on a UNIX server using TCP/IP to the controller, but a special protocol to the DBMS host.

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Global Schema Construction	FILE EXPERT, a full screen field description editor, is an integral part of the RAMIS Information System allowing for modification, creation, and deletion of data descriptions. Automatic load from dictionaries and COBOL copy books provided where necessary.	Imported database schemata converted into relational format may be further analyzed to ensure proper naming conventions, dependencies and data definitions. No Global Schema exists or is necessary.	Global database is constructed manually using SQL to define "links" to local databases. Links reference node, local database name, data type, and local database table (dataset) name. (Future enhancements include automated maintenance of global schema to reflect changes to local schemata, and replication of global schema to other nodes).	ADANET supports replicated, distributed, and isolated files. Also, it will support multi-file user views (ie., multiple physical files viewed as one logical file) through a product called VIEW PROCESSOR currently under development.
Query Processing (Retrieval)	RAMIS Reporter, a comprehensive report writer, provides for report development accesibility in four ways: non-procedural syntax, menu based, checklist, and ENGLISH (a natural language interface to RAMIS Reporter based on AI principles). All four methods allow for sorting, selecting, manipulating, and formatting of data to produce reports ranging from simple to the more complex production type.	Once data are imported from the non-Oracle based databases, queries may be performed via the SQL Language. Direct querying by non-Oracle based systems, besides those running DB2 or DBase III, is not permitted at this time.	SQL queries on the Global Distributed Data Dictionary are decomposed into local database SQL queries. The statistically-based distributed query optimizer manages the processing of queries at the local nodes by developing query execution plans that consider network costs. Thus, interim processing of a complex query may occur at the coordinator node. The distributed query optimizer enables joins across all the local database management systems supported by INGRES gateways.	All logical files and user views are defined in PREDICT, with unique database identifiers. An SQL or NATURAL query written using this view will be dynamically routed to the appropriate database at execution time. Translation of a user request is done through ADANET tables.

	SYBASE: SYBASE	TANDEM: NonStop SQL	TRW: TDIE	XEROX AIT: MULTIBASE
Global Schema Construction	On-line active data dictionary maintains relational model for Global Schemata. Local database schemata, subject to schema and data translation mechanisms, are maintained in Global Schema in relational form.	The idea of a Global Schema does not exist. Users create local schemata and can access remote schemata with standard data manipulation language. Users have access to distributed databases and can execute distributed joins to access a logically local database.	A front end tool called the Integration Advisor allows the interactive construction of a global integrating schema.	Uses the Functional Data Model to define the global schema and DAPLEX as the GDML. The local DBMS and their local host schema which are expressed in terms of the Global Data Model. The local schema are merged into global schema with an auxiliary Integration Schema which contains information for reconciling inconsistencies between local schemata.
Query Processing (Retrieval)	A query in SQL is passed directly to the SQL server. All other queries are executed as parameters to a stored procedure which triggers an event handler to a request written in the native database language which accesses the local database.	SQL queries are processed on the local database system. If a query consists of data from a distributed database, each system processes the query against the local database and transmits only the selected columns/rows to the querying system for final processing.	A local database SQL query can be processed against a remote IMS database (retrieval only).	A query expressed in DAPLEX on Global Schema is decomposed into queries on DAPLEX Local Schemata. These are further translated into Local DML queries. Adopts "no two variables in a query range over the same entity" strategy to generate queries on Local Schemata. Adopts data reduction and data movement techniques in Global Query optimization strategy. Merging and final processing takes place at Global Schema result node.

48

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Global Schema Construction	Uses CDD/Plus dictionary.	Supports full Relational catalog. Several tools for definition of database design are available.	Global schema is constructed manually and populated using dictionary files into the data structures of the system.	A single global schema does not exist, yet many composite databases may be defined.
Query Processing (Retrieval)	Query processing is done through an optimizer which takes into account Table Cardinality, Index Cardinality and type of Query. Directions on how to process queries can be specified in Network Manipulation Language (NML).	Full SQL support for queries from various query tools (SQL-Talk, WindowsTalk, Express Windows, APIs, SQLWindows, etc.). Read only capabilities exists in database.	E-C-R based GORDAS is decomposed into internal, relational algebra based query trees. These query trees are then optimized based on local processing and network costs into an execution strategy with the right communication commands embedded in the query tree. A SQL query interface is supported. Also, SQL queries will be converted into GORDAS queries first. The sub query-trees are converted at each local site into local DBMS commands.	Queries are optimized for minimal communication cost. Queries can be submitted against any number of composite databases provided the user is authorized to do so.

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Global Schema Construction	Several facilities exist in the FOCUS language to create schemata. One is File-Talk--a window driven facility to interactively build an MFD. Other examples are AUTOSQL and AUTOIDMS which directly take data information from database catalogues. The ACE (Application Control Environment) reconciles application and security requirements across the network.	Global Schema does not exist.	There is a single global schema and multiple user-oriented application schemas. The global schema is created by the database administration and can be edited and modified online.	Local schemata of one or more databases are graphically depicted for the user. If a particular application needs to aggregate data (eg. perform a join) across two or more physical databases, Metaphor's object oriented development environment (the capsule) can be used.
Query Processing (Retrieval)	The FOCUS language can directly access the aforementioned databases. It can also dynamically create views of new fields at the database level or summation level. Processing is optimized based on the functionality of the database, which includes record-by-record and set processing.	A query expressed in SQL is decomposed into BLR queries against local schemata. A program called Y-valve determines if the database is local or remote and routes it through the network if remote. In the case of InterBase or Digital Equipment Corp.'s Rdb, the BLR is used directly against the access method. In the case of Oracle, it is restored to SQL through the gateway and passed to the Oracle access method.	Query is parsed and validated against the global schema. Optimal sequence of steps is determined to restrict local queries, whenever possible. Ability for user to provide rules to override normal optimization on even interactively influence sequence. Each LQP handles requests for data from its database, as decomposed from the user's request by the Global Query Processor (GQP).	A query is performed by point-and clicking at objects presented as a logical view tailored for the specific end user. End user menus, browse lists, and computed fields are presented in a similar manner. SQL queries are automatically generated by the DIS tool set. Syntactical differences between the various flavors of SQL are handled transparently by the gateway server.

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Global Schema Construction	Global Schema definition is supported for object-oriented data storage and retrieval. Schema defined when specifying the interface definition for an object type.	Not maintained beyond those in the Database Models Knowledge Base.	Data Handler (patent pending) canonicalizes all data into a form that facilitates independence of user interface, functionality and data sources.	Defines entities to Global Directory. These are validated at definition time against the local schema of the target DBMS.
Query Processing (Retrieval)	Query processing (QP) across distributed object-oriented data storage is supported; QP over object-oriented replicated data is under development.	High level query is broken into a series of low level queries and function calls by the query processor (a constraint satisfaction knowledge based system to generate a plan to resolve the high level query. Knowledge from the Database Knowledge Base and the Functions Knowledge base is used to retrieve and alter data). Next, each query is called by the Query Language Generator (QLG). The QLG may be spawned to query databases simultaneously to enhance efficiency.	Query generator receives request for data, generates query, and routes to appropriate data servers and remote systems.	A global task running in the global layer spawns local tasks running in the local layers of each Supra involved in the query. Local tasks are subdivided to local database management systems by the HDMP.

57

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Global Schema Construction	There is no global schema per se. Multiple relational schemata may be referenced within the same program. Table names must be qualified in DML statements by a logical database name.	The DHDBMS would handle this area.	CDM*plus dictionary/directory uses the ANSI/SPARC 3 schema architecture to provide global schema mappings. This allows full access to all enterprise data descriptions and acomodates mapppings between schema.	Many global schemata may exist for different applications defined across various local databases. Definition done with an off-line tool. Local schemata are loaded. Each field is labelled with a domain (semantic labelling). Database administrator is assisted in specifying replicated and fragmented relations and in resolving different representations for data elements of the same domain.
Query Processing (Retrieval)	Joins that involve tables from different databases are resolved by simple nested-loop retrieval. Access to relational systems is performed by generating suitable DML and using the most efficient interface mechanism available (e.g. Rdb/VMS DSRI, DG/SQL HLI). Research on a distributed query engine is in progress.	Queries would be written in a combination of C and Shell Scripts. The query would call *gateway and the script would extract the specific data requested from the remote connection.	CDM*plus relational operations (query, update, insert and delete) are based on relational views as created in the external schema described above. The query is embedded in current applications replacing multiple queries and processing or it is created as part of new application development. All DBMS specific code is generated regardless of location and DBMS type.	Query in SQL is parsed and validated against global schema. An optimizer selects local databases needed to process a query and chooses the lowest cost strategy based on processing and network costs. Processing supports fragmented and replicated relations. The Optimizer sends commands in Distributed Intermediate Language to translate for one database. The translator does schema, data, and language translation.

53

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Global Data Model	Relational.	Relational.	Relational.	Relational-like.
Local Database Schema Conversion	Conversion is not required. RAMIS interfaces are designed to support databases 'as is'. Definitions are automatically loaded from the dictionary or catalog (e.g. DB2 catalog or IDMS - IDD) to the RAMIS data dictionary. A full screen editor provides modification functionality.	Non-Oracle Based schema is converted into a relational model via software utilities and SQL*Loader software which remaps imported data into Oracle format. Base relations and semantic information mapping is done before data are actually brought into Oracle.	None. Global schema is a set of links to local schemata. The link points to a local database table, or to individual fields (columns) in a local database table. Links observe the protections and integrity controls that local database administrators establish for their data.	ADANET defines and processes one or more physical databases at a logical level. ADANET maintains a mapping of logical identifiers or user views, with the corresponding physical database/file identifiers which comprise the logical view. Complete data transparency is provided.
Data Incompatibilities And Semantic Mismatches	Performs format, length and data type conversions where necessary. Allows for usage and actual data format types. Conversion occurs during usage, so there is no need to change existing data.	Supports different data types and units assigned to semantically equivalent fields through data definition in the Data Dictionary. Data conversion is performed as the data are physically imported.	Performs data type and error message conversions to a generic set, data dictionary conversions via a standard dictionary views interface; resolves syntactic and semantic mismatches of SQL ("Open SQL") for relational local databases, and via translation of SQL to local database language for non-relational DBMS.	There is none. ADANET acts as a translator.

	SYBASE: SYBASE	TANDEM: NonStop SQL	TRW: TDIE	XEROX AIT: MULTIBASE
Global Data Model	Relational (Entity-Relationship based).	Relational.	ECR. Also allows multiple global schemata.	Functional Data Model.
Local Database Schema Conversion	Needs conversion into relational form or information may be passed as the result of a preprogrammed event written in the local database language.	No conversion is required. NonStop SQL is designed to support a distributed database on Tandem NonStop Systems.	Transformation can either be into global model or one-to-one with the aid of a global integrating schema.	Local database schema called Local Host Schema (LHS) in any model needs to be redefined completely in the Functional Data Model to enforce uniformity. New schema is called Local Schema (LS). Local host schema remains intact.
Data Incompatibilities And Semantic Mismatches	Local database schemata converted to the relational form are placed directly in Global Schema. Functional translations must be programmed in the open server. Format conversions are handled automatically. Supports user defined data types for global manipulation language.	Data type and format are defined in local schemata which are accessible from remote systems.	Maintains information in a knowledge base that allows any level of semantic check to be applied through the use of rules or programs.	Resolved through an Integration Database Schema (called IS) for all the local schemata. Extent of coverage is not sufficient, but it can take care of differences like kilometers and miles, age in years and in qualitative form, (young, old), etc.

	DIGITAL EQUIPMENT CORP.: Rdb/VMS, DBMS	GUPTA TECHNOLOGIES, INC: SQL SYSTEM	HONEYWELL: DDTS	IBM: SAA, DB2, AND SQL/DS
Global Data Model	Network and Relational	Relational (extended)	E-C-R and Relational	Relational
Local Database Schema Conversion	The schema provided by DBMS is sufficient for conversions.	No conversion is required. Local database schema is the same as client-server schema and IBM's DB2 schema.	Local database schemata need to be converted into E-C-R model and integrated with each other manually. Using these converted and integrated schemata, the global and local data dictionaries will be populated.	Conversion to relational form is required.
Data Incompatibilities And Semantic Mismatches	Performs data conversion when possible.	Automatic Import/Export is provided for: Lotus 1-2-3, dBase 2/3, CONDOR, Enable, DIF, and ASCII files. SQLBase has automatic data conversion for "Unload/Load" to DB2.	Mappings between local database schema and the canonical schema are stored in the data dictionaries.	Supports a wide range of data types: integer, decimal, floating point, character, graphic, and date/time. A Data Extract Facility allows data to be extracted from other databases (including non-IBM: ADABAS, M204, IDMS, DATACOM/DB, and NOMAD2), and files for use by DB2 and SQL/DS.

55

	INFORMATION BUILDERS, INC.: FOCNET	INTERBASE SOFTWARE CORP.: INTERBASE	MASSACHUSETTS INSTITUTE OF TECHNOLOGY: CIS	METAPHOR DIS
Global Data Model	Hierarchical, Relational, and Network.	Relational.	Extended ER.	Relational.
Local Database Schema Conversion	Each local database has corresponding schema, called a Master File Description (MFD). FOCNET interfaces with DSRI and other systems directly. In turn, some databases, like Dbase III or Rdb, do not require a FOCUS Master File Description. In all cases, a Master File Description can be created automatically.	Needs conversion to relational form. Maps into existing data dictionary and extends it when necessary.	Entity and attribute restructuring: many-to-one and one-to-many mapping (both local to global and global to local). Automatic invocation of joins and projections, as needed.	A Workstation Tool's Data Dictionary (WTDD), stored on each database/host computer, is used by the graphical Query, Database Administration (DBA) and Reporter Tools to organize the physical database tables into logical views for end-user, developer access.
Data Incompatibilities And Semantic Mismatches	Each database has corresponding interfaces, even between SQL databases, such as DB2, Oracle, Ingres, and Sybase, Through the interface and the MFD, all data and semantic incompatibilities are reconciled.	Data format incompatibility between heterogenous systems is transparently handled by an XDR layer.	Major emphasis on supporting joins between disparate databases through synonym and translation facilities. Tools to aid in inter-database instance identification (through key semantic matching, and organizational affinity) are being developed.	Data from any respective database are translated into a common form within the Metaphor DIS environment. Syntactical differences between the various flavors of SQL are handled transparently by the gateway server.

56

	ON-LINE SOFTWARE INTERNATIONAL: RAMIS INFORMATION SYSTEM	ORACLE CORP.: ORACLE	RELATIONAL TECHNOLOGY: INGRES	SOFTWARE AG: ADABAS
Acronym Derived From	Rapid Access Management Information System	None	None	Adaptable Database Management System
Environment (Software and Hardware)	General. RAMIS runs in the following operating environments: DOS/VSE, DOS/CICS, VM/CMS, MVS/TSO, MVS-XA/TSO, MVS/BATCH, MVS/CICS, MVS-XA/CICS.	General. Oracle relational DBMS is currently supported on over 90 hardware platforms.	General. VMS and Unix platforms.	General. Production, development, and information center environments. Software AG's product line supports multiple hardware configurations and operating systems.
Local Data Models Supported	Hierarchical, Relational, and Network.	Relational.	Relational, Hierarchical, and Network.	An inverted list structure with relational characteristics. Data are stored completely separate from the relationships. Both hierarchical and network structures are supported.
Local DBMS Supported	DB2, IDMS/R, ADABAS, SQL/DS, TOTAL, IMS/DB, TERADATA, DATACOM/DB, DL-1, VSAM, ISAM, and QSAM.	DB2, SQL/DS, and DBase III.	Version 5: Rdb and VAX Record Management Service file system. Version 6: DB2, Rdb, RMS, IMS, and SQL/DS.	DL/1, VSAM, TOTAL, IMS/DB, SQL/DS, and DB2.
Global Data Manipulation Language	RAMIS non-procedural query language. An English-like syntax is also available through a complete menu-based architecture.	SQL	SQL. Also a forms interface for defining global model.	ADABAS Direct Calls, ADABAS SQL (SQL embedded in COBOL, FORTRAN, or PLI), or NATURAL (4th Generation Language).

	SYBASE: SYBASE	**TANDEM: NonStop SQL**	**TRW: TDIE**	**XEROX AIT: MULTIBASE**
Acronym Derived From	System Database	None	TRW Data Integration Engine	None
Environment (Software and Hardware)	General. VAX, SUN, Apollo, Pyramid, Stratus, IBM and various other platforms supported.	General. Designed to run on Tandem NonStop systems.	General. The TDIE currently runs on any platforms supporting C and UNIX/VMS.	General. Multibase runs on any operating system supporting ADA.
Local Data Models Supported	Hierarchical, Relational, and Network via Relational Interface.	Relational.	Hierarchical, Relational, and Network.	Hierarchical, Relational, Network, and File Systems.
Local DBMS Supported	Sybase.	NonStop SQL and Tandem NonStop Systems.	IDMS, IMS, DB2 on IBM. Oracle on VAX. Sybase on SUN.	IDS II, S2000, Oracle, and RIM.
Global Data Manipulation Language	Transact-SQL, an extension to ANSI SQL, permits transaction management.	ANSI SQL plus extensions.	None required.	DAPLEX. The local host may not support all capabilities provided by DAPLEX.

	BBN SYSTEMS & TECHNOLOGIES: CRONUS	BOOZ ALLEN: IDIMS	CAMBRIDGE TECHNOLOGY GROUP: SURROUND	CINCOM: SUPRA
Global Data Model	Object oriented.	Frame Based/Object Oriented. Supported through the Database Models Knowledge Base. The Query Language Generator uses information derived from the Knowledge Base to determine the appropriate query language and machine.	For each remote system access is made either via application on that system, database query, or direct file access. In short, any technique that does not require code to be written for remote system.	Relational
Local Database Schema Conversion	None required.	Local database schemata are converted into frame network in Database Models Knowledge Base.	None required.	Generic CREATE statement within the Global Data Mode. SUPRA offers extensions to the CREATE statement to handle semantic and conversion issues.
Data Incompatibilities And Semantic Mismatches	Data representation conversions between different architectures are supported.	Resolved through the use of functions modelled in the Functions Models Knowledge Base. The frames in the knowledge base have a representation of a function and how data are altered or manipulated by that function. Function applications, as well as database queries, are determined by the Query Processor's constraint satisfaction procedure.	Total system and data integrity guaranteed via canonicalized Data Handler (patent pending).	Handles semantics with extensions to the CREATE statement as well as supports popular data formats.

40 LAST PAGE

	COGNOS CORPORATION: POWERHOUSE	CONTROL DATA CORP: ASCENT*suite	CONTROL DATA CORP: CDM*PLUS	DATA INTEGRATION, INC.: MERMAID
Global Data Model	Relational with extensions for arrays, substructures, redefinitions and network relationships.	See above.	A 3-schema architecture with conceptual schema stored in IDEFix, and ER based modeling method.	Relational.
Local Database Schema Conversion	No local conversion supported. A proprietary schema repository (dictionary) is provided for systems which do not have schema data (such as flat or indexed files). Relational schema data are taken directly from the relational DBMS system tables and mapped into canonical form by database specific routines.	The programs written to communicate with the remote systems would have to know what the format of the data being extracted was and what the target format needed to be before it can make the conversion.	Local database schemata remain intact and are described to CDM*plus as Relational, Network or Hierarchical schema.	Structure conversion: One to Many relation mapping (global to local) with joins added, Many to One relation mapping using different projections, and One to Many and Many to One field mapping. Model conversion for network model: Relation (global) to repeating group and Joins (global) to links.
Data Incompatibilities And Semantic Mismatches	The data model describes most sequential, relative, and indexed files, as well as most relational and network databases. A large number of data types are supported providing 1:1 coverage. Conversions between different data types of the same kind are performed automatically.	See above.	Through the 3 schema definitions and inter-schema mappings, CDM*plus performs all the necessary data format and unit conversions across the schemata. CDM*plus NDML includes full referential integrity and domain constraint checking.	Deals with two types of data translation: functional type and enumerated type. Functions, written in C, deal with problems like unit and format conversions. Enumerated type translations deal with converting sets of values through an encode/decode table lookup. Translation is done on values in query and on data when a relation is first accessed.