AUGUST 1979 VOL. 3 NO. 2

a quarterly bulletin of the IEEE computer society technical committee on

Database Engineering

Contents:

a quarterly bulletin of the IEEE computer society technical committee on Database Engineering

Database Engineering Bulletin is a quarterly publication of the IEEE Computer Society Technical Committee on Database Engineering. Its scope of interest includes: data structures and models, access strategies, access control techniques, database architecture, database machines, intelligent front ends, mass storage for very large databases, distributed database systems and techniques, database software design and implementation, database utilities, database security and related areas.

Contribution to the Bulletin is hereby solicited. News items, letters, technical papers, book reviews, meeting previews, summaries, case studies, etc., should be sent to the Editor. All letters to the Editor will be considered for publication unless accompanied by a request to the contrary. Technical papers are unrefereed.

Opinions expressed in contributions are those of the individual author rather than the official position of the TC on Database Engineering, the IEEE Computer Society, or organizations with which the author may be affiliated.

Membership in Database Engineering Technical Committee is open to IEEE Computer Society members, student members, and associate members. (Application form in this issue.)

Performance Modeling of Distributed CODASYL Structures

Frank Germano, Jr.

Digital Equipment Corporation
146 Main St., ML3-2/E41, Maynard, MA 01754

University of Pennsylvania
Philadelphia, Pennsylvania 19107

Abstract

A series of analytic performance models has been developed to compare various software architectures or strategies of distributing CODASYL-like structures. Each model uses a given global schema structure and transaction mix to minimize operational costs, which include disk costs, cpu-memory costs, and communication costs. Although formulated as optimization models, only simple heuristics, implicit enumeration, and decision support facilities are used at this time to arrive at solutions. This note summarizes the nature of these models and the underlying distribution strategies.

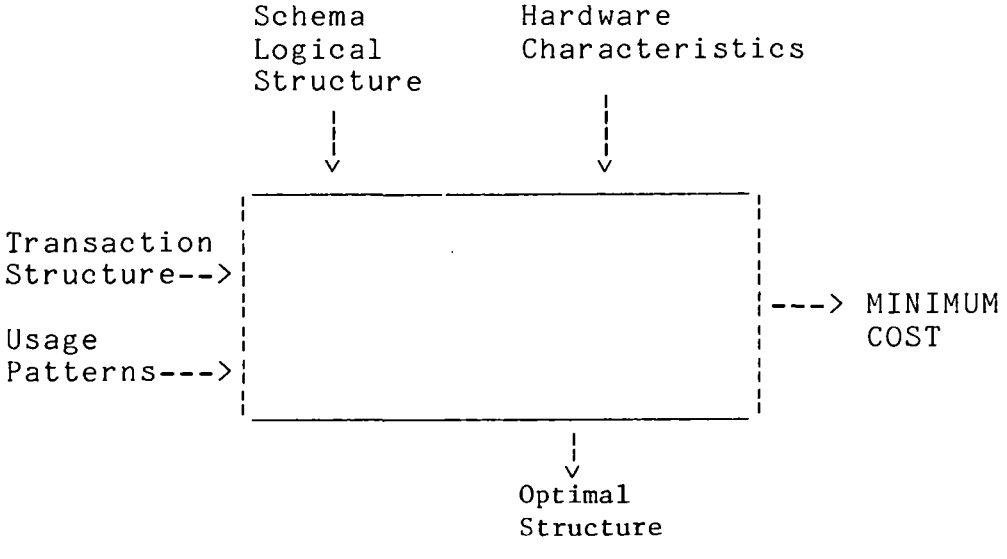
The implementor of DDBMS software can choose from a large number of possible software architectures. Although there are many important operational characteristics to consider when choosing an architecture, performance issues are a central factor. Analytic performance models are developed to assist the system implementator in this difficult comparison. Is the use of a high-level data language, parallelism, or data redundancy always indicated? What happens when communications costs are high or low? What effect does a change in the retrieval vs. update mix have on architecture choice?

Four architectures or strategies have been identified for distributing CODASYL-like structures.

- the Remote Area Model (RAM) -- use independent database management systems at each host and support global views
- the High-level Language Model (HLLM) -- use a high-level DML with remote databases and global views
- the Parallel Processing Model (HLPP) -- use a high-level DML, global views and parallel set component processing and polling.

The network environment supporting the distributed database is composed of multiple hosts, each capable of maintaining a local data structure. Underlying each architecture design is the desire to support transparent data access, i.e. the application programmer has no knowledge of data location. Each host operates components of the distributed database management system. In the remote area model this component is nothing more than a remote file system interface. In the other cases, the remote components support a three schema architecture containing user views, local structures and global (multi-host) views.

Figure 1 summarizes the general structure of each model. All are extensions of the Integrated Database Model [Gerritsen et al.].



Figure]. Model Structure

The model inputs include hardware parameters, including communications characteristics; database logical structure, represented by a global schema; and transaction information, including access path and usage patterns.

Based on this environment certain database decisions (data location and local structure decisions) are made to minimize operational costs. Operational costs include local disk storage costs, local CPU-MEMORY usage charges, and communications costs. Time performance impacts the CPU-MEMORY charges and the maximum turn-around time constraint on each transaction.

The transaction mix is represented by transaction structural characteristics and usage pattern. The structure of a transaction is represented by its access path, the series of sets accessed to get to the record once the entry record type has been established. Access to a transaction entry record is indicated to be "for all" of a given record type or "for one" identified by key equality. A transaction's usage pattern is represented by the number of times the transaction is run from each host in a given time period.

Schema logical structure includes the length of data records, their origin host(s), and record-set structure. An indication whether each record type in the transaction is added or updated is included.

Hardware parameters include the cost of disk block storage, CPU-MEMORY usage, and the cost of communications, a matrix representing the cost of sending "x" characters from host A to host B. A similar matrix representing communications time is also used.

The basic unit of distribution is the CODASYL record, except in the remote area model, where the area (file) is the unit of distribution. Data redundancy, handled as a special case of each architecture, is in terms of the unit of distribution. The unit of distribution is assigned to a host.

All models share the following local data structure decisions:

X -- record location via set of calc location

Y -- set implementation

Z -- record membership in singular set

B -- number of processing buffers at each host

P -- number of pages in each area

An additional decision, not present in the remote area model, is the record location mode binding time decision, R. The decision of where to locate record instances can be made at schema definition time or execution time. If all records of a given type are stored at a single host determined at execution time, the record location mode decision is static. If the record location mode decision depends upon execution time characteristics, the decision is dynamic. Two types of dynamic mode can be considered. The term dynamic-local will refer to the situation where all records originating at a site are stored locally at that site. The term dynamic-mobile will refer to the situation where a record's location is specified at execution time by the user or the system. Current model formulations do not support dynamic-mobile location mode because of the increased complexity.

The High-Level Data Language modeled is a series of adjacent or imbedded FOR EACH ... ENDFE constructions, a loop construction representing processing of all records of a given type in an area, calc-set or member-set [Germano]. The record selection clause of the FOREACH reduces the number of member records which must be sent back to a requestor. Processing includes the application of system functions (COUNT, MIN, MAX, SUM, AVERAGE) or user procedure to the retrieved records.

In order to capture the behavioral characteristics of a High-Level data language program the following information must be estimated for each transaction. First, for each record-type accessed along the access-path, what percentage of the data record is actually required (derived from items referenced)? Second, after the application of set-domain functions, how many characters need to be returned? And finally, what percentage of the records pass the qualification test?

Some sets defined in the global schema may have data instances at hosts different from their owner record. If a distributed set has a chain implementation mode, the set members cannot be retrieved if a single host holding member records goes down. For this reason alternative

representations of record sets must be used in a distributed environment.

A new construction is defined to support distributed sets. It is currently defined in terms of additional records which can automatically be generated by the schema definition processor. Defining additional records allows us to "piggyback" a distributed CODASYL implementation on top of existing DBMS implementations. Figure 2 summarizes this structure for distributed set j whose owner is of record type o and whose member is of type m.

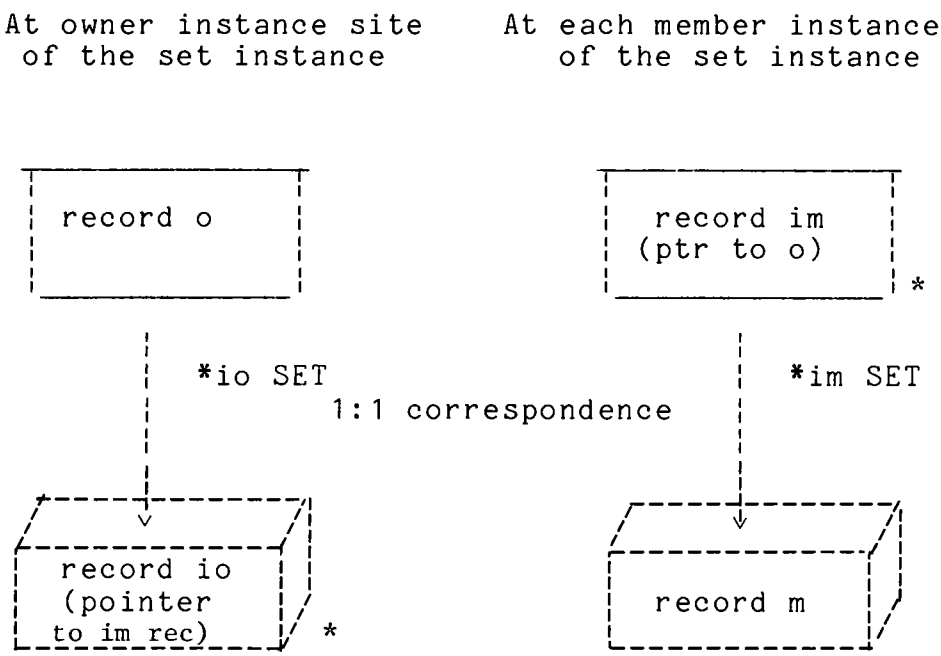


Figure 2. Implementation of Distributed Set j.

The record/set structures flagged by * represent information added to the local data structures to support distributed sets. This information is not present in the global schema. The io record holds a pointer to the remote distributed set header; the im record holds the pointer to the remote owner of the distributed set. The im record also serves to anchor the group of records of represented set j which reside at this host. This implementation of a distributed set forms the basis of a distributed directory.

The models, as currently formulated, are complex non-linear, 0-1 integer programming problems. Although sophisticated O.R. techniques may lead to efficient solution procedures, our current efforts have been directed towards using simple heuristics, implicit enumeration, and decision support facilities. We currently are using a two-stage approach: distribute the data and then optimize the local structures. Use of heuristics for some database decisions appears to be beneficial. If the models are used to support a Data Base Administrator in a distributed environment, the need for efficient solution procedures becomes more critical. Since we only wished initially to compare architectures we could be less sophisticated.

General conclusions are difficult to make, because results are dependent on how a data structure is used. Nevertheless, a few remarks are in order. It was no surprise when the remote database model dominated the remote area model for a few sample databases, but it was surprising to find domination of 10 to 100 times!

As with all modeling efforts of this nature certain objections can be raised. First, the amount of information required in large and expensive to capture. The schema information is available. Much of the transaction structure information can be derived automatically from well-designed transaction writing systems or manually as transactions are written. The run frequencies can be captured periodically in an on-going system or estimated as part of systems analysis for new systems. The communication cost function is dictated by policy, either external or internal.

Second, the mathematical structure of the models is complex. Even though we must rely on intuition and heuristics, and as the problems become more defined, better O.R. techniques, the information and understanding that we develop, can only assist the system implementor and data base administrator. They have so little to work with now.

Finally, the models have not been validated against live situations. At some point this should be attempted, not so much to verify the exact costs and times, but to hold true. Even without validation, however, the models offer a starting point for the system implementator and data base administrator. Despite the objections raised, development of the models has sharpened our insight into

some of the problems of distributing structures, and if for no other reason than this, the model development has been valuable.

REFERENCES

Chu, Wesley W., "Performance of a File Directory System for Databases," NCC 1976, pp577-587.

Germano, Frank Jr. and Meru Thakur, "SEEDFE: A FOR EACH Data Language for CODASYL Systems," Decision Sciences Working Paper 78-12-08, Dept. of Decision Sciences, University of Pennsylvania, Philadelphia, Pennsylvania.

Gerritsen, Rob, Thomas Gambino, and Frank Germano, Jr. "Cost Effective Database Design: An Integrated Database Model," Decision Science Working Paper 77-12-03, Dept. of Decision Sciences, University of Pennsylvania, Philadelphia.

Morgan, Howard Lee and K. Dan Levin, "Optimal Program and Data Locations in Computer Networks," CACM Vol 20, Number 5, May 1977, pp315-322.

REPORT ON IFIP TC-2 WORKING CONFERENCE

Ann Ellis Bandurski
David Taylor Naval R&D Center

One hundred percent of all constraints relating to a database (e.g., allowable data values, relationships, and processing) must be specified in the conceptual schema, according to Sjir Nijssen of Control Data Europe, so that no constraints remain embedded in application programs. The need for capturing a complete description of stored data, including its semantics and how it is distributed in a network environment, was a theme addressed repeatedly at the 1979 IFIP TC-2 Working Conference on Database Architecture.

Nijssen's comments were part of a report of the latest thinking of members of the IFIP Working Group 2.6 (Data Base Systems) which he chairs. The Working Group 2.6 report was presented along with 18 technical papers to approximately 70 invited discussants at this year's Working Conference in Venice, Italy June 26-29. The Working Conference was sponsored by IFIP Technical Committee 2 (Programming), 2.6's parent organization. Previous Working Conferences on Data Base Management issues have been held in 1974 in Cargese, Corsica, in 1975 in Wepion, Belgium, in 1976 in Freudenstadt, Germany, and in 1977 in Nice, France.

A three-level architecture, where a database is described in terms of conceptual, internal, and external schemas, was generally accepted as a model by attendees, and papers suggesting more complete descriptions did so within such a framework.

A Siemens group spoke on extensions to its Conceptual Schema Language (CSL) including a calendar system and semantic rule specification capabilities. The calendar system allows a fixed reference point to be defined (e.g., the birth of Christ) and is used in specifying data associations as valid within a time interval, beginning and ending with atomic events, or for specifying a temporal unit for change (e.g., weekly, monthly). A signification language within CSL is used to actually specify the semantic rules. Events can be specified and procedures defined which, when invoked, check whether actual parameters meet the requirements specified (e.g.,

formats, occurrence frequencies, consistency: valid states, persistence: valid change, existence).

Ron Stamper of the London School of Economics and Political Science suggested facilities for defining semantics within the LEGOL language for formal description of legislative rules. He further discussed the relevance of the facilities within LEGOL to the specification of semantic rules for a database through its conceptual schema. His Context Division would allow specification of jurisdiction in terms of geographical or organization hierarchies, of a hierarchy describing the problem subject area which a law addresses, and of periods of language usage so that written laws may be understood in terms of the usage of the language in which they were written at the time they were written. His second division, a Rules Division, would allow specification of "prescriptive laws" (i.e., man-made laws) as separate from "descriptive laws" (i.e., natural laws). The division covered in most detail is the third: the Surrogate Division. This division describes classes of entities which can be "things," "conditions," or "states" and which are used as surrogates in an intermediate structure between the laws which specify things about the real world and all of the facts about the real world which are relevant to those laws.

Carlo Zaniolo of Sperry Research Center, Sudbury, Massachusetts, addressed one of the issues in 3-level architecture theory: multi-model external schemas; specifically, mapping between a CODASYL conceptual schema and relational and hierarchical external schemas. He discussed both bottom-up conceptual schema design based on first defining application views (external schemas), which is possible when schemas need not be designed around an existing database, and top-down schema design based on an existing database where new external schemas are designed after the conceptual schema. To design a relational view of a CODASYL schema, Zaniolo first generates a relation from each (applicable) record type. Then, 1:n relationships, represented by sets in CODASYL, are introduced by migrating Duplicates-not-allowed data items (unique keys) or data item combinations from owners to members.

Lazlo Mercz of Control Data France also addressed the question of relational interfaces to CODASYL conceptual schemas. He spoke of the need for the definition of operators at each level of the 3-schema architecture, which would be useful at the external schema level for

transaction level. The CODASYL conceptual schema requires additional semantic definition facilities before sufficient information will be available to form relational views. For example, CODASYL provides no mechanism for specifying that two values in different record types are from the same domain and are therefore comparable. An expanded semantic specification facility is also warranted to increase system control when Manual, Optional, and Set-selection options are used. The information-bearing properties of CODASYL sets also need to be explicit.

In a discussion led by Tom Steel of AT&T, it was pointed out that some facilities which can be considered integrity constraints already exist in the CODASYL DDL, but that additional facilities are needed for this language to fulfill the requirements for conceptual schema definition. Existing CODASYL facilities include Automatic (similar to a trigger--if one action is specified, others are automatically performed also), Mandatory, Fixed, Source, and Duplicates-not-allowed data items. The question was brought up of how to accomplish meaningful updating from a relational user view and whether it is possible to let a user assume that if he can't see a portion of the database, he doesn't have to worry about it. It was also pointed out that constraints may be specified in a conceptual schema, but unless they are connected to an integrated semantic description of the database, it may become unclear why they were originally included, and other, inconsistent constraints may be specified without their relationship being apparent.

Francois Bancilhon of France's IRIA Laboratory presented his work on assuring database integrity under updates from relational external schemas. His criterion for evaluation of update acceptability was to check whether, when a user adds a tuple and then deletes it, not only the view but also the database is returned to its original state. His findings showed that, if and when additional information is (can be) provided defining the complement within a relation of the users view of it, then sufficient information is available to the system to know what should be done with the update specified by the user. Derivation of a complement depends on the operators allowed to specify a view (join and projection were examined) and on the integrity constraints involved (dependencies).

Other papers presented database system and language design (especially for distributed database systems), algorithms

for recovery, lock management, deadlock avoidance and commitment in shared database systems, and theoretical models for problem solution.

The proceedings of the Working Conference will be available to the public through North-Holland Publishing Co., Amsterdam (U.S. distribution by American Elseview, New York). The proceedings are edited by Giampio Bracchi and Sjir Nijssen and are dedicated to the memory of Mike Senko.

For additional information on the proceedings or the IFIP Working Group 2.6, I suggest contacting G.M. Nijssen, Control Data Belgium N.V., 46 Avenue des Arts, 1040 Bruxelles, Belgium or going through T.B. Steel, Jr., AT&T, 295 North Maple Avenue, Basking Ridge, N.J. 07920.

1. Name of Project: Interactive Data Base Design

2. Organization : Computer Systems Research Group
121 St. Joseph Street
University of Toronto
Toronto, Ontario
CANADA M5S 1A1

3. Personnel : Fred H. Lochovsky, Edward Chan

4. Keywords : data base design, entity-relationship model, graphics, enterprise description

5. Description : The purpose of the project is to build computer assisted tools that can be used by a database administrator in database design. The usefulness of graphics-based tools is being investigated. The entity-relationship model of Chen is being used as a basis for the enterprise description. As a first step, we are implementing a package for interactively defining and modifying a schema using the entity-relationship model. The package will run on a PDP-11/45 using UNIX and a graphics package developing an integrated design tool for generating database schemas.

6. Implications : Our eventual goal is to build integrated tools to assist in the logical database design process. Requirements collection will be done via a computerized forms system and the data will be maintained in a computerized form. This data will be used in the generation of the enterprise description using the entity-relationship model. The DBA will be able to interact with this description to modify it as needed. Finally, from this description, end-user schemas will be generated automatically according to a specified data model, e.g., relational, network, etc. The generation will be under DBA control and different designs can be

evaluated. By this process the DBA is free to concentrate on the design rather than the management of the data for the design.

7. References :

Tsichritzis, D.C., and Lochovsky, F.H., "Designing the database", Datamation 24(8), pp. 147-151, 1978.

Chen, P. P-S., "The Entity-Relationship Model: Toward a Unified View of Data", ACM TODS 1, pp. 9-36, 1976.

Chen, P. P-S., "The Entity-Relationship Model - A Basis for the Enterprise View of Data", Proc. NCC, AFIPS 46, pp. 77-84, 1977.

Chen, P. P-S., "Applications of the Entity-Relationship Model", NYU Symposium on Database Design, pp. 25-33, 1978.

1. Name of Project: System for Distributed Databases (SDD-1)

2. Organizations:

Contractor: Computer Corporation of America
575 Technology Square
Cambridge, Massachusetts 02139

Sponsor: Defense Advanced Research Projects Agency (ARPA)

3. Personnel: James B. Rothnie (Principal Investigator)
Christopher L. Reeve (Project Manager)
Timothy A. Anderson
Philip A. Bernstein
Stephen A. Fox
Nathan Goodman
Michael M. Hammer
Terry A. Landers
David W. Shipman
Eugene Wong

4. Keywords: Distributed databases, distributed processing, concurrency control, distributed query processing, reliability

5. Description:

The SDD-1 project is a three year effort to design and implement a distributed database management system. The research focus of this project has been in the areas of query processing, concurrency control and reliability in the distributed environment ([ROTHNIE et al] and [ROTHNIE and GOODMAN]). The following results have been achieved:

1. A distributed query processing algorithm has been designed and implemented. This algorithm optimizes query execution by minimizing communications traffic and exploiting parallelism ([GOODMAN et al) and [WONG]).

2. Distributed concurrency control mechanisms based on pre-analyzed transaction classed and timestamps have been developed and implemented. These mechanisms have two advantages over distributed locking: they are deadlock free and they decrease the required number of inter-site synchronization messages

([BERNSTEIN and SHIPMAN a], [BERNSTEIN and SHIPMAN b], [BERNSTEIN et al a], and [BERNSTEIN et al b]).

3. A set of reliability mechanisms have been designed and are in the process of being implemented. The major component of the reliability design is a layer of software on top of the ARPANET software called the Reliable Network (RelNet). The RelNet provides SDD-1 with a level of reliability above that provided by the ARPANET ([HAMMER and SHIPMAN]).

The initial version of SDD-1 runs on four sites on the ARPANET. The system runs on TENEX and TOPS-20 systems. The system will be completed by the end of 1979.

6. References:

[BERNSTEIN and SHIPMAN a]
Bernstein, P.A. and D. Shipman; "Concurrency Control in SDD-1: A System for Distributed Databases; Part II: Analysis of Correctness"; ACM Transactions on Database Systems, to appear.*

[BERNSTEIN and SHIPMAN b]
Bernstein, P.A. and D.W. Shipman, "A Formal Model of Concurrency Control Mechanisms for Database Systems," Proc. 1978 Berkeley Workshop on Distributed Databases and Computer Networks.

[BERNSTEIN et al. a]
Bernstein, P.A., Rothnie, J.B., Goodman, N., Papadimitriou, C.A.; "The Concurrency Control Mechanism of SDD-1: A System for Distributed Databases (The Fully Redundant Case)", IEEE Trans. on Soft. Eng., May 1978.*

[BERNSTEIN et al. b]
Bernstein, P.A.; Shipman, D.W.; Rothnie, J.B. Concurrency Control in SDD-1: A System for Distributed Databases; Part I: Description"; ACM Transactions on Database Systems, to appear.*

[GOODMAN et al]

Goodman, N., P.A. Bernstein, C.L. Reeve, J.B. Rothnie, and E. Wong, "Query Processing in SDD-1: A System for Distributed Databases", submitted for publication.*

[HAMMER and SHIPMAN]
Hammer, M.M.; and Shipman, D.W., " The Reliability Mechanisms of SDD-1: A System for Distributed Databases", submitted for publication.*

[ROTHNIE and GOODMAN]
Rothnie, J.B.; and Goodman, N. "An Overview of the Preliminary Design of SDD-1: A System for Distributed Databases", 1977 Berkeley Workshop on Distributed Data Management and Computer Networks, Lawrence Berkeley Laboratory, University of California, Berkeley California, May 1977.*

[ROTHNIE et al]
Rothnie, J.B., P.A. Bernstein, S.A. Fox, N. Goodman, M.M. Hammer, T.A. Landers, D.W. Shipman, C.L. Reeve, and E. Wong, "SDD-1: A System for Distributed Databases", ACM Transactions on Database Systems, to appear.*

[WONG]
Wong, E. "Retrieving Dispersed Data from SDD-1: A System for Distributed Databases", 1977 Berkeley Workshop on Distributed Data Management and Computer Networks, Lawrence Berkeley Laboratory, University of California, Berkeley California, May 1977.*

* These documents are also available from CCA at a price of $5 each.

1. Name of Project: Information Systems Design for Navy Logistics Systems (ISDNLS).

2. Organization : David W. Taylor Naval Ship Research and Development Center, Code 1821, Bethesda, MD 20084.

3. Personnel : David K. Jefferson

4. Keywords : Information system design, requirements analysis, database design, database conversion.

5. Description : The objective of this project is to develop a comprehensive methodology and set of computer tools for information systems design and implementation. Work to date has concentrated primarily upon the integration of PSL/PSA (ISDOS) and database design tools (March and Severance, 1978). The Interactive Database Design Methodology (IDDM) (Wilens and Volz, 1978) provides an interface between the PSA database and the database design tools. Current work involves three areas: a general simulation capability (using IDDM to access the PSA database), tools for managing software development, and tools for conversion of DML statements from one DBMS to another.

6. Implications : Results indicate that development of a tool is hard, integration of tools is harder, and development and maintenance of a methodology is hardest. Future design tools should be built within a design methodology, and should utilize common parameter databases and user interfaces.

7. References :

ISDOS Project, "PSL/PSA", Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109. (Users guides, reference manuals, etc. are all available)

March, S., and D. Severance, "A Mathematical Modeling Approach to the Automatic Selection of Database Designs," Proc. 1978 ACM SIGMOD Conference.

Wilens, M., and R. Volz, Interactive Database Design Methodology: Design Concepts Manual, Database Systems Research Group, Graduate School of Business Administration, The University of Michigan, 48109, May, 1978.

IEEE Computer Society
Technical Committee on
Database Engineering

Membership Application Form

To become a member of the TC/DBE and be on the mailing list for Database Engineering, please return this form or a copy of it to:

IEEE TC/DBE
Computer Corporation of America
575 Technology Square
Cambridge, MA 02139

NAME _____
(please print)

ORGANIZATION _____

ADDRESS _____

_____

_____

TELEPHONE _____

Meetings of Interest

August 23-24 - Dartmouth College, Hanover, N.H. Databases in the Humanities and Social Sciences. Contact: Joseph Raben, Computers and the Humanities, Queens College CUNY, Flushing, N.Y. 11367.

August 28-30 - San Francisco. Fourth Berkeley Workshop on Distributed Databases and Computer Networks. General Chairman: Dennis Hall, Lawrence Berkeley Laboratory, University of California, Number One Cyclotron Road, Berkeley, CA 94720.

Program Co-chairmen: Michael Stonebraker and Carl Sunshine. Papers were invited in the areas of protocols, routing, gateways, performance issues, concurrency control, consistency of multiple copies of data, query processing, case studies, and human factors.

October 3-5 - Rio de Janiero, Brazil. Fifth International Conference on Very Large databases. U.S. Conference Chairman: Prof. Stanley Y.W. Su, Department of Computer and Information Sciences, University of Florida, 512 Weil Hall, Gainesville, Florida 32611.

U.S. Program Chairman: Prof. Howard L. Morgan. Twenty-three sessions include 7 panel sessions, a tutorial, work in progress abstracts (2), and the database Design Workshop report.

Sessions on: user interfaces, theoretical aspects (2), applications (2), conceptual models, performance and modeling, logical design, design concepts; security, consistency, and conversion; distributed databases, query handling strategies.

Panels on: database research and directions, optical disk storage media, applications of VLDB in developing countries, database machines, natural language interfaces, multi-schema

architectures, and software engineering.

November 6-8 - Chicago. IEEE Computer Society's Third International Computer Software and Applications Conference. General Chairman: Dr. William Smith, Executive Director, Toll Electronic Switching and Operator Services Division, Bell Laboratories, Naperville, IL 60540, 312-690-2389.

Program Chairman: Prof. K.S. Fu, Purdue University. Papers were invited by June 1 in database management, data communication, transaction and information management systems as well as software development and management and application-oriented areas.

November 26-30 Los Angeles. Relational Databases Course. Contact: UCLA Extension, Short Course Program Office, P.O. Box 24902, Los Angeles, CA 90024.

Course Coordinator: Michel Melkanoff, Computer Sciences Department, UCLA. Melkanoff, Michael Blasgen, Peter de Jong, Jack Minker, and Carlo Zianolo will lecture on System R, QBE, logic and databases, the Maryland relational system with voice response, and a new design methodology for RDBs.

December 10-12 Los Angeles. International Conference on Entity-Relationship Approach to Systems Analysis and Design. Conference Chairman: Ephraim McLean, Graduate School of Management, University of California, Los Angeles, CA 90024.

Program Chairman: Prof. Peter Chen. Papers were invited by August 1 on theory and graphical representations, application to system design, application to requirement definition, and case studies and management issues.

1980 March 11-14 Pacific Grove, CA. Workshop on Computer Architecture for Non-Numeric Processing.

Chairman: Frank King, Co-chairman: Sakti Ghosh, both of IBM Research Laboratory, 5600 Cottle Road, San Jose, CA 95193.

Program Chairman: Prof. David Hsiao, Department of Computer and Information Science, Ohio State University, Columbus, Ohio 43210, (614)422-3083. Program Vice-Chairman: Glen Langdon, IBM Research Laboratory, (408)256-6454. Papers invited by 2 November 1979 especially on office automation, intelligent terminals, facsimile devices, robotic machines, text processors, database computers, and satellite business systems.

1980 October 6-9 Tokyo, Japan.

October 14-17 Melbourne, Australia. Eighth IFIP World Computer Congress. (flights arranged to second site) Program Committee Member (database and Information Systems Program Area): Prof. D.C. Tsichritzis, Computer Systems Research Group, University of Toronto, Toronto, Canada M5S 1A4. Papers invited by 1 December 1979 on information requirements specification, database systems architecture, database design, distributed databases, database and data communication systems, decision control systems, database translation, data models and data languages, and information retrieval.

CALL FOR PAPERS

The 1980 Workshop on Computer Architecture for Non-Numeric Processing

CHAIRMAN:
Frank King,
IBM Research Lab, San Jose, California, USA

CO-CHAIRMAN:
Sakti Ghosh,
IBM Research Lab, San Jose, California, USA

PROGRAM COMMITTEE CHAIRMAN:
David K. Hsiao,
Department of Computer and Information Science, The Ohio State University, Columbus, Ohio 43210, USA (614) 422-3083

PROGRAM COMMITTEE VICE CHAIRMAN:
Glen Langdon,
IBM Research Lab, 5600 Cottle Road, San Jose, California 95193, USA (408) 256-6454

PROGRAM COMMITTEE MEMBERS:
Robert Anderson,
The Rand Corp., USA
Jon Bentley,
Carnegie-Mellon University, USA
Kjell Bratbergsengen,
University of Trondheim, Norway
Helmut L. Clemm,
Siemens, West Germany
George P. Copeland,
Tektronix, USA
Harvey A. Freeman,
Sperry Univac, USA
Michael Hammer,
MIT, USA
Otto Leilich,
University of Braunchweig, West Germany
Raymond Liuzzi,
RADC, USA
Amar Mukhopadhyay,
University of Iowa, USA
Sylvian Ray,
University of Illinois, USA
David C. Roberts,
CIA, USA
Jean Rohmer,
IRIA, France
J. F. Rulifson,
Xerox, USA
Donna Williams,
Office Systems Integration, USA
Teresa T. Yang,
DEC, USA

STANDING COMMITTEE (perpetual):
P. Bruce Berra
Lee A. Hollar
Robert R. Korfhage
G. Jack Lipovski
Stewart Schuster
Stanley Y. W. Su

SPONSORED BY:
ACM SIGARCH, ACM SIGIR, ACM SIGMOD, IBM, IEEE TCARCH, IEEE TCDBE.

THIS FIFTH WORKSHOP in the series is intended to identify and encourage research and development of computer architecture for non-numeric processing. A special emphasis this year will be aimed at new work in the areas of office automation, intelligent terminals, facsimile devices, robotic machines, text-processors, database computers and satellite business systems.

Papers of up to 4,000 words each are sought. Send four copies of each paper to either Program Committee Chairman or Vice Chairman by November 2, 1979. Notification of acceptance by December 1, 1979. Final revisions of papers due February 15, 1980. All accepted papers will be included in the Workshop Proceedings and also distributed by SIGARCH, SIGIR, and SIGMOD. □

March 11-14

LOCAL ARRANGEMENTS CHAIRPERSON
Liz Hoover

Asilomar Conference Center
Pacific Grove, California

Fifth International Conference on Very Large Data Bases

Advance Program

Very Large Data Bases

WEDNESDAY OCTOBER 3, 1979

09.30 - 10.00 – SESSION 1. TUTORIAL SESSION
Co-chairmen: Antonio Furtado, PUC, Brasil and Howard L. Morgan, Univ. of PA, USA
"Locking and Recovery in a Shared Database System: An Application Programming Tutorial," C.J. Date, IBM, USA

10.30 - 12.00 SESSION 2: USER INTERFACES
Co-chairmen: S. Navathe, NYU, USA and C. Humes, Jr., IBM and Univ. of Sao Paulo, Brasil
"Spatial Management of Data," C. Herot, CCA, USA
"Towards More Informative User Interfaces," J. Janas, Inst. fur Informatik, W. Germany
"Logical Design of Deductive Natural Language Consultable Databases," V. Dahl, Univ. de Buenos Aires, Argentina
"A System for Interactive Error Detection," G. Wilson and S. Salazar, CCA, USA

10.30 - 12.00 – SESSION 3. THEORETICAL ASPECTS I
Co-chairmen: E. Neuhold, Univ. of Stuttgart, W. Germany and P. Veloso, PUC, Brasil
"Towards a Universal Relation Interface," S. Osborn, Univ. of Western Ontario, Canada
"Multivalued Dependencies with Null Values in Relational Databases," Y. Lien, Bell, Labs, USA
"Algorithmic Applications for a New Result on Multivalued Dependencies," D. Parker and C. DeLobel, UCLA, USA
"Generalized Mutual Dependencies and the Decomposition of Database Relations," A. Mendelzon and D. Maier, Princeton Univ., USA

14.00 - 15.30 – SESSION 4. PANEL - DATA BASE RESEARCH AND DIRECTIONS
Co-chairmen: D. Deutsch, NBS, USA and C. S. dos Santos, UFRGS, Brasil
Panelists: J. Lynch, Burroughs, USA, G.M. Nijssen, Control Data, Europe, J. Rothnie, CCA, USA and D. Tsichritzis, U. of Toronto, Canada

14.00 - 15.30 – SESSION 5. APPLICATIONS
Co-chairmen: G. Langdon, IBM, USA and M. Magidin, SPE, Mexico
"Functional Specification for a Decision Support System," R. Studer, Univ. Stuttgart, W. Germany
"Implementation of the IRRD Data Base in Portugal," A. Lucas, Lnec, Portugal
"A Description Language and Pilot-System Executive for Information-Transport Systems," E. Sandewall, Univ. Linkoping, Sweden

16.00 - 17.30 – SESSION 6. CONCEPTUAL MODELS
Co-chairmen: B. Yormark, Interactive Systems, USA and J.P. Schiffini, IBM, Brasil
"Software Requirement Definition and Data Models," A. Solvberg, Univ. Trondheim, Norway
"An External Schema Facility for CODASYL 1978," E. Clemons, Univ. of PA, USA
"On the Role of 'Understanding Models' in Conceptual Schema Design," J. Bubenko, Chalmers Inst., Sweden
"Statistics for the Usage of a Conceptual Date Model as a Basis for Logical Database Design," Ole Oren, Center for Industrial Research, Norway

16.00 17.30 – SESSION 7. PANEL - OPTICAL DISK STORAGE MEDIA
Co-chairmen: C. Fields, DARPA, USA and A. Akonteh, PUC, Brasil
Panelists: G. Weltman, Perceptronics, USA, other panelists to be announced

18.00 - 19.30 SESSION 8. PERFORMANCE AND MODELING
Co-chairmen: S. Kimbleton, NBS, USA and R.T. Price, UFRGS, Brasil
"An Optimization of a Recovery Mechanism for Sequential Updates," J. Arditi, Weizmann Inst., Israel
"On Search Performance for Conjuntive Queries in Compressed Fully Transposed Ordered Files," P. Svenssen, Swedish National Defense Research Inst., Sweden
"Minimizing the I/O Operations for UNDO Logging in Database Systems," A. Reuter, Tech. Hoch. Darmstadt, W. Germany

18.00 - 19.30 – SESSION 9. PANEL - APPLICATIONS OF VLDB IN DEVELOPING COUNTRIES
Co-chairmen. S. Ghosh, IBM, USA and I. Fang, ELEBRA, Brasil
Panelists: A. Cardenas, UCLA, USA, J.J. Mandiburu, Argentina, I. Mijares, ITESM, Mexico, A. Schimann, Columbia and G. Wiederhold, Stanford U., USA

THURSDAY OCTOBER 4, 1979

08.00 - 10.00 – SESSION 10. LOGICAL DESIGN
Co-chairmen: S. Berild, Univ. of Stockholm Sweden and O. Barros, Univ. de Chile, Chile
"The ADAPT System: A Generalized Approach Towards Data Conversion," M. Bach, N. Goguen, and M. Kaplan, Bell Labs. USA
"Functional Dependency Model for Logical Database Design," B. Housel, V. Waddle, and S.B. Yao, IBM, Purdue U., and NYU, USA
"Monitoring of Integrity Constraints in a CODASYL-like DBMS," R. Melo, PUC, Brasil
"A Designer for DBMS Processable Logical Database Structures," K. Irani, S. Purkayastha and T. Teorey, Univ. of Michigan, USA

08.30 - 10.00 – SESSION 11: ABSTRACTS OF WORK IN PROGRESS I
Co-chairmen: E. Nahouni, IBM, USA and J. Vidart, Univ. de Simon Bolivar, Venezuela
Participants To be announced

10.30 - 12.00 – SESSION 12: DESIGN CONCEPTS
Co-chairmen: L. Kerschberg, Bell Labs, USA and V. Setzer, Univ. de Sao Paulo, Brasil
"The Entity Join," W. Kent, IBM, USA
"Fundamental and Secondary Issues in the Design of Non-Procedural Relational Language," A. Pirotte, MBLE, Belgium
"Tools for Information System Dynamics Management," C. Rolland, S. Leifert and C. Richard, U. Paris I, France
"Interactive Specification and Formal Verification of User's Views in Data Base Design," C. Baldissera, S. Ceri, G. Pelagatti and G. Bracchi, IEE, Italy

10.30 - 12.00 – SESSION 13. DATA BASE MACHINES
Co-chairmen. G. Champine, Sperry Univac, USA and W. Padua, Fo. UFMG, Brasil
"Database Concurrent Processor," J. J. Ruz Ortiz, Univ. Complutense, Spain
Panelists: D.K. Hsiao, Ohio State Univ. USA and D. Childs, USA, others to be announced

15.00 - 16.30 –SESSION 14: PANEL - NATURAL LANGUAGE INTERFACES
Co-chairmen: A. K. Joshi, Univ. of PA, USA and F. Curado, UNICAMP, Brasil
Panelists: To be announced

15.00 - 16.30 – SESSION 15: DATA BASE SECURITY, CONSISTENCY AND CONVERSION
Co-chairmen: R.W. Taylor, IBM, USA and R. Lapyda, Univ Sao Paulo, Brasil
"Data Base Management Systems Security and INGRES," D. Downs and G. Popek, UCLA, USA
"Proving Consistency of Database Transactions," G. Gardarin and M. Melkanoff, Univ. Paris VI and UCLA, France and USA
"Database Program Conversion: A Framework for Research," R. Taylor, J. Fry, B. Shneiderman, D.C.P. Smith, and S. Su, DPCTG, CODASYL, USA

17.00 - 18.30 –SESSION 16: APPLICATIONS
Co-chairmen: A. Shoshani, LBL, USA and S. I. Roschke, SERPRO, Brasil
"A Methodologic Guideline for the Installation of Database Development and Data Administration Tools; Two Case Studies in Brasil," H. Quintella, G. de Gusmao, M. Oliveira, L.T. de Albequerque and F.R. Leao, IBICT/CNPq and SERPRO, Brasil
"A DBMS for Large Statistical Databases," M. Turner, R. Hammond and P. Cotten, Statistics Canada, Canada

17.00 - 18.30 – SESSION 17: PANEL - MULTI-SCHEMA ARCHITECTURES
Co-chairmen: E. Clemons, Univ. of PA, USA and R. Melo, PUC, Brasil
Panelists: G. Bracchi, IEE, Italy, others to be announced

20.00 BRASILIAN FOLK ART SHOW

FRIDAY OCTOBER 5, 1979

08.30 - 10.00 – SESSION 18: DATA BASE DESIGN WORKSHOP REPORT
Co-chairmen: V. Lum, IBM, USA and I. Mijares, ITESM, Mexico
"1978 New Orleans Data Base Design Workshop Report," V. Lum, S. Ghosh, D. Jefferson, S. Su, J. Fry, T. Teorey, M. Schkolnick and S.B. Yao

08.30 - 10.00 – SESSION 19: ABSTRACTS OF WORK IN PROGRESS II
Chairman: A. Morales, Univ. Catolica de Valparaiso, Chile
Participants: To be announced

10.30 - 12.00 – SESSION 20: DISTRIBUTED DATA BASES
Co-chairmen: J. Rothnie, CCA, USA and D. Menasce, PUC/RJ, Brasil
"Transaction and Catalog Management of the Distributed File Management System DISCO," H. Breitwieser and U. Kersten, Tech. Hoch. Karlsruhe, W. Germany
"On Optimistic Methods for Concurrency Control," H.T. Kung and J. Robinson, Carnegie-Mellon Univ., USA
"Decentralized Authorization in a Database System," C. Wood and E. Fernandez, IBM, USA
"Two-Phase Deadlock Detection Algorithm in Distributed Databases," S. Kawazu, S. Minami, K. Itoh and K. Teranaka, NTT, Japan

10.30 - 12.00 – SESSION 21: QUERY HANDLING STRATEGIES
Co-chairmen: R. Gerritsen, Intl. Database Systems, USA and T. Mikami, Central Res. Labs, Japan
"The Problem of Multiple Paths in a Database Schema," P. Buneman, Univ. of PA, USA
"Ambiguity in Processing Boolean Queries on TDMS Tree Structures: A Study of Four Different Philosophies," W.T. Hardgrave, Univ. of Maryland, USA
"Parallel Processing of Relations: A Single-Assignment Approach," J. Schmidt, Univ of Hamburg, W. Germany
"Query Processing in a Relational Database Management System," E. Wong and C. Yousseffi, Univ. of Calif., Berkeley, USA

13.30 - 15.00 – SESSION 22. THEORETICAL ASPECTS II
Co-chairmen: D. McLeod, USC, USA and J.L.R.H. de Araujo, UFRJ, Brasil
"Distribution Models of Relations," T. Merrett and E. Otoo, McGill Univ., Canada
"Application of Sub-Predicate Tests in Database Systems," R. Munz, H. Schneider and F. Steyer, Tech. Univ. Berlin, W. Germany
"Program Schemas for Relational Data Base Queries and Their Derivation by Transformations," V. Setzer, Univ. de Sao Paulo, Brasil
"Semantic Control of Questions Expressed in Predicate Calculus Language," R. Demolombe, Onera-Cert, France

13.30 - 15.00 – SESSION 23: PANEL - SOFTWARE ENGINEERING
Co-chairmen: R. Yeh, Univ. of Texas, USA and C. Guimares, UNICAMP, Brasil
Panelists: To be announced

Rio de Janeiro, Brazil
October 3-5, 1979

Fifth International Conference on Very Large Data Bases

Rio de Janeiro, Brazil
October 3-5, 1979

Very Large Data Bases

This fifth conference in the series, which meets at Rio de Janeiro on October 3-5, 1979, is intended to identify and encourage research, development and applications of data base systems. Its objectives are the promotion of an understanding of current research, the furthering of exchanges of information and experience gained in the design, construction and use of databases, and the provision of a forum for the discussion of future research and development. In addition, a special focus this year will be aimed at applications of data bases in developing countries and on mini- and micro-computer data base management systems.

Travel Information

Participants traveling from North America interested in group travel may choose from three attractive plans prepared by IPANEMA Tours, Inc. [Tel. (800) 252-0681 in California and (800) 421-0932 in the continental U.S.A.]
A VISA IS REQUIRED OF U.S. CITIZENS and may be required for other nationalities. Brazilian currency is the cruzeiro. US. dollars are the prefered foreign exchange currency.
A limited number of travel grants are available. High priorities will be given to those participants who contribute to the technical program. Requests should be sent to the U.S. Conference Chairman (Prof. Stanley Y.W. Su). Brazilian applicants should apply to Steering Committee Chairman (Mr. P. Schiffini).

Conference Sponsors

Association for Computing Machinery (ACM)-SIGMOD, SIGBDP, SIGIR
International Federation for Information Processing
IEEE Computer Society
SUCESU-Share Users Association, Brazil
IBM, Brazil

Accommodations

Rooms have been reserved in the conference hotel, the Hotel Nacional. It is a "five star" hotel located on the beach. In order to guarantee room reservations at the low conference rates, a one night non-refundable deposit is required and must be received before September 4. Hotel prices per day including breakfast are:
U.S. $37.00 single occupancy
U.S. $41.00 double occupancy
Participants are urged to make hotel reservations early. Those traveling under group arrangement must make the hotel arrangement with IPANEMA, Tours, Inc. The phone number for the Hotel Nacional is 5521/399-0100.

XII Brazilian National Data Processing Congress

During the week of October 8 to 12, 1979 - immediately after the Fifth International VLDB Conference - the XII Congresso Nacional de Processamento de Dados-SUCESU will be held in São Paulo, Brazil. The technical program will include about 80 presentations plus parallel activities such as equipment displays, seminars, symposia, panels, etc. A special day (Monday) is reserved in the technical program for the Data Base topics. The detailed technical program will be published one month before the Congress.

Registration

Reduced rates for pre-conference registration are available until September 1, 1979. To pre-register, fill out the Registration Form provided below and send it to the appropriate registration chairman with your remittance. Participants in Brazil should send the completed form and a check in cruzeiros payable to the Latin American Registration Chairman, in the amount equivalant to the registration fee in U.S. dollars shown below (buying rate). All others, including participants from Central America, North America, Europe, and Asia, should send the completed form and international money order in U.S. dollars to the United States Registration Chairman. This money order should be made payable to the "Fifth International Conference on Very Large Data Bases."

Brazilian Registration Chairman
Mr. Alvaro J.M. de Castro
IBM do Brasil LTDA
Caixa Postal 1830-ZC-00
20.000 - Rio de Janeiro, Brasil
Telephone 5521/243-2830

U.S. Registration Chairman
Prof. Joseph Brownsmith
Dept. of Computer & Information Sciences
University of Florida
Gainesville, FL 32611
Telephone 904/392-2371

Registration Form

The registration fee includes one copy of the Conference Proceedings, refreshments, and social events.
I cannot attend the Conference, but would like a copy of the Proceedings.
Member of Sponsoring Organization ☐ $15
Non Member ☐ $20
U.S. participants unable to send a money order or cashier's check may send a personal check (in dollars) to the U.S. Registration Chairman.

Name _____

Affilation _____

Telephone _____

Mailing Address _____

My check for registration is enclosed and
☐ includes ☐ does not include
a one night non-refundable deposit of U.S. $37.00 for single occupancy room or U.S. $41.00 for double occupancy room. (Deposit must be received by the proper registration chairman before Sept. 4, 1979.)
Please reserve these hotel accommodations for the following period.
Arrival date _____ Departure date _____

Please check the appropriate box.

	Received on or before Sept. 4	Received After Sept. 4
Member of Sponsoring Organization	☐ $75	☐ $100
Non-Member	☐ $125	☐ $150
Full-time Student	☐ $50	☐ $75

Organization _____
School _____

Conference Officers

General Conference Chairman
Mr. R.J. Libero
IBM do Brasil

Steering Committee Chairman
Mr. Jose P. Schiffini
IBM do Brasil

Latin American Conference Chairman
Prof. Valdemar W. Setzer
Instituto de Matematica e Estatistica
Universidade de São Paulo

Latin American Program Chairman
Prof. Antonio L. Furtado
Departmento de Informatica
Pontificia Universidade Catolica-RJ

U.S. Conference Chairman
Prof. Stanley Y.W. Su
Department of
Computer and Information Sciences
University of Florida

U.S. Program Chairman
Prof. Howard L. Morgan
Dept. of Decision Sciences
University of Pennsylvania

Tutorial Coordinators
Prof. Eric Clemons
University of Pennsylvania
Dr. Glenn Langdon
IBM-San Jose, U.S.A.

Latin American Publicity Chairman
Mr. Raul Isiris
SUCESU-R.J.

U.S. Publicity Chairman
Prof. Roger Elliott
University of Florida

Latin American Coordinator
Dr. Oscar Barros
Univ. de Chile

European Coordinator
Dr. Eckhard Falkenberg
Siemens AG
ZFE FL INF

Asian Coordinator
Dr. Toru Mikami
Central Research Labs

Latin American Registration Chairman and Treasurer
Mr. Alvaro J.M. De Castro
IBM do Brasil LTDA

Local Arrangement Chairman
Mr. Ronaldo C.M. Latsch
SUCESU-R.J.

U.S. Registration Chairman
Prof. Joseph Brownsmith
University of Florida

U.S. Treasurer
Mrs. Mary Lynn Slone
University of Florida

U.S. Travel Coordinator
Mr. Ernest Staveley
The Ohio State University

VLDB Foundation Representative
Prof. David K. Hsiao
The Ohio State University

Program Committee Members

Prof. Joao L.R.H. de Araujo
Universidade Federal do Rio de Janerio, Brasil

Mr. Stig Berild
University of Stockholm, Sweden

Dr. Rob Gerritsen
International Data Base Systems, Inc. U.S.A

Celio Cardoso Guimaraes
Universidade Estadual de Campinas, Brasil

Dr. Barry Housel
IBM Research, San Jose, U.S.A.

Dr. Larry Kerschberg
Bell Labs, U.S.A.

Prof. Jane Liu
University of Illinois, U.S.A.

Dr. Mario Magidin M.
Sistems y Processos Electronicos, Mexico

Prof. Ignacio Mijares
Instituto Technologico de Monterrey, Mexico

Prof. Amilcar Morales
Univ. Catolica, Valparaiso, Chile

Prof. Shamkant B. Navathe
New York University, U.S.A.

Sr. Roberto T. Price
Universidade Federal de Rio Grande do Sul, Brasil

Dr. Mario Schkolnick
IBM Research, San Jose, U.S.A.

Dr. Ari Shoshani
Lawrence Berkeley Lab., U.S.A.

Prof. Toby Teorey
University of Michigan, U.S.A.

Dr. Jorge Vidart
Universidad Simon Bolivar, Venezuela

Ms. Beatrice Yormark
Interactive Systems Corp., U.S.A.

Steering Committee Members

Prof. Bruce Berra, Chairman
ACM-SIGIR:

Dr. Vincent Lum, Chairman
IEEE-TCDBE

Mr. John Lyon, Chairman
ACM-SIGMOD

Mr. Charles C. Tucker, Chairman
ACM-SIGBDP

Prof. Herbert Weber, Chairman
VLDB-4

Prof. Raymond Yeh, Chairman
IEEE-TSE