# Table of Contents

Editor's Notes .................................................. 1
Chairman's Column ............................................. 2
SIGMOD Survey .................................................. 5
Notes from the Vice Chairperson .............................. 6
Calls for Papers ................................................... 10
Comments on "Note on the Expected Size of a Join" by T. Merrett .......................... 13
Letters .............................................................. 14
A Formal Definition of the Relational Model
by C. Date ....................................................... 18
A Precise Definition of Basic Relational Notions
and of the Relational Algebra by A. Pirotte ...................... 30
Horizontal Decomposition and Their Impact on Query Solving by P. De Bra & J. Paredaens ........ 46
Notes on the State of the Data Base Audit
by J. Clark & F. Davis ......................................... 79
On Some Arguable Claims in B. Shneiderman's Evaluation of Natural Language Interaction with Database Systems by N. Rowe ......................... 92
Response to N. Rowe's Review by B. Shneiderman ............ 98
SI~MOD RECORD
A Quarterly Publication of the Association for Computing Machinery (ACM)
Special Interest Group on Management of Data (SIGMOD)

SIGMOD Officers

Chairperson
Michael Stonebraker
M265 - Cory Hall
University of California
Berkeley, CA 94720
(415) 642 - 5799

Vice-Chairperson
Patricia G. Selinger
IBM Corp. - K55/282
5600 Cottle Road
San Jose, CA 95193
(408) 256 - 1295

Secretary-Treasurer
John L. Berg
Comp. & Info. Tech.
Standard Oil
MC: 1801
PO Box 5910-A
200 E. Randolph Drv
Chicago, IL 60680
(312) 856-4896

SIGMOD NEWSLETTER EDITOR
Tom Cook
9930 SW 61st Avenue
Portland, OR 97219
(503) 244-9729

CONTRIBUTIONS to the SIGMOD RECORD may be sent to the Editor. They should be camera-ready and typed single spaced. All papers printed herein are unrefereed working papers, unless otherwise noted.

MEMBERSHIP IN SIGMOD (at $3 per year) is open to ACM members, associate members and student members; non-ACM members may belong to SIGMOD (at $11 per year) provided their major professional allegiance is in a field other than information processing or computing. Library subscriptions are available for $9 per year.

CHANGES OF ADDRESS, SUBSCRIPTION INQUIRIES and OTHER BUSINESS INQUIRIES should be directed to ACM SIGMOD, 11 West 42nd Street, New York, NY 10036.

REGULAR BACK ISSUES of the SIGMOD RECORD (formerly PDI) are available from ACM Headquarters for $2 per copy prepaid. SIGMOD WORKSHOP PROCEEDINGS are also available for different prices.
Editor's Notes

This issue of the RECORD is a month later than intended. Due to some last minute editorial changes as well as some good old computer problems, this issue gathered a bit more dust than intended. The editorial changes involved "paging out" a few pages near the beginning of the issue. Don't let the numbers deceive you -- you are NOT missing any pages.

The technical content of this issue is, like the predecessor issues, very strong. In addition, there is a SIGMOD survey which you are all encouraged to fill out, as well as some interesting letters.

The articles by Date and Pirotte concern the fundamental definitions underlying the relational model. The De Bra and Paredaens article also deals with some theoretical aspects of the relational model. The Vander Auwera and Bernard paper is about a query language designed for bibliographic databases. The Clark and Davis paper is quite refreshing: it deals with a very important topic (database auditing) which often gets ignored in the flashy world of models and languages, and it provides an extensive bibliography on auditing. The last article and response (by Rowe and Shneiderman respectively) are actually a response and re-response to an article that appeared in a different journal, but the discussions read well alone and provide some interesting points on query languages.

LAST MINUTE NOTICE: I just got word that the Second Statistical Database Workshop will NOT be held in May (1983), but will be held in the fall. A complete announcement will be in the next RECORD.
Chairman's Column

I have just returned from a junket to the 2nd International Conference on Databases in Jerusalem. I guess that is one of the only "perqs" of academic life. Of course, I returned to a zero percent cost of living adjustment for Univ. of California professors and a University awash in cutbacks.

Israel is hot (beyond belief), looks a little like death valley and is culturally diverse (the Arabs, Jews, Armenians and Christians coexist in Jerusalem). One leaves with an appreciation of the magnitude of the Middle East political problems and an admiration for the Israelis who have turned death valley into an oasis.

The conference was noteworthy for having the mayor of Jerusalem speak at the conference reception, for having a dedicated local arrangements committee who made everything happen smoothly, and for having about 150 attendees whereby one stood some chance of meeting everybody!

The SIGMOD annual conference was a resounding success thanks largely to the efforts of Stanley Su, Mario Schkolnick and a collection of volunteers from Martin-Marietta lead by Jan Martin who handled local arrangements. About 250 attendees soaked up the Florida sunshine (and humidity). Personally, I found the paper by Raymond Lorie and Roger Haskins on extending System R to support CAD data to be the most interesting research reported.

At the business meeting the following policies were approved by the members present.

1) The SIGMOD conference should not be held outside North America in the next couple of years.

2) Outside funds should be hustled to subsidize travel of foreign authors; however, such authors should not be subsidized out of registration fees.

3) A poll of the membership should be taken to determine their sentiments on "the VLDB issue". The SIGMOD RECORD should be used to conduct this poll.

4) Registration fees should be minimized for future SIGMOD conferences. If possible, part of the SIGMOD surplus should be used to subsidize conference attendees.

5) The proceedings of the SIGMOD annual conference should be sent to all members if financially practical.

If I can get my act together by the deadline set by my taskmaster, Tom Cook, a survey form will appear in this
We also had a meeting of the SIGMOD Executive Committee. The discussion focused on the long term future of the SIGMOD annual conference and other activities which we should engage in. The consensus was that we should focus all our energy on making the conference technically excellent and on encouraging the "data base heavies" to attend. Other activities should only be undertaken if the officers have more energy than required for the above undertaking.

There was extensive discussion on whether the field was getting "stale" and if so, what to do about it. Most folks thought that the emphasis would slowly shift away from traditional topics toward a host of new ones and SIGMOD should simply "move with the times".

Humor of the month award goes to Larry Rowe for:

Q: Name two actors and a dog
A: Rin Tin Tin, Lassie and IMS

More details on our next annual conference will be forthcoming in the next issue. Meanwhile save May 23-26 and watch this space....

Also, SIGMOD will hold elections next summer. We have to recruit two people to run for each of our offices (i.e. Chairman, Vice-Chairman, and Secretary-Treasurer). The job descriptions of each position are:

Chairman: Make sure that the annual conference gets off the ground. Deal with ACM on a myriad of random issues. Absorb any flak created by VLDB or other sources. Run the annual business meeting.

Vice-Chairman: Coordinate SIGMOD participation in the ACM annual conference. Coordinate site selection for the SIGMOD annual conference and assist the chairman with "arm twisting" an organizing committee.

Secretary-Treasurer: Deal with ACM on the SIGMOD budget. Track our financial health and give a report at the annual business meeting.

We need at least two candidates for each office. Moreover, at the moment both Pat Selinger and I are not planning to run again. Our illustrious secretary-treasurer, John Berg, claims he is too busy and it may not be possible to "encourage" him to run for a SIGMOD office. Consequently, we need at least 5 and perhaps 6 volunteers to run. If you are a SIGMOD member and would like to run (or would like to volunteer one of your friends) would you drop me a note.
Survey of SIGMOD Membership

The following questions apply to the relationship between ACM and the VLDB Endowment as indicated in the last issue of the SIGMOD Record.

1) I support the ACM position concerning the VLDB Conference.
   yes ------ no ------

2) I support the VLDB Endowment position concerning the VLDB Conference.
   yes ------ no ------

3) I favor a VLDB Conference jointly co-sponsored by ACM and the VLDB Endowment.
   yes ------ no ------

4) I favor the creation of another data base conference that would attract a general audience and be of the stature of the current SIGMOD conference and VLDB conference and be sponsored by ACM or IEEE or both.
   yes ------ no ------

Please tear this page out of the newsletter and mail it to:

Michael Stonebraker
SIGMOD Survey
M265 Cory Hall
University of California
Berkeley, Ca. 94720
NOTES FROM THE VICE CHAIRPERSON

Here is a reminder of the various upcoming database and general conferences. For more details, see the ACM Communications.

CALLS FOR PAPERS


Feb. 15, 1983: Third International Conference on Entity-Relationship Approach to be held Oct. 6-8, 1983 in Anaheim, Calif.

CALENDAR OF CONFERENCES


Aug. 9-13, 1982: Second International Conference in Computer Science, Santiago, Chile.


Sept. 20-24, 1982: COMPCON Fall 82, Washington, D.C.

Oct. 8-9, 1982: ACM Mountain Region Conference, Albuquerque, N.M.

Oct. 24-26, 1982: 18th Annual Convention, Texas Association for Educational Data Systems, Austin, Tex.


Nov. 8-12, 1982: COMPSAC 82, Sixth International Computer Software and Applications Conference, Chicago, Ill.


Pat Selinger
IBM Research Lab K55/61F
5600 Cottle Rd.
San Jose, CA 95193
CALL FOR PAPERS

SECOND ACM SIGACT-SIGMOD SYMPOSIUM ON
Principles of Database Systems

Atlanta, March 21-23, 1983

The conference will cover new developments in both the practical and theoretical aspects of database systems. Papers are solicited which address either fundamental principles or important innovations in the theory, design, specification, or implementation of database systems and query languages.

Some suggested, although not exclusive, topics of interest are: concurrency control, database design, database security, data models, data structures, dependency theory, distributed database, error control, file organization, query languages, schema design, and performance evaluation.

You are invited to submit ten copies of a detailed abstract (not a complete paper) to the program chairman:

Professor Philip A. Bernstein
Aiken Computation Laboratory
33 Oxford Street
Harvard University
Cambridge, Massachusetts 02138

All submissions will be read by each member of the program committee and evaluated on the basis of significance, originality, and overall quality. Each abstract should (1) contain enough information to enable the program committee to identify the main contribution of the work; (2) explain the importance of the work — its novelty and its practical or theoretical relevance to database management; and (3) include comparisons with and references to relevant literature. Abstracts should be no longer than ten double-spaced pages.

Program Committee

P.A. Bernstein, Harvard University
C. Beeri, The Hebrew University of Jerusalem
A.K. Chandra, IBM
J. Minker, University of Maryland
C.H. Papadimitriou, M.I.T. and National Technical University of Athens
D.J. Rosenkrantz, SUNY Albany
K.C. Sevcik, University of Toronto
J.M. Smith, Computer Corporation of America
M. Yannakakis, Bell Laboratories

The deadline for submission of abstracts is October 15, 1982. Authors will be notified of acceptance or rejection by December 6, 1982. The accepted papers, typed on special forms, will be due at the above address by January 4, 1983. Proceedings will be distributed at the conference, and will be subsequently available for purchase through ACM.

General Chairman: Dr. Ronald Fagin    Local Arrangements: Professor Marc H. Graham
IBM Research    Georgia Institute of Technology
San Jose, California    Atlanta, Georgia
CALL FOR PAPERS

RESEARCH and DEVELOPMENT in INFORMATION RETRIEVAL
June 6-8, 1983 Washington, D. C.

SIXTH ANNUAL INTERNATIONAL ACM SIGIR CONFERENCE
IN COOPERATION WITH GESELLSCHAFT FÜR INFORMATIK
HOSTED BY THE NATIONAL LIBRARY OF MEDICINE

Papers are invited on theory, methodology or applications of information retrieval. Suggested areas include: Automatic Indexing, File Organization, System Evaluation, System Modeling, User Interface, Complexity Problems, Enhanced Retrieval Techniques, Natural Language Processing, Integrating IR with other Information Systems, and Applications of New Technology.

Authors should submit a full, draft manuscript (2000-5000 words) by December 1, 1982. Reports of work in progress, which are strongly encouraged, should contain outlines of sections dependent on results. Papers, as well as suggestions for panel sessions or demonstrations, should be sent to:

MATTHEW B. KOLL, SIGIR 83
DECISION SCIENCES
GEORGE MASON UNIVERSITY
4400 UNIVERSITY DRIVE
FAIRFAX, VIRGINIA 22030
(703) 323-2758

Conference Chairman:
Michael J. McGill
National Science Foundation

Program Chairman:
Matthew B. Koll
George Mason University

Local Arrangements:
Tamas E. Doszkocs
National Library of Medicine

Treasurer:
Robert E. Williamson
National Library of Medicine

Program Committee:
Peter Bollman
Technische Universitat Berlin

Christine L. Borgman
Stanford University

Duncan A. Buell
Louisiana State University

W. Bruce Croft
University of Massachusetts

Jose-Marie Griffiths
King Research, Inc.

Robert N. Oddy
Syracuse University

C.J. van Rijsbergen
University College Dublin

Jean Tague
University of Western Ontario

Clement T. Yu
University of Chicago

IMPORTANT DATES

SUBMISSION DEADLINE: DECEMBER 1, 1982
ACCEPTANCE NOTIFICATION: FEBRUARY 9, 1983
FINAL COPY DUE: MARCH 16, 1983
CONFERENCE: JUNE 6-8, 1983
Call For Papers And Workshop Information
Seventh Workshop on Computer Architecture for Non-Numeric Processing
Snowbird, Utah, March 6-9, 1983

Sponsored by: ACM SIGARCH, SIGIR, and SIGMOD
Computer Science Department, University of Utah

As the costs to design, implement, and maintain a large scale system change from where hardware costs predominate to where software costs do, it is reasonable to explore computer architectures which differ from the classic numerically-oriented machine. Front end processors, performing protocol translations, are now common in data communications systems, and a variety of backend processors for database and information retrieval applications are available or currently under development. New architectures are being developed for robotics, searching and sorting, artificial intelligence, highly available systems, workstations, and text processing.

In the past, this workshop has been a primary avenue for those engaged in research and development of a variety of specialized non-numeric systems to discuss their current activities and future directions. It has proved invaluable to students conducting research in computer architecture, allowing them to present preliminary results of their work, and receive comments and suggestions from others in the field.

Registration Information

The workshop will be held at the Snowbird Ski and Summer Resort, located in Little Cottonwood Canyon near Salt Lake City, Utah. Transportation is available from the Salt Lake airport. The registration fee of $300 includes double occupancy rooms for three nights, breakfast, and the workshop banquet. The remaining meals and any skiing expenses are the responsibility of the participants. Rooms will be available for check-in at 3:00 PM Sunday, March 6, and the first workshop session will be held at 7:30 that evening.

Because of severely limited space, attendance at the workshop will be by preregistration only. Rooms have only been reserved for the workshop participants. If you wish to bring along family members, you should contact the Workshop Chairman or Snowbird as soon as possible. For those wishing to arrive on Saturday, rather than Sunday, a limited number of rooms are available, at an additional $45.

To register for the workshop, write the Workshop General Chairman by January 1, 1983, indicating your name and affiliation, mailing address and telephone number, your interest and background in computer architecture for non-numeric processing, and whether you have submitted a paper. Include the appropriate registration fee ($300 normally, $345 for early arrival). Acceptances will be sent out on January 15, 1983, and registration checks for those we are unable to accommodate will be returned. Priority on registration will be given to those submitting a paper for presentation.

Instructions for Authors

We invite papers on current or proposed work in all areas of specialized computer architecture for non-numeric applications, such as: data communications, information management and retrieval, workstations, highly available systems, robotics, artificial intelligence, searching and sorting, and text processing. Presentations will last from 30 to 45 minutes, with half the time devoted to the presentation of the paper and half to discussion. In addition, a final session, to be organized at the workshop, will consist of a number of short presentations for those wishing to present or discuss a concept, but unable to prepare a full length paper.

Authors should to submit four copies of their paper to the Program Chairman by November 15, 1982. Papers should be approximately 5000 words in length, and include a short abstract. Consideration will also be given to extended abstracts of about 1000 words; appropriate references and figures should be included. All submissions will be acknowledged and authors will be notified of acceptance by December 31, 1983. It is anticipated that the workshop proceedings will be published as a special joint issue of the newsletters of the sponsoring SIGs.

Workshop General Chairman:
Lee A. Hollaar
Department of Computer Science
University of Utah
Salt Lake City UT 84112
(801) 581-3203

Workshop Program Chairman:
Roger L. Haskin
IBM Research Laboratories, K52-282
5600 Cottle Road
San Jose CA 95193
(408) 256-6353
Comments on "Note on the expected size of a join" by Arnon S. Rosenthal [Ro]

The usual assumption leading to the estimate $|R| \cdot |R'|/|D|$ for the size of the join of relations of $|R|$ and $|R'|$ tuples respectively is that the tuples are distributed uniformly on the $|D|$ values of the common attribute. Rosenthal shows that weaker assumptions lead to the same expected size, although it is difficult to verify whether these assumptions hold or not.

Rosenthal makes a reference to our paper [MO] which states that we make stronger assumptions, implying that our approach can be used only for a smaller set of relations. I would like to correct this impression. In fact, the results of [MO] can be used for a much larger set of relations because, while it assumes uniformity in order to get the size estimate above, it assumes uniformity only within local regions (called "cells" in [MO]) of the space of tuples for each relation. The distribution of tuples from region to region can be just as non-uniform as we are likely to find in practice. Not only do we give estimates for the size of the natural join but also for every operation of Codd's relational algebra [Co] and for our own extensions [M1, M2, etc].

Rosenthal criticizes the assumption of independence, which is shared by [Ro] and [MO]. This is probably, like uniformity, a worse defect in a model with only one cell than in a model with many. As Rosenthal points out (private communication), his results further strengthen our own method, because he shows that we did not have to assume uniformity to use our size estimate. The best method for estimating the size of a natural join is that of [MO], bolstered by the new result of [Ro].

T.H. Merrett, School of Computer Science, McGill University, Montreal

References


June 25, 1982

Mr. Tom Cook
Tektronix, Inc.
P. O. Box 500
M/S: 50-384
Beaverton, OR 97077

Re: SIGMOD and the Data Base Practitioner

Dear Tom:

During the recent SIGMOD conference in Orlando, I discussed the issue of the conference's appeal (or lack of appeal) to the data base practitioner with John Lyon, John Berg, and others. I believe the main issue is whether or not SIGMOD should be responsible to the professional interests of practitioner (in addition to the other groups it serves). If so, it is failing in this effort; if not, should an attempt be made to spawn a separate organization?

I would appreciate seeing readers' opinions on this matter in the RECORD. I for one think it vital to maintain a link between the research/development community and the practicing community in this very fast moving field in order to:

- surface important practical problems;
- relate operational results for new methods (e.g., early installation results of SQL/DS);
- introduce advanced concepts to the more progressive practitioners.

There is a lot of advanced thinking and methodology which is not finding its way rapidly enough into practice. Many of my clients, even in large "sophisticated" installations, think "normalization" just means not having repeating groups, and don't know what a domain is.
Should SIGMOD offer programs for the practitioner? If so, they need to be well conceived and directed. The data base management exhibits at the Orlando conference were a joke. Surely we can do better than that.

Sincerely,

[Signature]

Robert M. Curtice

RMC:cad
Dear Dr. Cook,

in his Chairman's Column in SIGMOD RECORD 12:3, Michael Stonebraker asked for this year's data base PhDs. Well, I've just finished my thesis "Transaction-oriented Recovery Concepts for Distributed Data Base Systems". Unfortunately, my thesis is written in german. Nevertheless, because of the following reasons, an abstract might be of interest for the SIGMOD community:
- Since the SIGMOD RECORD is read worldwide, some of its readers will be familiar with the german language.
- Other people will be interested in what is going on in the "rest of the world".
- Parts of the thesis are also available as papers or reports written in english.

Yours sincerely

Encl.

This german thesis as well as several excerpts written in english are available directly from the author: Bernd Walter, University of Stuttgart, Azenbergstr.12, D-7000 Stuttgart 1, Fed. Rep. of Germany
A part of this thesis was also presented at this year's Berkeley Workshop and is also to appear in the Computer Networks Journal.

Abstract: Recovery concepts are introduced which enable a DDBS to detect and to handle failures of distributed transactions of arbitrary complexity.
In the first part a robust and efficient protocol for checking the availability of remote sites is described. This so-called RSC-protocol maintains at each site in a network a list of all remote sites currently available; it is robust against any number of site crashes and/or communication breakdowns including network partitioning. The protocol is application independent and fits into open systems architectures. So-called 'Timed Predicate Transition Nets' (a new variant of higher level petri nets) are introduced and used to analyze the RSC-protocol. It is shown that the protocol is bounded and free of deadlocks.
The second part starts with a discussion of general mechanisms to support the robustness of DDBSs. Then it is shown how to build robust transaction processing protocols using the services supported by the RSC-protocol.
Finally a local recovery strategy is given which fits into the global concepts developed in the thesis. Various strategies for implementing stable storage in a DDBS environment are evaluated using an analytical cost model.