Lecture Notes in Computer Science 2097
Edited by G. Goos, J. Hartmanis and J. van Leeuwen

Springer
*Berlin*
*Heidelberg*
*New York*
*Barcelona*
*Hong Kong*
*London*
*Milan*
*Paris*
*Singapore*
*Tokyo*

Brian Read (Ed.)

# Advances in Databases

18th British National Conference on Databases, BNCOD 18
Chilton, UK, July 9-11, 2001
Proceedings

Springer

Series Editors

Gerhard Goos, Karlsruhe University, Germany
Juris Hartmanis, Cornell University, NY, USA
Jan van Leeuwen, Utrecht University, The Netherlands

Volume Editor

Brian Read
CLRC Rutherford Appleton Laboratory
Information Technology Department
Chilton, Didcot, Oxfordshire OX11 0QX
United Kingdom
E-mail: Brian.Read@rl.ac.uk

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Advances in databases : proceedings / 18th British National Conference on Databases, BNCOD 18, Chilton, UK, July 9 - 11, 2001. Brian Read (ed.). - Berlin ; Heidelberg ; New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ; Tokyo : Springer, 2001
(Lecture notes in computer science ; Vol. 2097)
ISBN 3-540-42265-X

CR Subject Classification (1998): H.2-4

ISSN 0302-9743
ISBN 3-540-42265-X Springer-Verlag Berlin Heidelberg New York



Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de


Printed in Germany

Typesetting: Camera-ready by author, data conversion by PTP-Berlin, Stefan Sossna
Printed on acid-free paper SPIN: 10839443 06/3142 5 4 3 2 1 0

# Foreword

The ever-expanding growth of Information Technology continues to place fresh demands on the management of data. Database researchers must respond to new challenges, particularly to the opportunities offered by the Internet for access to distributed, semi-structured and multimedia data sources.

This volume contains the proceedings of the 18th British National Conference on Databases (BNCOD 2001), held at the Rutherford Appleton Laboratory in July 2001. In recent years, interest in this conference series has extended well beyond the UK. In selecting just eleven of the submitted papers for presentation, the programme committee has included contributors from The Netherlands, Germany, Sweden, Canada and USA. In addition, two specially invited speakers address subjects of topical interest.

Our first invited speaker is Professor Dr. Rudi Studer from the University of Karlsruhe. At AIFB, the Institute for Applied Informatics and Formal Description Methods, he and his colleagues are in the forefront of work on the Semantic Web. This aims to make information accessible to human and software agents on a semantic basis. The paper discusses the role that semantic structures, based on ontologies, play in establishing communication between different agents. The AIFB web site has been developed as a semantic portal to serve as a case study.

The massive increase in data volumes from big science such as remote sensing and high energy physics means that we now contemplate the storage and processing of petabytes. Grid technology, specifically the „Data Grid“ is seen as attractive. It is thus timely that our second invited speaker addresses strategy in this field. He is Professor Tony Hey, now recently appointed as Director of the UK e-Science Core Programme and well placed to expound the vision.

The contributed papers are presented in four groups. The first of these addresses performance and optimisation. This issue has always been at the core of database technology. The first paper, by Regan and Delis, reports on a practical study of space management in logs. They evaluate a technique for reclaiming log space from short transactions while retaining recoverability for long running ones. The increasing popularity of XML presents new challenges. Zhu and Lü propose an algorithm for an effective storage placement strategy for XML documents that facilitates their efficient parallel processing. The trade off between data quality and performance is an interesting topic tackled by Caine and Embury. They study algorithms for integrity checking delayed from when the system is too busy to off-peak, „lights out“ hours.

The second group of papers concentrates on objects in databases and software engineering. The great variety of CASE tools prompt the adoption of standardised meta-models and transfer formats. In proposing an extension to OCL, Gustavsson and Lings further the interchange of models by defining a common, model independent notation for design transformations. Next, Zhang and Ritter investigate the state of database support for software development using object-oriented programming languages. They highlight the shortcomings in this respect of the current object-relational database paradigm and suggest how it might beneficially be enhanced. The third paper returns to the engineering design environment and tackles concurrent version control. Al-Khudair, Gray and Miles present a generalised object-oriented model that captures the evolution of design configurations and their components by supporting versioning at all levels.

In the third group of papers, we again consider optimisation. More specifically, contributors consider efficient querying in the newer domains of multimedia and distributed data sources. The requirements and techniques of the worlds of information retrieval and transactional databases are very different. The Dutch team of Blok, de Vries, Blanken and Apers present a case study on the „top-N" queries familiar in content retrieval in the context of a database approach to the management of multimedia data. The key issues addressed, such as speed and quality of answers and the opportunities for scalability are supported by experimental results. A similar problem is of concern to Sattler, Dunemann, Geist, Saake and Conrad. They seek control over the potentially excessive data returned from a query over heterogeneous data sources. By extensions to multi-database languages, they explore ways of asking for just the „first n" results, or of asking for a *sample* of the complete result. Still with the theme of information systems relying on database technology, Waas and Kersten are concerned with a web multimedia portal based on the Monet database system. Here the optimisation challenge is query throughput. The authors report on the performance of a simple and robust scheme for the scheduling of queries in a large, parallel, shared-nothing database cluster.

The two papers in our final group are both about querying objects. However, they are very different. Trigoni and Bierman present an inference algorithm for OQL that identifies the most general type of a query in the absence of schema type information. This is relevant to where heterogeneity is encountered – for example, in any open, distributed, or even semi-structured, database environment. Distributed databases and virtual reality are combined in the ambitious work reported by Ammoura, Zaiane and Ji. They explore data mining in a virtual data warehouse. Rendering multi-dimensional data aggregates as objects, the user flies through the data to explore and query different views.

## Acknowledgements

The members of the programme committee, under the energetic leadership of Carole Goble, reviewed the submitted papers with critical thoroughness. Their enthusiastic commitment to the continued success of BNCOD is much appreciated. Susan Hilton and her team played an essential part in the practical organisation. Thanks are also due to Alex Gray and the steering committee for the invitation to host the conference and valuable guidance.

April 2001 Brian Read

# Conference Committees

## Programme Committee

| | |
|---|---|
| Carole Goble (Chair) | University of Manchester |
| David Bell | University of Ulster |
| Richard Connor | University of Strathclyde |
| Richard Cooper | University of Glasgow |
| Suzanne Embury | University of Wales, Cardiff |
| Carole Goble | University of Manchester |
| Alex Gray | University of Wales, Cardiff |
| Mike Jackson | University of Wolverhampton |
| Anne James | Coventry University |
| Keith Jeffery | CLRC Rutherford Appleton Laboratory |
| John Keane | UMIST |
| Graham Kemp | University of Aberdeen |
| Jessie Kennedy | Napier University |
| Brian Lings | University of Exeter |
| Rob Lucas | Keylink Computers Ltd |
| Mark Levene | University College London |
| Nigel Martin | Birkbeck College, University of London |
| Werner Nutt | Heriot-Watt University |
| Norman Paton | University of Manchester |
| Alex Poulovassilis | Birkbeck College, University of London |
| Brian Read | CLRC Rutherford Appleton Laboratory |
| Nick Rossiter | University of Newcastle upon Tyne |
| Michael Worboys | Keele University |

## Additional Referees

| | |
|---|---|
| Carsten Butz | Heriot-Watt University |
| Alvaro Fernandes | University of Manchester |
| Tony Griffiths | University of Manchester |
| Paulo Pinheiro da Silva | University of Manchester |
| Sandra Sampaio | University of Manchester |

## Organising Committee

| | |
|---|---|
| Brian Read (Chair) | CLRC Rutherford Appleton Laboratory |
| Susan Hilton | CLRC Rutherford Appleton Laboratory |
| Kevin O'Neill | CLRC Rutherford Appleton Laboratory |

## Steering Committee

| | |
|---|---|
| Alex Gray (Chair) | University of Wales, Cardiff |
| Nick Fiddian | University of Wales, Cardiff |
| Carole Goble | University of Manchester |
| Peter Gray | University of Aberdeen |
| Roger Johnson | Birkbeck College, University of London |
| Jessie Kennedy | Napier University |
| Brian Lings | University of Exeter |
| Michael Worboys | Keele University |

# Table of Contents

## Invited Papers



## Performance and Optimisation



## Objects - Design & Development

## Query Optimisation



## Querying Objects