Lecture Notes in Computer Science 1859

Edited by G. Goos, J. Hartmanis, and J. van Leeuwen

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

Marcus Jürgens

Index Structures for Data Warehouses

Springer

Series Editors

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

Author

Marcus Jürgens
Freie Universität Berlin
Fachbereich für Mathematik und Informatik, Institut für Informatik
Takustraße 9, 14195 Berlin, Germany
E-mail: marcus.juergens@gmx.de

Cataloging-in-Publication Data applied for

Die Deutsche Bibliothek - CIP-Einheitsaufnahme

Jürgens, Marcus:
Index structures for data warehouses / Marcus Jürgens. - Berlin ; Heidelberg ;
New York ; Barcelona ; Hong Kong ; London ; Milan ; Paris ; Singapore ;
Tokyo : Springer, 2002
(Lecture notes in computer science ; 1859)
Zugl.: Berlin, Freie Univ., Diss., 2000
ISBN 3-540-43368-6

CR Subject Classification (1998): H.3.1, H.2.7, H.3, H.2

ISSN 0302-9743
ISBN 3-540-43368-6 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 Boller Mediendesign
Printed on acid-free paper SPIN: 10846408 06/3142 5 4 3 2 1 0

Preface

This thesis investigates which index structures support query processing in typical data warehouse environments most efficiently. Data warehouse applications differ significantly from traditional transaction-oriented operational applications. Therefore, the techniques applied in transaction-oriented systems cannot be used in the context of data warehouses and new techniques must be developed.

The thesis shows that the time complexity for the computation of *optimal* tree-based index structures prohibits its use in real world applications. Therefore, we *improve* heuristic techniques (*e. g.* R^*-tree) to process range queries on aggregated data more efficiently. Experiments show the benefits of this approach for different kinds of typical data warehouse queries. Performance models *estimate* the behavior of standard index structures and the behavior of the extended index structures. We introduce a new model that considers the distribution of data. We show experimentally that the new model is more precise than other models known from literature. Two techniques *compare* two tree-based index structures with two bitmap indexing techniques. The performance of these index structures depends on a set of different parameters. Our results show which index structure performs most efficiently depending on the parameters.

Acknowledgements

I am very grateful to have had the opportunity to write my Ph. D. Thesis under the supervision of Professor Hans-Joachim Lenz who brought the area of data warehouses to my attention. In countless meetings he gave me helpful feedback. I would like to thank Professor Heinz Schweppe for his constructive suggestions and the invitation to cooperate with the database group at the Freie Universität Berlin. Professor Freytag supported me with beneficial ideas and outstanding comments.

The graduate school in Distributed Information Systems would not be possible in this efficient form without its speaker Professor Oliver Günther. His commitment gives this school a constructive and pleasant environment.

I would like to express my thanks to all members of database groups participating in this graduate school for their interesting and encouraging talks and discussions. In particular, I am grateful for the constructive discussions with Agnès Voisard and Annika Hinze. The *Deutsche Forschungsgemeinschaft (DFG)* supported me as a fellowship recipient. Professor Joseph Bronstad and Leslie Hazelwood did not give up trying to correct my English.

Contents