

Jens Kohler

**Optimizing Query Strategies in  
Fixed Vertical Partitioned and  
Distributed Databases and their  
Application in Semantic Web  
Databases**



Bulgarian Academy of Sciences  
Institute of Information and Communication Technologies

Jens Kohler

Optimizing Query Strategies in Fixed Vertical Partitioned  
and Distributed Databases and their Application in  
Semantic Web Databases

Doctoral Thesis

Doctoral Program: Informatics

Professional Area: 4.6 Informatics and Computer Science

Supervisor: Prof. Dr. Kiril Simov

Sofia, 2017



Berichte aus der Informatik

**Jens Kohler**

**Optimizing Query Strategies in Fixed Vertical  
Partitioned and Distributed Databases and their  
Application in Semantic Web Databases**

Shaker Verlag  
Aachen 2018

**Bibliographic information published by the Deutsche Nationalbibliothek**

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Zugl.: Bulgarian Academy of Science, Diss., 2018

Copyright Shaker Verlag 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publishers.

Printed in Germany.

ISBN 978-3-8440-6097-3

ISSN 0945-0807

Shaker Verlag GmbH • P.O. BOX 101818 • D-52018 Aachen

Phone: 0049/2407/9596-0 • Telefax: 0049/2407/9596-9

Internet: [www.shaker.de](http://www.shaker.de) • e-mail: [info@shaker.de](mailto:info@shaker.de)

# Table of Contents

|                                                                                                            |           |
|------------------------------------------------------------------------------------------------------------|-----------|
| <b>List of Abbreviations</b>                                                                               | <b>1</b>  |
| <b>Introduction</b>                                                                                        | <b>3</b>  |
| Importance of the Topic . . . . .                                                                          | 4         |
| Overview of the Main Results in the Area . . . . .                                                         | 5         |
| Goals and Tasks of the Thesis . . . . .                                                                    | 7         |
| Contributions of the Thesis . . . . .                                                                      | 9         |
| Scope and Limitations . . . . .                                                                            | 10        |
| Formal Conventions . . . . .                                                                               | 11        |
| Structure of the Thesis . . . . .                                                                          | 13        |
| <b>1 Problem Definition</b>                                                                                | <b>15</b> |
| 1.1 Data Set Definition . . . . .                                                                          | 15        |
| 1.2 Data Access . . . . .                                                                                  | 16        |
| 1.2.1 Selection . . . . .                                                                                  | 17        |
| 1.2.2 Projection . . . . .                                                                                 | 18        |
| 1.2.3 Join . . . . .                                                                                       | 20        |
| 1.2.4 Naming Conventions . . . . .                                                                         | 21        |
| 1.3 Problem Formulation . . . . .                                                                          | 21        |
| 1.4 Hypotheses . . . . .                                                                                   | 25        |
| <b>2 Definition of the FVPD Methodology and its Original Implementation in the <i>SeDiCo</i> Framework</b> | <b>29</b> |
| 2.1 Fixed Vertical Partitioning and Distribution (FVPD) Definition .                                       | 29        |
| 2.1.1 Vertical Data Partitioning . . . . .                                                                 | 30        |
| 2.1.2 Correctness of FVPD methodology . . . . .                                                            | 33        |
| 2.2 Data Distribution: The <i>SeDiCo</i> Approach . . . . .                                                | 40        |
| 2.2.1 FVPD Join . . . . .                                                                                  | 43        |
| 2.2.2 Row Reconstruction in <i>SeDiCo</i> . . . . .                                                        | 44        |

|          |                                                   |           |
|----------|---------------------------------------------------|-----------|
| 2.2.3    | FVPD CRUD Operations . . . . .                    | 46        |
| <b>3</b> | <b>Related Work</b>                               | <b>49</b> |
| 3.1      | Data Security and Privacy . . . . .               | 50        |
| 3.1.1    | Privacy . . . . .                                 | 51        |
| 3.1.2    | Implications for <i>SeDiCo</i> . . . . .          | 53        |
| 3.2      | Cloud Computing . . . . .                         | 55        |
| 3.2.1    | Service Models . . . . .                          | 57        |
| 3.2.2    | Deployment Models . . . . .                       | 58        |
| 3.2.3    | Implications for <i>SeDiCo</i> . . . . .          | 59        |
| 3.3      | Object-Relational Mapping (ORM) . . . . .         | 60        |
| 3.3.1    | Impedance Mismatch . . . . .                      | 61        |
| 3.3.2    | Hibernate as an ORM Implementation . . . . .      | 63        |
| 3.3.3    | Implications for <i>SeDiCo</i> . . . . .          | 65        |
| 3.4      | Caching . . . . .                                 | 65        |
| 3.4.1    | Middle-tier Database Caching . . . . .            | 67        |
| 3.4.2    | Requirements for a Cache Implementation . . . . . | 68        |
| 3.4.3    | Cache Workflow . . . . .                          | 68        |
| 3.4.4    | Caching Schemes . . . . .                         | 69        |
| 3.4.5    | Implications for <i>SeDiCo</i> . . . . .          | 70        |
| 3.5      | Database Performance Benchmarking . . . . .       | 77        |
| 3.5.1    | Implications for <i>SeDiCo</i> . . . . .          | 77        |
| <b>4</b> | <b>Conceptualization</b>                          | <b>79</b> |
| 4.1      | Query Rewriting Approach . . . . .                | 79        |
| 4.1.1    | FVPD Join . . . . .                               | 81        |
| 4.2      | Caching Approach . . . . .                        | 83        |
| 4.2.1    | Server-Based Caching . . . . .                    | 84        |
| 4.2.2    | Local Caching . . . . .                           | 86        |
| 4.2.3    | Remote Caching . . . . .                          | 88        |
| 4.3      | SSD-Based Approach . . . . .                      | 88        |
| <b>5</b> | <b>Implementation</b>                             | <b>91</b> |
| 5.1      | Query Rewriting Implementation . . . . .          | 93        |
| 5.1.1    | FVPD Join Implementation . . . . .                | 94        |
| 5.2      | Caching Implementation . . . . .                  | 95        |
| 5.2.1    | Server-Based Caching . . . . .                    | 97        |
| 5.2.2    | Local Caching . . . . .                           | 99        |

|          |                                                          |            |
|----------|----------------------------------------------------------|------------|
| 5.2.3    | Remote Caching . . . . .                                 | 100        |
| 5.3      | SSD-Based Implementation . . . . .                       | 101        |
| <b>6</b> | <b>Evaluation</b>                                        | <b>102</b> |
| 6.1      | Evaluation Environment . . . . .                         | 102        |
| 6.2      | Basic Database Performance Evaluation . . . . .          | 106        |
| 6.2.1    | Conclusion . . . . .                                     | 107        |
| 6.3      | <i>SeDiCo</i> Framework Performance Evaluation . . . . . | 108        |
| 6.3.1    | Conclusion . . . . .                                     | 108        |
| 6.4      | Query Rewriting Evaluation . . . . .                     | 109        |
| 6.4.1    | Conclusion . . . . .                                     | 111        |
| 6.5      | Caching Evaluation . . . . .                             | 113        |
| 6.5.1    | Conclusion . . . . .                                     | 115        |
| 6.6      | SSD-based Evaluation . . . . .                           | 116        |
| 6.6.1    | Conclusion . . . . .                                     | 117        |
| <b>7</b> | <b>Summarization of the Main Results</b>                 | <b>119</b> |
| <b>8</b> | <b>Framework Application in Semantic Web Databases</b>   | <b>125</b> |
| 8.1      | Introduction . . . . .                                   | 126        |
| 8.1.1    | RDF . . . . .                                            | 129        |
| 8.1.2    | SPARQL . . . . .                                         | 131        |
| 8.2      | Problem Formulation . . . . .                            | 133        |
| 8.3      | Formal Definitions . . . . .                             | 135        |
| 8.3.1    | Open and Closed World Assumption . . . . .               | 135        |
| 8.3.2    | Correctness . . . . .                                    | 136        |
| 8.3.3    | Complexity . . . . .                                     | 142        |
| 8.4      | Related Work . . . . .                                   | 143        |
| 8.4.1    | Caching . . . . .                                        | 146        |
| 8.4.2    | Benchmarking . . . . .                                   | 146        |
| 8.5      | Approach . . . . .                                       | 147        |
| 8.6      | Implementation . . . . .                                 | 151        |
| 8.7      | Evaluation . . . . .                                     | 158        |
| 8.7.1    | Evaluation Environment . . . . .                         | 158        |
| 8.7.2    | Local SPARQL 1.0 Evaluation . . . . .                    | 159        |
| 8.7.3    | Remote SPARQL 1.0 Evaluation . . . . .                   | 160        |
| 8.7.4    | Local and Remote SPARQL 1.1 Evaluation . . . . .         | 161        |
| 8.8      | Conclusion . . . . .                                     | 162        |

|                                                               |            |
|---------------------------------------------------------------|------------|
| 8.9 Outlook and Future Work . . . . .                         | 163        |
| <b>Summary and Outlook</b>                                    | <b>166</b> |
| Summary . . . . .                                             | 166        |
| List of Publications Related to the Thesis . . . . .          | 168        |
| List of Theses Supervised by the Author . . . . .             | 176        |
| Approbation of the Results . . . . .                          | 177        |
| Key Scientific and Applied Scientific Contributions . . . . . | 179        |
| Outlook . . . . .                                             | 181        |
| <b>Declaration of Originality</b>                             | <b>186</b> |
| <b>Acknowledgments</b>                                        | <b>187</b> |
| <b>References</b>                                             | <b>188</b> |
| <b>Appendix A List of Tables</b>                              | <b>203</b> |
| <b>Appendix B List of Figures</b>                             | <b>205</b> |
| <b>Appendix C Listings</b>                                    | <b>208</b> |
| <b>Appendix D <i>SeDiCo</i> Application Screenshots</b>       | <b>209</b> |