

Position on Doctoral Thesis of Jens Kohler on

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

Research Project SeDiCo, Bachelor Theses and Publications

Jens Kohler has been working on a research Project called SeDiCo (A framework for a *SE*cure and *DI*stributed Cloud Data *StO*re) since 2012. SeDiCo has initially been a research project founded by the “Karl Steinbuch Research Program” of the Media and Film Society (MFG) in Baden-Württemberg (Germany). Jens Kohler wrote the proposal with my support. Our University of Applied Sciences in Mannheim was the only project partner.

The main objective of this research project was the design and development of a distributed framework for vertical partitioned distributed databases in order to improve security in cloud databases. The main idea is to split the columns of security relevant database tables into two or more partitions in a way that the information of each partition itself is worthless as long as the other partitions are not available. This makes it possible to store those partitions in the cloud, e.g. the critical attributes in a private cloud and the rest in a public cloud.

This project succeeded, and we could prove the feasibility of such a framework by a prototypical implementation. Nevertheless, we recognised severe performance problems when evaluating the prototype with real world data, in particular when doing queries on large amounts of data. This was the point for us to continue our research focusing on performance improvement, which forms the focus of Jens Kohler’s Doctoral Thesis.

Jens Kohler read lots of literature and many papers, and I had many interesting and long discussions with him about possible approaches for performance improvement. He did most of the conceptual work, while I felt more as a coach and supporter to him. For implementation and testing, he supervised several Bachelor Theses together with me and other colleagues.

Based on these works, we have written many common publications for national and international conferences as well as journals. Jens Kohler always was the main author and also presented the results at the conferences and put them for up discussion in order to get additional feedback. For most of the papers, I was the co-author, discussed the contents with Jens Kohler and optimized the papers. For some of the papers, also the authors of the Bachelor Theses delivered some content.

Doctoral Thesis: Research Focus

Jens Kohler’s Doctoral Thesis has a clear research focus on the query optimization in fixed vertical partitioned and distributed databases in the cloud. This topic is very up-to-date and praxis relevant, since it helps to build secure and performant databased software systems in the cloud. The used technologies like MySQL, Java, Hibernate, Eucalyptus, Cloud Stack) are state of the art and widespread.

On the other hand, the author gives a well-grounded foundation of the underlying theory, in particular about Codd’s relation model and its implications for query optimization. This even includes a correctness proof of the applied query-rewriting algorithm. The indentation of theoretical basics and practical usability is one of the strengths of this thesis.

Furthermore, he transforms the concept of query rewriting to the domain of semantic web, which gives a great outlook for further research activities.

Doctoral Thesis: Structure

In chapter 1, Jens Kohler gives a detailed introduction into the problem and formulates some hypotheses to be confirmed or contradicted later on. These hypotheses give a clear structure to the remaining chapters.

Chapter 2 lays the theoretical foundations including Codd's relation model, fixed vertical partitioning and query rewriting.

Chapter 3 discusses relevant concepts, methods and technology based on an extensive literature research. A summary with the implications for his thesis concludes each subchapter.

In chapter 4, he works out his concept for query optimization based on the theoretical foundations from chapter 2 and the relevant concepts identified in chapter 3.

Chapter 5 delivers some implementation details. Especially the UML sequence diagrams are quite helpful to understand the underlying algorithms.

Chapter 6 gives a detailed evaluation of the developed query optimization framework and the three different optimization strategies (query rewriting, caching and SSD-based).

Chapter 7 summarizes these results and their implications for the hypotheses formulated in chapter 1.

Finally, chapter 8 gives an outlook to apply the methodology and framework developed in this thesis to the application domain of semantic web.

In summary, the structure of the thesis is very clear, from the formulation of the research problem and the hypotheses over the theoretical and practical foundations via conceptualization and implementation up to the evaluation and verification of the hypotheses. The last chapter with the application to Semantic Web Databases on my opinion is more an outlook, since it leaves many questions open and a detailed investigation would exceed the scope of this thesis.

Doctoral Thesis: Content

The thesis gives an accurate definition for all used terms based on scientific literature. Scientific methods and models (e.g. Codd's relation model) are used correctly, and a proof of the developed query-rewriting algorithm is given. Tables mapping the theoretical concepts to the technical artefacts are quite helpful for understanding. The same is true for the UML sequence diagrams modelling the algorithms for caching and joining the results. The evaluation results are clearly arranged and commented.

To summarize, the thesis is very reproducible, and all thoughts are justified. Whenever a methodical or technological decision was made, the author considered several alternatives from literature as well as own ideas. He evaluated them based on requirements and feasibility and decided for the best suited.

Some minor negative aspects concerning the content shall also be mentioned. In the introduction, Jens Kohler describes the Design Science approach, but never references it again in the following chapters. In some of the UML sequence diagrams in chapter 5, Cache loading could

be done in parallel for all partitions to further improve performance. Furthermore, in figure 5.2 alternative and parallel branches are missing.

Doctoral Thesis: Literature Selection and Analysis

Jens Kohler did a very thoroughly literature research considering scientific foundations, methods and solutions as well as technologies. The referenced literature is of high quality. This includes scientific papers as well as up-to-date online sources

Whenever possible, he cites literature to confirm his statements, and includes these citations well into the text. His citation style is always correct and consistent. He also references many papers we have written together, with him as the main author, in particular on the basic Fixed Vertical Partitioning and Distribution framework.

Doctoral Thesis: Formal Aspects

Jens Kohler worked very accurate when writing his doctoral thesis. The layout is perfect, and figures help understanding difficult content. Although the thesis is written in English, linguistic quality is very high, and there are very few orthography and grammar mistakes. A detailed list of all used abbreviations is given as well as lists of tables, figures and listings.

Nevertheless, some of the figures are a bit blurred, and sometimes an additional figure could have improved understanding.

Summary

Summing up, Jens Kohler has delivered a very good doctoral thesis, which is scientifically founded as well as highly praxis-relevant. He has worked very thoroughly, and I had many interesting scientific discussions with him. Whenever he focused a problem, he developed innovative ideas to solve them and often discussed them with me.

Several bachelor students have written their thesis under our supervision and learned a lot on the project. Furthermore, we produced many papers on the subject and got a lot of constructive feedback. Jens Kohler always took this feedback as inspiration for new ideas to move on the SeDiCo project as well as his thesis, and not as criticism.

He has worked very self-reliant and motivated. I had more the role of a coach than a supervisor on him. The title and subject of this doctoral thesis were also his own idea.

All in all, Jens Kohler has delivered a very good doctoral thesis, with some minor deficiencies.

Prof. Dr. Thomas Specht
University of Applied Sciences
Computer Science Department
Paul-Wittsack-Str. 10
68163 Mannheim
Germany