

# Report on the Second International Workshop on Self-Managing Database Systems (SMDB 2007)

Anastassia Ailamaki  
Carnegie Mellon University  
natassa@cmu.edu

Surajit Chaudhuri  
Microsoft Research  
surajitc@microsoft.com

Sam Lightstone  
IBM Software Development  
Laboratory  
light@ca.ibm.com

Guy Lohman  
IBM Almaden Research Center  
lohman@almaden.ibm.com

Pat Martin  
Queen's University  
martin@cs.queensu.ca

Ken Salem  
University of Waterloo  
kmsalem@uwaterloo.ca

Gerhard Weikum  
Max-Planck Institut fr Informatik  
weikum@mpi-sb.mpg.de

## 1 Introduction

Information management systems are growing rapidly in scale and complexity, while skilled database administrators are becoming rarer and more expensive. Increasingly, the total cost of ownership of information management systems is dominated by the cost of people, rather than hardware or software costs. This economic dynamic dictates that information systems of the future be more automated and simpler to use, with most administration tasks transparent to the user.

Autonomic, or self-managing, systems are a promising approach to achieving the goal of systems that are increasingly automated and easier to use. The aim of the workshop was to provide a forum for researchers from both industry and academia to present and discuss ideas related to self-managing database systems.

SMDB 2007 was the first event organized by the new **IEEE Computer Society Data Engineering Workgroup on Self-Managing Database Systems** (<http://db.uwaterloo.ca/tcde-smdb/>). The workgroup, which was founded in October 2005, is intended to foster research aimed at enabling information management systems to manage themselves seamlessly, thereby reducing the cost of deployment and administration.

## 2 Workshop Overview

The workshop was held in Istanbul, Turkey on Monday April 16, 2007 prior to the start of the International Conference on Data Engineering. The workshop's program committee consisted of the members of the SMDB workgroup executive committee plus four other well-known researchers in the area. SMDB 2007 received 19 submissions and each paper was reviewed by 3 program committee members. 11 papers were accepted to the workshop, resulting in an acceptance rate of 58%. In an effort to make the workshop as inclusive as possible 4 submissions were accepted as poster papers and were given a shorter presentation time in the workshop. The average attendance at the workshop during the day was 32 participants.

---

*Copyright 2007 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE.*

**Bulletin of the IEEE Computer Society Technical Committee on Data Engineering**

---

### 3 Technical Program

The technical program was organized into four sessions: principles and overviews; self-healing; self-optimizing and poster papers. Links to the talks and papers can be found at the workshop Web page (<http://db.uwaterloo.ca/tcde-smdb/SMDB2007program.html>).

The first session on principles and overviews included three research papers highlighting key research issues in self-managing database systems. Soror et al. [1] considered the impact that the trend to virtualization will have on tuning database management systems (DBMSs). They presented a formalization of the resource allocation problem in this environment and discussed an approach to cost modeling that employs a query optimizer with a "what-if" mode. Lightstone et al. [2] argued for a greater use of control theory in self-managing database systems and presented examples in IBM's DB2 where it was used effectively to control utility throttling and self-managing memory. They found the main advantages of control theory are its stability and its ability to handle noise. Chen et al. [3] observed that the currently available coarse-grained resource provisioning solutions do not necessarily make effective use of the available resources. They proposed a fine-grained approach that uses outlier detection to pinpoint sources of overload problems and then migrates these queries. The fourth paper in the session, by Bowman et al. [4], provided an overview of the self-management features of SQL-Anywhere from Sybase iAnywhere. SQL-Anywhere is designed to be deployed as an embedded DBMS within zero-administration environments.

The second session, which was on self-healing, included two papers. Cook et al. [5] defined the general problem of supporting self-healing in database-centric multi-tier services and outlined a research agenda for solving the problem. They specifically identified performance availability problems as reasonable targets for self-healing and supported an approach based on robust learning algorithms. Modani et al. [6] described an approach to automated diagnosis in which symptoms of a new problem are matched to a database of symptoms of previously diagnosed problems. They exploited the fact that function call stacks can serve as symptoms of a class of problems and proposed algorithms for effectively matching call stack patterns.

The third session of the workshop included four papers on self-tuning. Papadomanolakis and Ailamaki [7] observed that existing index selection tools rely on heuristics that are hard to analyze. They proposed a model for index selection based on integer linear programming that offers higher solution quality, efficiency and scalability without sacrificing any of the precision offered by existing index selection tools. Sattler et al. [8] examined a different aspect of the index selection problem. They proposed an approach that continuously collects statistics for recommended indexes and performs on-the-fly index generation during query processing using new query plan operators IndexBuildScan and SwitchPlan. Schnaitter et al. [9] also considered an aspect of automatic index selection, namely the selection of indexes as the workload on a DBMS shifts characteristics. They described COLT, which is a novel framework that continuously monitors the workload of a database system and enriches the existing physical design with a set of effective indices. Qin et al. [10] looked to improve the cost models employed by query optimizers by improving the accuracies of I/O cost estimates of database access methods. They presented an adaptive black box statistical cost estimation method.

The fourth session of the workshop included four poster papers on a variety of topics in the area of self-managing database systems. The poster papers were each given a shorter presentation time than the regular papers. Niu et al. [11] described an approach to automatically adapting DBMS workloads such that service level objectives of the various applications are met. Duchateau et al. [12] presented an automatic schema matching approach. They specifically focused on a B-tree index structure to improve the performance of the matching algorithm. Teisanu et al. [13] outlined the problem of designing effective workload management and provided a formal definition that supports the further development of algorithms and architectures for effective on-line database tuning strategies. Lang et al. [14] described a caching algorithm for scans on buffer pools that keeps track of ongoing scans and the state of each scan. They showed that this approach could achieve improved buffer pool hit rates.

### 4 Summary

The Second Workshop on Self-Managing Database Systems was very successful. The high quality of the papers and the discussions generated during the workshop are strong indicators of the vitality and growing importance of the area of self-managing information management systems.

The Workgroup on Self-Managing Database Systems looks forward to organizing the third edition of the workshop along with ICDE 2008 in Cancun. They seek to encourage a wider range of submissions and a broader participation by academics and industrial partners in the area.

## References

- [1] A. Soror, A. Abounnaga, and K. Salem, "Database virtualization: A new frontier for database tuning and physical design," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 388 – 393, Istanbul Turkey, 2007.
- [2] S. Lightstone, M. Surendra, Y. Diao, S. Parekh, J. Hellerstein, K. Rose, A. Storm, and C. Garcia-Arellano, "Control theory: a foundational technique for self managing databases," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 394 – 403, Istanbul Turkey, 2007.
- [3] J. Chen, G. Soundararajan, M. Mihailescu, and C. Amza, "Outlier detection for fine-grained load balancing in database clusters," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 404 – 413, Istanbul Turkey, 2007.
- [4] I. Bowman, P. Bumbulis, D. Farrar, A. Goel, B. Lucier, A. Nica, G. Paulley, J. Smirnos, and M. Young-Lai, "Sql anywhere: A holistic approach to database self-management," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 414 – 423, Istanbul Turkey, 2007.
- [5] B. Cook, S. Babu, G. Candea, and S. Duan, "Toward self-healing multitier services," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 424 – 432, Istanbul Turkey, 2007.
- [6] N. Modani, R. Gupta, G. Lohman, T. Syeda-Mahmood, and L. Mignet, "Automatically identifying known software problems," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 433 – 441, Istanbul Turkey, 2007.
- [7] S. Papadomanolakis and A. Ailamaki, "An integer linear programming approach to database design," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 442 – 449, Istanbul Turkey, 2007.
- [8] K.-U. Sattler, M. Luehring, K. Schmidt, and E. Schallehn, "Autonomous management of soft indexes," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 450 – 458, Istanbul Turkey, 2007.
- [9] K. Schnaitter, S. Abiteboul, T. Milo, and N. Polyzotis, "On-line index selection for shifting workloads," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 459 – 468, Istanbul Turkey, 2007.
- [10] Y. Qin, K. Salem, and A. Goel, "Towards adaptive costing of database access methods," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 469 – 477, Istanbul Turkey, 2007.
- [11] B. Niu, P. Martin, W. Powley, P. Bird, and R. Horman, "Poster session: Adapting mixed workloads to meet slos in autonomic dbms," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 478 – 484, Istanbul Turkey, 2007.
- [12] F. Duchateau, Z. Bellahsene, M. Roantree, and M. Roche, "Poster session: An indexing structure for automatic schema matching," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 485 – 491, Istanbul Turkey, 2007.
- [13] A. Teisanu, M. Consens, M. Kandil, S. Lightstone, D. Zilio, and C. Zuzarte, "Poster session: Problem definition for effective workload management," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 492 – 497, Istanbul Turkey, 2007.
- [14] C. Lang, B. Bhattacharjee, T. Malkemus, and I. Stanoi, "Poster session: Improved buffer size adaptation through cache/controller coupling," in *Proceedings of ICDE Workshops (SMDB 2007)*, pp. 498 – 506, Istanbul Turkey, 2007.