

DIAS: Data-Intensive Applications and Systems
Laboratory
School of Computer and Communication Sciences
Ecole Polytechnique Federale de Lausanne
Building BC, Station 14
CH-1015 Lausanne
URL: <http://dias.epfl.ch>



Automated Physical Design for Scientific Workloads

(Fall Semester Sep-Dec 2009)

Keywords: Physical Design, Scientific Data Management, Index Interaction, Continuous Tuning

Problem: Database Management Systems (DBMSs) have been widely deployed the last years and the more complicated the database applications become the more significant the physical design is. Selecting indexes, materialized views, horizontal and vertical partitioning that would optimize performance for a particular workload while taking into consideration limited resources is a challenging task. Due to the dynamic and complex nature of this problem, automating the design of a database system towards building self-tuning systems has gained importance. Moreover, scientific fields such as astronomy and biology create huge amount of data that have to be organized and analyzed. The volume of these data increases every year and as a result more sophisticated techniques to manage data are becoming more and more demanding.

Project: Researchers in the DIAS lab, have developed a tool for automating indexing, called **PARINDA**. The tool allows to build appropriate schema and indexes for an input database and queries in an interactive way. The aim of this project is to extend PARINDA with the following objectives:

1. Extend the tool to incorporate on continuous tuning [1], and index interaction [2]
2. If time permits, enrich the schema design with horizontal partitioning [3]
3. Evaluate performance and efficiency of the tool using scientific workloads from the Sloan Digital Sky Survey (4), the materials sciences and the brain-Mind Institute.

Supervisor: Prof. Anastasia Ailamaki

References

1. K. Schnaitter, S. Abiteboul, T. Milo, and N. Polyzotis. Colt: continuous on-line tuning. In SIGMOD '06: Proceedings of the 2006 ACM SIGMOD international conference on Management of data, pages 793–795, 2006.
2. Karl Schnaitter, Neoklis Polyzotis, Lise Getoor: Index Interactions in Physical Design Tuning: Modeling, Analysis, and Applications. PVLDB 2(1): 1234-1245 (2009).

3. Stratos Papadomanolakis, Anastassia Ailamaki, "AutoPart: Automating Schema Design for Large Scientific Databases Using Data Partitioning," *ssdbm*, pp.383, 16th International Conference on Scientific and Statistical Database Management (SSDBM'04), 2004.
4. <http://www.sdss.org/>