

Performance Evaluation and Experimental Assessment — Conscience or Curse of Database Research?

Ioana Manolescu
INRIA Futurs, France
ioana.manolescu@inria.fr

Stefan Manegold
CWI, Netherlands
manegold@cwi.nl

Outline

Performance, performance and performance used to be the three things that really mattered in database research. Most of our published works indeed include an experimental evaluation of the proposed techniques. However, such evaluations are sometimes seen as a "must-have" eating up the valuable space where one could describe new ideas. The experimental evaluations end up being short, lacking important information to interpret and/or reproduce the results, and often end without clear conclusion.

Potential issues of controversy range from the general question, whether experimental assessment and performance evaluation are considered part of research or rather part of engineering, to practical problems such as how experiments should be designed, conducted, documented, and interpreted to make the results reproducible and meaningful for others.

The distinguished panelists will discuss their viewpoints on the following questions:

- What are the requirements for a credible experimental assessment ?
- When reviewing a submitted work, how much should we accept "based on trust" and what would be the "minimal proof requirements" ?
- Are current experimental benchmarks up to the task? Are they sufficiently used?
- Should we modify the reviewing process for big conferences to:
 - raise the requirements for experimental studies in "normal" papers?
 - solicit more disclosure of data and code?
 - highlight that a really deep experimental study actually makes a contribution by itself? and if so, how?
- Would a requirements list or even template for an "Experimental Assessment" section help to ensure standardized and complete presentation of both experimental results and the

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, to post on servers or to redistribute to lists, requires a fee and/or special permission from the publisher, ACM.

VLDB '07, September 23-28, 2007, Vienna, Austria.
Copyright 2007 VLDB Endowment, ACM 978-1-59593-649-3/07/09.

necessary information to reproduce them? Or would such an attempt of standardization be "too constraining"?

- How should we consider the relationship between commercial systems and non-commercial research prototypes, when assessing the performance of research prototypes? Is the comparison possible, given the large difference in the amount of man-month invested on the two sides ? Is the comparison meaningful, given all the constraints under which commercial systems operate (multi-user, heavy-weight, real-life application constraints...)?
- Are there classes of papers that do not need experimental validation?
- Are there other metrics than performance that could/should be assessed empirically? If so which (e.g., usability/applicability, expressiveness), and how?
- EXPDB started in 2006 [1] (and continues in 2007 [2]) as a satellite SIGMOD workshop series focused on performance evaluation in database research. Should such issues be discussed in a separate forum, or should we "bring them to the mainstream" ?

Participants

Torsten Grust (Technical U. Munich, Germany) Torsten Grust is professor for Database Systems at the Technische Universitaet Muenchen, Munich, Germany. His research interests range from database-supported XML processing, database query languages and query optimization to functional programming languages and compiler construction.

Martin Kersten (CWI, The Netherlands) Martin Kersten is head of the department of Information Systems at CWI, the Dutch national research institute for mathematics and computer science in Amsterdam, and full professor in Multimedia Databases at the University of Amsterdam. His research interests include database architectures, data mining techniques, multimedia databases as well as design and development database systems, both research prototypes and open source systems.

Paul Larson (Microsoft, USA) Paul Larson is a senior researcher in the Database Group at Microsoft Research, part of Microsoft. His research is in the area of databases, in particular mid-tier database caching, query processing and query optimization.

Guido Moerkotte (U. Mannheim, Germany) After positions at the University of Karlsruhe and the RWTH Aachen, Guido Moerkotte became a full professor at the University of Mannheim.

His main interests are query optimization and processing, and XML databases.

Yannis Papakonstantinou (UCSD, USA) Yannis Papakonstantinou is an Associate Professor of Computer Science and Engineering at the University of California, San Diego. His research is in the intersection of database and Internet technologies. A recipient of the 1998 NSF CAREER award, in 2000 Yannis founded Enosys Software, which built the first generally available distributed XQuery processor.

1. REFERENCES

- [1] EXPDB 2006 home page. Available at <http://gemo.futurs.inria.fr/events/EXPDB2006>.
- [2] EXPDB 2007 home page. Available at <http://homepages.cwi.nl/~manegold/ExpDB2007>.