



Chapter III

Object-Oriented Database Benchmarks

Jérôme Darmont
Université Lumière, France

Michel Schneider
Université Blaise Pascal, France

We present in this chapter an overview of the benchmarks aimed at evaluating the performances of Object-Oriented Databases (OODBs). We particularly focus on the Object-Clustering Benchmark (OCB), which is both generic and originally designed to specifically evaluate the performances of clustering policies within OODBs. OCB is generic because its sample database may be customized to fit any of the databases introduced by the main previous benchmarks, e.g., Object Operation 1 (OO1) or OO7. The first version of OCB was purposely clustering-oriented due to a clustering-oriented workload, but OCB has been thoroughly extended to be able to suit other purposes. Eventually, OCB's code is compact and easily portable. OCB has been validated through two implementations: one within the O₂ OODB and another one within the Texas persistent object store. The performances of two different clustering policies implemented in Texas, Dynamic, Statistical, Tunable Clustering (DSTC) and Detection & Reclustering of Objects (DRO), have also been compared with OCB.

The need to evaluate the performances of Object-Oriented Database Management Systems (OODBMSs) is important both to designers and users. Performance evaluation is useful to designers to determine elements of architecture, choose between caching strategies, and select Object Identifier (OID) type, among others. It helps them validate or refute hypotheses regarding the actual behavior of an OODBMS. Thus, performance evaluation is an essential component in the development process of efficient and well-designed object stores. Users may also employ performance evaluation, either to compare the efficiency of different technologies before selecting an OODBMS or to tune a system.

The main work presented in this chapter was initially motivated by the evaluation of object clustering techniques. The benefits induced by such techniques on global performances are widely acknowledged and numerous clustering strategies have been proposed.

As far as we know, there is no generic approach allowing for their comparison. This problem is interesting for both designers (to set up the corresponding functionalities in the system kernel) and users (for performance tuning).

There are different approaches used to evaluate the performances of a given system: experimentation, simulation, and mathematical analysis. We focus only on the first two approaches. Mathematical analysis is not considered because it invariably uses strong simplification hypotheses (Benzaken, 1990; Gardarin, Gruser, & Tang, 1995) and its results may well differ from reality.

Experimentation on the real system is the most natural approach and *a priori* the simplest to complete. However, the studied system must have been acquired, installed, and have a real database implanted in it. This database must also be significant of future exploitation of the system. Total investment and exploitation costs may be quite high, which can be drawbacks when selecting a product.

Simulation is casually used in substitution or as a complement to experimentation. It does not necessitate installing nor acquiring the real system. It can even be performed on a system still in development (*a priori* evaluation). The execution of a simulation program is generally much faster than experimentation. Investment and exploitation costs are very low. However, this approach necessitates the design of a functioning model for the studied system. The reliability of results directly depends on the quality and the validity of this model. Thus, the main difficulty is to elaborate and validate the model. A modeling methodology can help and secure these tasks.

Experimentation and simulation both necessitate a workload model (database and operations to run on this database) and a set of performance metrics. These elements are traditionally provided by benchmarks. Though interest for benchmarks is well recognized for experimentation, simulation approaches usually use workloads that are dedicated to a given study, rather than workloads suited to performance comparisons. We believe that benchmarking techniques can also be useful in simulation. Benchmarking can help validate a simulation model as compared to experimental results or support a mixed approach in which some performance criteria necessitating precision are measured by experimentation and other criteria that does not necessitate precision are evaluated by simulation.

There is no standard benchmark for OODBs, even if the more popular benchmarks, OO1, HyperModel, and OO7 are *de facto* standards. These benchmarks are aimed at engineering applications (Computer-Aided Design, Manufacturing, or Software Engineering). These general purpose benchmarks feature quite simple databases and are not well suited to the study of clustering, which is very data dependent. Many benchmarks have been developed to study particular domains. A fair number of them are more or less based on OO1, HyperModel, or OO7.

In order to evaluate the performances of clustering algorithms within OODBs, a new benchmark was designed: OCB (Darmont, Petit, & Schneider, 1998). It originally had a generic object base and was clustering oriented through its workload. It actually appeared afterwards that OCB could become more general, provided the focused workload was extended, as described in this chapter.

Full specifications for a new version of OCB are provided in this chapter. More precisely, we address the following points: the generalization of the OCB workload so that the benchmark becomes fully generic, a comparison of OCB to the main existing benchmarks, and a full set of experiments performed to definitely validate OCB. These performance tests were performed on the O₂ OODB (Deux, 1991), the Texas persistent object store (Singhal, Kakkad, & Wilson, 1992), and the DSTC (Bullat & Schneider, 1996) and the DRO

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/object-oriented-database-benchmarks/4321

Related Content

Integration of Relational and NoSQL Databases

(2018). *Bridging Relational and NoSQL Databases* (pp. 239-281).

www.irma-international.org/chapter/integration-of-relational-and-nosql-databases/191985

Conceptual Data Modeling Patterns: Representation and Validation

Dinesh Batra (2005). *Journal of Database Management* (pp. 84-106).

www.irma-international.org/article/conceptual-data-modeling-patterns/3333

Indexing Textual Information

Ioannis N. Kouris, Christos Makris, Evangelos Theodoridis and Athanasios Tsakalidis (2009). *Database Technologies: Concepts, Methodologies, Tools, and Applications* (pp. 196-204).

www.irma-international.org/chapter/indexing-textual-information/7911

Data Dissemination

Ludger Fiege (2005). *Encyclopedia of Database Technologies and Applications* (pp. 105-109).

www.irma-international.org/chapter/data-dissemination/11130

A Cost-Based Range Estimation for Mapping Top-k Selection Queries over Relational Databases

Anteneh Ayanso, Paulo B. Goes and Kumar Mehta (2009). *Journal of Database Management* (pp. 1-25).

www.irma-international.org/article/cost-based-range-estimation-mapping/37210