Benchmarking Database Design for Mixed OLTP and OLAP Workloads

Anja Bog
Hasso Plattner Institute,
University of Potsdam
August-Bebel-Str. 88,
14482 Potsdam, Germany
anja.bog@hpi.uni-potsdam.de

Kai Sachs
SAP AG
Dietmar-Hopp-Allee 16,
69190 Walldorf, Germany
kai.sachs@sap.com

Alexander Zeier
Hasso Plattner Institute,
University of Potsdam
August-Bebel-Str. 88,
14482 Potsdam, Germany
alexander.zeier@hpi.uni-potsdam.de

ABSTRACT

Current database benchmarks are either focusing on online transaction processing (OLTP) or on online analytical processing (OLAP) systems. This traditional separation has to be reevaluated to reflect current trends in the design of database systems. We see a need for a realistic benchmark workload taking both aspects into account. Therefore, we defined a mixed workload and illustrate ways to apply our workload to evaluate the influence of database design on system performance.

Categories and Subject Descriptors

H.2.1 [**Database Management**]: Logical Design—*Data models*

General Terms

Design, Performance

Keywords

Benchmarking, OLAP, OLTP, Workload

1. INTRODUCTION

Traditionally, database applications have been separated into OLTP and OLAP. The reasons for this separation are not application-specific requirements but technical aspects due to hardware limitations and resulting conflicting optimization goals. As a consequence, in today's enterprises, data has to be synchronized between OLTP systems, responsible for handling data from different business interactions, and OLAP systems, thus providing operational reporting. This leads to a significant overhead of keeping different systems up to date and providing analytical data on time.

To date, this is an even more challenging task than several years ago: on the one hand, the adoption of new technologies such as wireless-sensor networks and RFID will strongly increase the amount of transactional data [5], on the other hand, the industry is asking for real-time reporting of these (and other) data. As a potential solution defining a common database approach for OLTP and OLAP systems, e.g., based on in-memory column databases, was proposed in [8]. Existing application scenarios that benefit from such an approach are, e.g., dunning or demand planning. They rely on up-to-date and fine granular transactional data, which conforms to typical transactional applications, but their access patterns are similar to typical analytical applications, i.e., accessing a large number of rows to compute the results. To evaluate the benefits of a common database approach compared to traditional solutions, benchmarks and representative workloads are needed. However, to the best of our knowledge, current benchmarks are either focusing on OLTP or on OLAP systems, but none of them uses a mixed workload scenario. As a consequence, we identified the need for a realistic benchmark workload taking both aspects – transactional and analytical processing – into account.

In our poster, we describe such a mixed workload and illustrate ways to apply our workload to evaluate the influence of database design on system performance at the example of a popular open source database.

ICPE'11, March 14–16, 2011, Karlsruhe, Germany.
ACM 978-1-4503-0519-8/11/03.

2. BACKGROUND

In this section we provide a short introduction into the background of database design and briefly discuss existing database benchmarks and workloads.

2.1 OLTP and OLAP Database Design

For both, OLTP and OLAP, basic rules for the creation of optimized database designs exist. The design goals for optimizing OLTP and OLAP systems are in conflict, meaning that a design, which is optimized for OLAP performance, degrades OLTP performance and vice versa [3].

OLTP data schemes are optimized mainly for the efficient recording of business transactions. As a consequence high levels of redundancy are avoided by normalization to reduce the risk of data inconsistencies and update dependencies [9]. Additional structures for optimized query performance like indexes, materialized views, or precomputed aggregates introduce redundancy, which adds overhead to the insertion of new data, updates, and deletes [6], but is relativized with a growing share of read access and the increasing size of tables while access patterns are constant.

The OLAP data schema, i.e., the star schema, is a query-centric design in contrast to the update-centric design needed for OLTP applications [1]. Therefore, it is optimized for fast browsing and aggregation of mass data and avoids joins over large tables through denormalization. Thereby, the star schema introduces a large amount of redundantly stored data within the data set.

2.2 Workload Mix in Database Benchmarks

The benchmarks of the Transaction Processing Performance Council (TPC) became the de-facto standard in the database area. Their currently active benchmarks are based on a static workload mix. TPC-W [7], a web application benchmark that is marked as obsolete by TPC, supports different workload mixes to show diverse user behavior ranging from browsing products as typical analytical behavior to ordering known products without the need of previous browsing, which is typical transaction processing behavior. To validate the influence of the database design on a certain workload mix, a benchmark is needed that simulates different workload mixes of OLTP and OLAP-style behavior. For a detailed comparison of existing database benchmarks we refer to [2].

3. A NOVEL WORKLOAD MIX USED IN BENCHMARKING

To bridge the gap between OLTP and OLAP workloads we introduced a benchmark named **c**_omposite_ **b**_enchmark for_ **t**_ransaction processing and operational_ **r**_eporting_ (CBTR) in [2]. It revives the workload mix concept of TPC-W and simulates varying shares of transactional and analytical behavior to produce different workloads. Similar to TPC's OLTP benchmark TPC-C [4], the scenario of this benchmark is composed of the order-to-cash process and the related part of accounting. This scenario is equally important for transaction processing, e.g., incoming orders, outgoing invoices, incoming payments, and finding late payments to trigger dunning, as it is for reporting, e.g., analyzing the levels of order fulfillment, determining all open orders of customers to ensure timely deliveries, or calculating the average processing time of an order for validating the own service levels and keeping customer satisfaction at a high level. Furthermore, it is easily comprehensible since the process of sales, invoicing, and payment is omnipresent in everyday life.

As part of the benchmark, an example database design in first normal form was defined. However, a benefit of the benchmark is that we can apply it on different database designs without changing the workload definition itself.

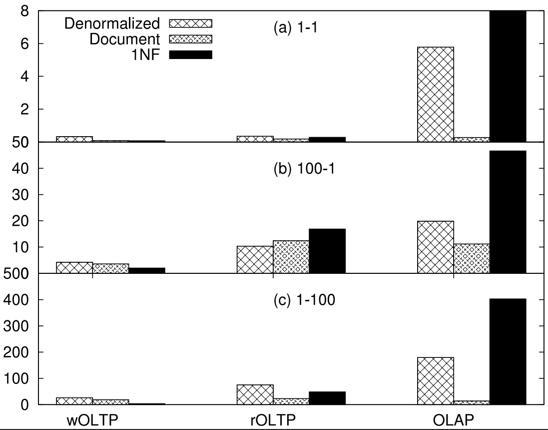
4. DATABASE DESIGN VARIANTS

In the following we illustrate how the benchmark can be used to analyse the impact of different database designs. For this purpose, we introduce two design variations and compare them against each other in a case study. Both design variants favor analytical processing.

The first variation, called _Document (DO)_, decreases the number of joins needed in OLAP queries as well as read-only transactions (_rOLTP_) by denormalizing the header and item tables of the transactional data set. Header and item tables are a typical pattern encountered in OLTP database designs. It reduces redundancy as, e.g., common data of one sales order is not stored together with each of its line items, but is referenced instead. The second variation, called _Denormalized (DE)_, is an even further denormalized schema reducing eight transactional tables from 1NF to two tables according to the table sets needed for the OLAP queries.

We measured the performance of both designs on MySQL InnoDB and analysed three different workload settings (all clients run concurrently): _(a) 1–1:_ one client is running transactional queries; one is running analytical queries; _(b) 100–1:_ 100 clients are running transactional queries; one is running analytical queries; _(c) 1–100:_ one client is running transactional queries; 100 are running analytical queries.

As illustrated in Figure 1, both schema variants are highly beneficial for OLAP queries, while inducing only slight overhead for the write-access OLTP queries (_wOLTP_), with _Document_ being the clear favorite.



	1-1			100-1			1-100		
	_DE_	_DO_	_1NF_	_DE_	_DO_	_1NF_	_DE_	_DO_	_1NF_
wOLTP	0.32	0.09	0.07	4.21	3.58	2.00	26.23	18.62	3.70
rOLTP	0.35	0.18	0.28	10.33	12.43	16.85	75.20	23.01	48.61
OLAP	5.78	0.27	7.97	19.82	11.17	46.60	179.56	13.78	402.69

Figure 1: Average response times (sec.) of queries for database design variants

5. CONCLUSION

In many application scenarios, both transactional and analytical processing exist in parellel and, furthermore, are closely related to each other. We identified a lack of benchmark workloads for such scenarios. Therefore, we described a benchmark including a novel workload targeting these kinds of applications. We showed, how the underlying database design can be adjusted to a workload in order to stress certain aspects of the underlying database and, at the example of InnoDB, we applied our workload to analyse the impact of different database designs on query response time.

As part of our future work, we plan to prepare a comprehensive performance study by applying the benchmark to several database design variants, including those introduced in this paper, on different databases. We assume that the design variants will show even faster acceleration for OLAP queries using column-oriented in-memory databases as the table width, which has a great impact on row-stores, is negligible when accessing only a few columns in column-stores.

6. REFERENCES

[1] S. A. Becker. _Developing Quality Complex Database Systems: Practices, Techniques, and Technologies._ Idea Group, Inc., 2001.

[2] A. Bog, H. Plattner, and A. Zeier. A mixed transaction processing and operational reporting benchmark. _Information Systems Frontiers Journal, Springer_, 2010.

[3] C. French. "One Size Fits All" Database Architectures Do Not Work for DSS. In _SIGMOD_, 1995.

[4] J. Gray, editor. _The Benchmark Handbook for Database and Transaction Systems._ Morgan Kaufmann, 1993.

[5] A. Hinze, K. Sachs, and A. Buchmann. Event-Based Applications and Enabling Technologies. In _DEBS_, 2009.

[6] S. S. Lightstone, T. J. Teorey, and T. Nadeau. _Physical Database Design._ Morgan Kaufmann Publishers, 2007.

[7] D. A. Menasce. TPC-W: A Benchmark for E-Commerce. _IEEE Internet Computing_, 6, 2002.

[8] H. Plattner. A Common Database Approach for OLTP and OLAP Using an In-Memory Column Database. In _Proc. of the 35th SIGMOD_, 2009.

[9] J. Rossberg and R. Redler. _Pro Scalable .NET 2.0 Application Designs._ Apress, 2005.