# Performance Characterization of Databases under Virtualization

*Shoban Chandrabose, Jing Fan, Shike Mei

## 1 Problem & Motivation

The last decade has seen the emergence of large scale public and private clouds platforms to reap the benefits of statistical multiplexing of shared resources. Additionally, these cloud platforms have made the process of deployment and scaling of applications effortless. Hence, unsurprisingly, many new applications and old legacy enterprise applications are being deployed or migrated to these cloud platforms, and database applications are no exception to this trend [3].

Given the importance of this mode of deployment, there has been little work done on understanding the performance of databases on these cloud platforms under virtualization. A combination of factors could affect the performance of databases including the mechanisms used to virtualize physical resources (CPU, memory, disk, network) to virtual machines. The presence of other virtual machines contending for shared resources (CPU, cache, disk) could also affect the performance of databases.

Ailamaki et al. [2] analyzed the performance of traditional databases on modern processors, and identified key issues in their designs, which triggered revisiting the base assumptions. We, in the same vein, plan to analyze the performance of databases under virtualization. We approach this measurement problem in two directions: (a) identify the overhead of virtualization without contention from other VMs; (b) identify the effects of contention on shared resources. We then plan to design optimizations that mitigate the overheads due to virtualization and contention of shared resources.

## 2 Methodology

To characterize the overhead of virtualization, we plan to compare the performance of database running in bare-metal hardware with that running on a VM utilizing the same resources. To characterize the impact of interference due to contention on shared resources, we plan to compare the performance of database running on a VM uncontended for resources with that running with other VMs contending for shared resources. We plan to measure the performance of databases primarily on three types of workloads, namely, simple workload, online transaction processing (OLTP) workload, and online analytic processing (OLAP) workload. Simple workload would comprise of simple database queries like single table aggregation/sorting and simple primary key foreign key join. For OLTP and OLAP workloads, we plan to use the standard benchmarks (TPC-C and TPC-H respectively) from Transaction Processing Performance Council [1]. We plan to evaluate the performance under two virtualization stacks, namely Xen and KVM.

*Authors are listed in last name alphabetical order.

## 3 Plan

The achievable goals of the project are listed into three categories: guaranteed, maybe, and possibly, each being progressively difficult or unlikely to deliver.

**Guaranteed.** Measurement of overheads of virtualization compared to running database on bare-metal without interference from any VMs contending for resources. Measurement of interference due to contention for disks.

**Maybe.** Measurement of interference due to sharing physical CPU cores (i.e., VMs virtualized on same physical cores), and sharing cache hierarchy (i.e., VMs possibly virtualized on different physical cores but sharing some level of cache). Optimizations that mitigate the overheads of virtualization or interference of other VMs possibly involving the use of super pages or smart TLB updating strategies.

**Possibly.** Measurements of performance of databases on public clouds using local or remote disks and using spinning disks or SSDs. Optimizations that mitigate overheads due to suboptimal disk scheduling contending on shared disks.

## References

[1] Transaction Processing Performance Council Benchmarks. `http://www.tpc.org/`.

[2] Anastassia Ailamaki, David J. DeWitt, Mark D. Hill, and David A. Wood. DBMSs on a Modern Processor: Where Does Time Go? In *Proceedings of the 25th International Conference on Very Large Data Bases*, VLDB '99, pages 266–277, San Francisco, CA, USA, 1999. Morgan Kaufmann Publishers Inc.

[3] Michael Stonebraker, Andrew Pavlo, Rebecca Taft, and Michael L Brodie. Enterprise Database Applications and the Cloud: A Difficult Road Ahead. In *Cloud Engineering (IC2E), 2014 IEEE International Conference on*, pages 1–6. IEEE, 2014.