

Blekinge Institute of Technology  
Blekinge Institute of Technology Doctoral Dissertation Series No. 2018:02  
ISSN 1653-2090  
ISBN 978-91-7295-348-2

# PERFORMANCE ASPECTS OF DATABASES AND VIRTUALIZED REAL-TIME APPLICATIONS

Christine Niyizamwiyitira

Akademisk avhandling  
som för avläggande av Teknologie doktorsexamen vid Blekinge Tekniska  
Högskola kommer att offentligens försvaras i J1650, Blekinge Tekniska  
Högskola, 371 79 Karlskrona, Karlskrona, 2018-09-21 , kl. 13.00

Handledare:  
Professor Lars Lundberg, Institutionen  
för datalogi och datorsystemteknik

Mikael Svahnberg, Institutionen för  
programvaruteknik

Opponent:  
Professor Richard Chbeir, University  
of Pau and Pays de l'Adour (UPPA),  
France



Blekinge Tekniska Högskola  
Institutionen för datalogi och datorsystemteknik  
Karlskrona

## Abstract

**Context:** High computing system performance depends on the interaction between software and hardware layers in modern computer systems. Two strong trends that effect different layers in computer systems are that single processors are now more or less completely replaced by multiprocessors, which are often organized into clusters, and virtualization of resources. The performance evaluation of different software on such physical and virtualized resources, is the focus of this thesis.

**Objectives:** The objectives of this thesis are to investigate the performance evaluation of SQL and No SQL database management systems, namely Cassandra, CouchDB, MongoDB, PostgreSQL, and RethinkDB; and soft real-time application namely, voice-driven web. Scheduling algorithms for resource allocation for hard real-time applications on virtual processor are also investigated.

**Methods:** Experiment is used to measure the performance of SQL and No SQL management systems on cluster. It is also used to develop a prototype and predicts processor performance of voice-driven web on multiprocessors. Theoretical methods are used to model and design algorithms to schedule real-time applications on the virtual processor machine. Simulation is used to quantify the performance implications of certain parameter values in our theoretical results and to compare expected performance with theoretical bounds in our schedulability tests.

**Results:** The performance of Cassandra, CouchDB, MongoDB, 2

PostgreSQL, and RethinkDB is evaluated in terms of writing and reading throughput and latencies in cluster computing. For reading throughput, all database systems are horizontally scalable as the cluster's nodes number increases, however, only Cassandra and couchDB exhibit scalability for data writing. The overall evaluation shows that Cassandra has the most writing scalable throughput as the number of nodes increases with a relative low latency, whereas PostgreSQL has the lowest writing latency, and MongoDB has the lowest reading latency.

The architectures' tradeoffs of voice-driven web show that the voice engine should be installed on the server instead of being on the mobile device, and performance evaluations show that speech engine scales with respect to the number of cores in the multiprocessor with and without hyperthreading.

The thesis presents scheduling techniques for real-time applications that runs in virtual machines which are time sharing the processor. Each virtual machine's period and execution time that allow real-time applications to meet their deadlines can be defined using these techniques. Simulation results show the impact of the length of different VM periods with respect to overhead. The tradeoffs between resources consumption and period length are also given. Furthermore, a utilization based test for scheduling real-time application on virtual multiprocessor is presented. This test determines if a task set is schedulable or not. If the task set is schedulable the algorithm provides the priority for each task. This algorithm avoids Dhall's effect, which may cause task sets with even very low utilization to miss deadlines.

**Conclusions:** The thesis presented the performance evaluation of reading and writing throughput and latencies for SQL and NoSQL management systems in the cluster computing. The thesis quantifies the tradeoffs of voice-driven web architectures and the performance scalability of the speech engine with respect to number of cores of the multiprocessor. Furthermore, this thesis proposes scheduling algorithms for real-time 3

application with hard deadline on virtual processors, either as a single core processor or as a multicore processor.

**Keywords:** SQL and NoSQL database, Bigdata management systems, Structured and non-structured Database Evaluation, Voice-driven web, Multicore performance prediction, Hard real-time Scheduling, Virtual Multiprocessor Scheduling