



The Impact of the Storage Tier: A Baseline Performance Analysis of containerized DBMS

Daniel Seybold, Christopher B. Hauser, Georg Eisenhart,
Simon Volpert, and Jörg Domaschka

The Impact of the Storage Tier: A Baseline Performance Analysis of containerized DBMS

Daniel Seybold✉, Christopher B. Hauser, Georg Eisenhart, Simon Volpert, and Jörg Domaschka

Institute of Information Resource Management, Ulm University, Germany
{firstname.lastname, joerg.domaschka}@uni-ulm.de

Abstract. Containers emerged as cloud resource offerings. While the advantages of containers, such as easing the application deployment, orchestration and adaptation, work well for stateless applications, the feasibility of containerization of stateful applications, such as database management system (DBMS), still remains unclear due to potential performance overhead. The myriad of container operation models and storage backends even raises the complexity of operating a containerized DBMS. Here, we present an extensible evaluation methodology to identify performance overhead of a containerized DBMS by combining three operational models and two storage backends. For each combination a memory-bound and disk-bound workload is applied. The results show a clear performance overhead for containerized DBMS on top of virtual machines (VMs) compared to physical resources. Further, a containerized DBMS on top of VMs with different storage backends results in a tolerable performance overhead. Building upon these baseline results, we derive a set of open evaluation challenges for containerized DBMSs.

Keywords: container, YCSB, benchmarking, DBMS, MongoDB

1 Introduction

The raise of containers, containerization, and container orchestration [3] has a great influence on the structure of distributed applications, and greatly eased the operation of such systems by finally leveraging the realisation of continuous deployment. Support for containers is offered beside traditional virtual machine offerings by Amazon Elastic Container Service¹ and OpenStack Magnum².

Much of the success of containers is a consequence of the fact that they enable a quick installation of pre-packaged software components, which is a prerequisite for handing overload (scale out), bug fixing (software upgrade), and replacing failed components (fault tolerance). All of these concepts work fine for mostly stateless components such as load balancers, web and application servers, message queues, and also caches. Yet, despite recent attention in the

¹ <https://aws.amazon.com/ecs/>

² <https://wiki.openstack.org/wiki/Magnum>

field [2, 14], it is currently unclear to what extent containerization is suited for and beneficial to the operation of stateful applications. Database management systems (DBMS) are an important representative of this type of applications and a crucial part of Big Data and IoT applications.

While the containerization of DBMS particularly eases the usage of features of modern DBMS such as horizontal scalability or high availability, at least two challenges remain: (a) The general runtime overhead of containerized DBMS is unknown. (b) The container eco-system offers a myriad of different storage backends and their impacts on performance are also unclear.

Only with an answer to these baseline questions, it is beneficial to think about more sophisticated questions such as placement of state and data migration. This paper is an initial step to identify further research and engineering challenges with respect to containerized DBMS. Our contributions are as follows: (i) We introduce three different operational models for DBMS ranging from bare metal to containers in virtual machines. (ii) We analyse the landscape of storage backends for containers and their pros and cons. (iii) For three operational models and two storage backends we evaluate the performance for the well known MongoDB³ DBMS under various workloads. In contrast to related work, our main focus is not on a performance comparison between containerized and virtualised execution. (iv) Based on the outcome of the evaluation, we propose open challenges for modelling and evaluating DBMS performance.

The remainder of this document is structured as follows: Section 2 discusses the containerization of stateful applications. Section 3 defines the evaluation methodology, while Section 4 presents the evaluation environment. Section 5 discusses the results and derives open evaluation challenges for containerized DBMS. Section 6 presents related work, and Section 7 concludes.

2 Challenges for Containerization of Stateful Applications

Containerization in the context of cloud computing is besides *hardware virtualization* for virtual machines (VMs) so called *operating-system (OS) virtualization* for containers. In hardware virtualization a hypervisor manages the resource allocation and operating state of virtual machines. OS-Virtualization uses operating system features to create lightweight isolated environments, known as containers. Container engines allocate resources and access to *e.g.* networking and storage, the popular Docker⁴ engine. Orchestrators manage VMs or containers across hypervisors or container engines [1, 3, 15]. These virtualization approaches provide the different operational models depicted in Fig. 1(a) where each operation model combines the benefits and drawbacks of the respective virtualization approaches: Hardware virtualization securely isolates with fixed hardware-oriented offers; OS virtualization provides less strong isolation with soft hardware limits [12].

In multi-tier applications, stateful components require to store data temporarily or even durably, *i.e.* by instantiating (distributed) DBMS or stateful

³ <https://www.mongodb.com/de>

⁴ <https://www.docker.com>

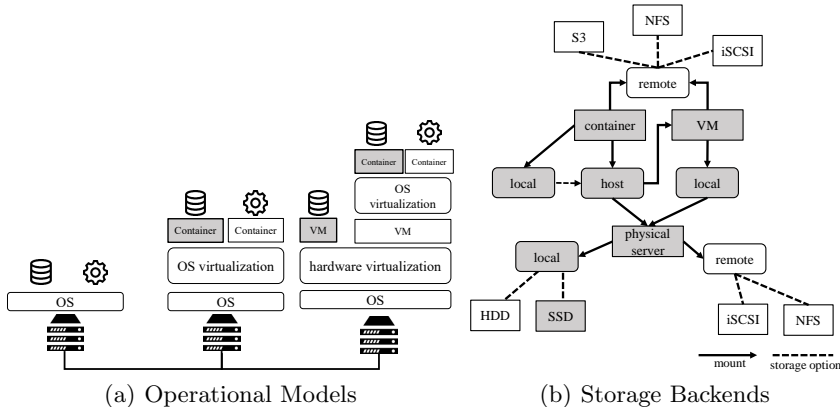


Fig. 1. Containerization of Stateful Applications

caches. Fig. 1(b) lists potential storage backends for VMs and containers, which leads to challenging decisions when deploying stateful containers, especially in large-scale set-ups. Challenges in this field include *(i)* performance aspects such as throughput and latency, *(ii)* support for scalability, *i.e.* considering parallel read/write access, and *(iii)* failure strategies and recovery mechanisms. Since VMs and containers isolate customers to share infrastructure, performance interferences may occur whenever resources (*e.g.* storage) are utilised, which are not directly under control of the container engine and the underlying kernel.

The focus of our work evaluates the performance aspect of containerized DBMS with respect to different storage backends for containerized DBMS on physical hardware over DBMS in VMs to containerized DBMS on top of VMs. The performance and runtime overhead of these approaches are evaluated in the following.

3 Evaluation Methodology

In this section, we define an extensible evaluation methodology for the identification of potential performance overhead of common operation models for containerized DBMS. In the following, the methodology is defined on a conceptual level, while Section 4 describes the technical implementation.

3.1 System Architecture

In order to provide a concise analysis of the potential performance overhead of containerized DBMS in the cloud, we define an extensible system architecture, which comprises three common operational models for DBMS, highlighted in the grey boxes in Fig. 1(a). Each operational model is defined by its virtualisation, *i.e.* OS virtualisation for container, hypervisor for VMs or container on top

of VMs. Further, we apply two storage backends for the containerized DBMS, namely *local* for using the container filesystem and *host* for using the hosting resource filesystem as depicted in Fig. 1(b). For the VM-based DBMS, we apply the local filesystem provided by the hypervisor. The resulting resource configurations are depicted in Table 1. While *remote* storage is also a common storage configurations for containerized DBMS, it is omitted in this work to reduce the interference factor of the network and will be targeted in future evaluations. In addition, we do not use any container-specific network virtualisation as the focus relies on compute, memory and storage.

Table 1. Operational Models and Storage Backends

ID	Operational Model [<i>physical</i> (<i>P</i>), <i>container</i> (<i>C</i>), <i>VM</i>]	Storage Backend [<i>local</i> (<i>L</i>), <i>host</i> (<i>H</i>), <i>remote</i> (<i>R</i>)]
P-C-L	physical + container	local
P-C-H	physical + container	host
VM-L	VM	local
VM-C-L	VM + container	local
VM-C-H	VM + container	host

3.2 Workload and DBMS

In favour of emulating container-centric workloads, we define a *write-heavy* (*w-h*) workload, emulating the storage of sensor data and a *read-heavy* (*r-h*) workload, emulating a social media application with mostly reads and barely update operations. Both workloads are defined in a *memory-bound* version, *i.e.* the whole data set fits into memory and a *disk-bound* version, *i.e.* the data is larger than the available memory. As workload generator, we select the Yahoo Cloud Serving Benchmark (YCSB) [4], which is widely used in performance studies on NoSQL DBMSs. YCSB offers web-based workloads based on create, read, update and delete (CRUD) operations, enabling the emulation of container-centric workloads [10].

As exemplary containerized DBMS, we select document-oriented MongoDB for our evaluation as it is a NoSQL DBMS⁵. MongoDB emphasizes its operation on virtualised resources⁶. Records are stored as *documents* within *collections*. While MongoDB supports a distributed architecture, we select a single node setup for our evaluation to reduce potential interference factors such as network jitter or MongoDB specific data distribution algorithms. Yet, our methodology can easily be extended for a distributed setup and also MongoDB can be exchanged with any desired DBMS.

⁵ <https://db-engines.com/de/ranking>

⁶ <https://www.mongodb.com/containers-and-orchestration-explained>

3.3 Metrics

For each evaluation scenario the following metrics are collected to analyse the results: *throughput* in operations per seconds and *latency* per operation type in μs . Each evaluation scenario is repeated ten times to ensure significant results and for the all metrics the minimum, maximum, average and standard deviation are provided. In addition, system metrics (CPU, RAM, I/O, network) are monitored during each evaluation scenario for MongoDB and the YCSB to provide reliable results by ensuring that none of the system resources creates a bottleneck.

3.4 Evaluation Execution

Our methodology comprises the memory-bound (mb) and disk-bound (db) evaluation scenarios. Each scenario starts with the w-h workload, followed by the r-h workload. Each workload is executed against the resource combinations of operational models and storage backends presented in Table 1. Hence, the execution (E) of the memory-bound and disk-bound scenarios can be expressed as *scenario:E(mb(wh,r-h))*, e.g. *P-C-L:E(mb(wh,r-h))* and *P-C-L:E(db(w-h,r-h))*

4 Evaluation Environment

Based on the introduced evaluation methodology in Section 3, the following presents its implementation for a private, OpenStack-based cloud⁷ (version Pike) with full and isolated access to all physical and virtual resources. In order to reduce potential resource interference and to guarantee reproducible results, we use the availability zones feature of OpenStack to dedicate one physical host for spawning the required VMs and containers. The resulting evaluation environment for the specified evaluation scenarios is depicted in Fig. 2. In the following, the implementation details for the resources, MongoDB and YCSB are presented.

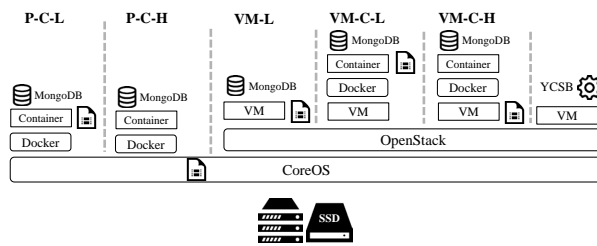


Fig. 2. Evaluation Environment

⁷ <https://www.openstack.org/>

4.1 Resources

As depicted in Fig. 2, all containers and VMs are located on the same physical host, which has enough resources for running the YCSB VM and the DBMSs without resource interference (*i.e.* no overbooking). Further, this set-up only uses the host-internal network interfaces and avoids the overhead of the Open-Stack network service. Accordingly, all containers are configured to use the host network interface via `--network host`. The available resources of the respective physical host, container and VMs are described in Table 2. In order to ensure comparable results, the container resources on the physical host (*i.e.* P-C-L and P-C-H) are limited to 4 cores and 4GB RAM. The containers on CoreOS use the kernel version 4.14.19-coreos while the VM and container inside the VMs use the kernel version 4.4.0-127-generic.

Table 2. Evaluation Scenario Resources

Resource	Virtualisation	OS	Cores	RAM	FS	Storage
Physical host	-	CoreOS 1632	16 ⁸	64GB	Ext4	512GB ⁹
MongoDB container	Docker 18.04	Ubuntu 16.04	4	4GB	overlay2	40GB
MongoDB VM	KVM, QEMU 1.5.3	Ubuntu 16.04	4	4GB	Ext4	40GB
YCSB VM	KVM, QEMU 1.5.3	Ubuntu 16.04	4	2GB	Ext4	10GB

4.2 MongoDB and YCSB

The evaluation scenarios are based on a vanilla deployment of MongoDB and the YCSB to ensure a baseline performance evaluation of MongoDB containerization. The relevant configurations for MongoDB and the YCSB are listed in Table 3. Further, the YCSB operation distribution for the *w-h* workload are 100% write operations and for the *r-h* workload 95% read operations and 5% update operations. Table 3 also highlights overall collection size of each workload version as the number of records for the *memory-bound* version results in a 2 GB MongoDB collection, while the *disk-bound* version results in a 10 GB MongoDB collection. The MongoDB binding of YCSB is configured with the write concern option¹⁰ `w=1` and `j=false`, *i.e.* write operations are acknowledged by MongoDB after they are put into memory.

⁸ 2xIntel XeonE5-2630v3 8-Core Haswell 2.4Ghz

⁹ 2x 256GB SSD of type SAMSUNG MZ7WD240HAFV-00003

¹⁰ <https://docs.mongodb.com/manual/reference/write-concern/>

¹¹ <https://github.com/brianfrankcooper/YCSB/releases/tag/0.12.0>

Table 3. YCSB VM Details

MongoDB Configuration	Value	YCSB Configuration	Value
Version	3.6.3 (CE)	Version	0.12 ¹¹
services	1 x mongod	record size	1 KB
storage engine	WiredTiger	# of records (memory-bound)	2.000.000
replication	off	# of records (disk-bound)	10.000.000
		# of operations	10.000.000
		# of threads	20
		distribution	zipfian

4.3 Portability and Reproducibility

The execution of each scenario is fully automated by utilizing ready to deploy artifacts, which are together with the results publicly available¹²; their release as open research data is currently under way. For the Docker images we make use of the Docker native capabilities of building images based on Dockerfiles. The VM images are generated by Packer¹³. Packer processes a Packerfile, which is similar to a Dockerfile, but uses a multitude of different virtualization providers to generate and store the image. In our case we are using OpenStack Glance¹⁴. This approach enables fellow researchers to reproduce, validate and extend our scenarios by changing the cloud provider or benchmark a different DBMS.

5 Results and Discussion

In the following, we present and discuss the results of the memory- and disk-bound evaluation scenarios (*cf.* 1) based on defined metrics in Section 3.3.

5.1 Evaluation Results

The throughput results are depicted in Fig. 3 and latency results in the Fig. 4. Each plot represents the results of the respective scenario, *i.e.* memory-bound or disk-bound and the respective workload, *i.e.* w-h or r-h. For the latency plots of the r-h workloads, the first bar of each operational model always represents the read latency while the second bar represents the update latency. As remark, the results reflect the best case operational models as the DBMS and the YCSB are operated on the same, isolated physical host (*cf.* Section 4.1).

The results shows a significant throughput and latency overhead for operating a DBMS on top of VMs instead of using physical hardware. These results

¹² <https://github.com/omi-uulm/Containerized-DBMS-Evaluation>

¹³ <https://packer.io>

¹⁴ <https://docs.openstack.org/glance/pike/>

confirm previous performance studies of memory-bound workloads for former Docker versions [5]. A novel insight is shown by the results for the DBMS operated in a container on VM the (VM-C-L,VM-C-H) as the performance only decreases slightly compared to DBMS directly operated on VMs (VM-L), *e.g.* VM-C-H achieves 6% less throughput than VM-L for the w-h workload and 13% for r-h of the disk-bound scenario. Hence, if VMs are the only available resource, operating the DBMS in container on top of the VMs can be beneficial to exploit container orchestrators or the soft resource limits to operate additional containerized applications next to the DBMS on the same VM [12].

The second insight of the results is the performance overhead of the internal Overlay2 filesystem of Docker. The container on physical hardware with the Overlay2 filesystem (P-C-L) shows significantly less throughput and higher latencies compared to the container using the host filesystem (P-C-H). This finding most clearly applies for the r-h workload of the disk-bound scenario (*cf.* Fig. 3(d) and Fig. 4(d)). The Overlay2 overhead is also present on container running on VMs (VM-C-H) but to a lower extent

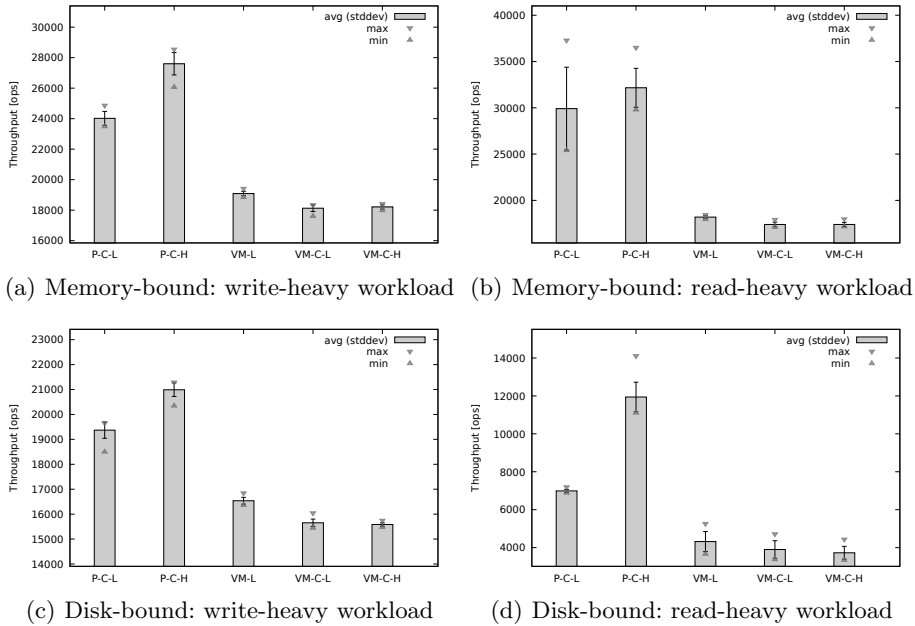


Fig. 3. Throughput Results

5.2 Open Evaluation Challenges

The results of our baseline evaluation, show that containers are suitable to operate DBMS, even on top of VMs. Yet, the operational models reveal significant

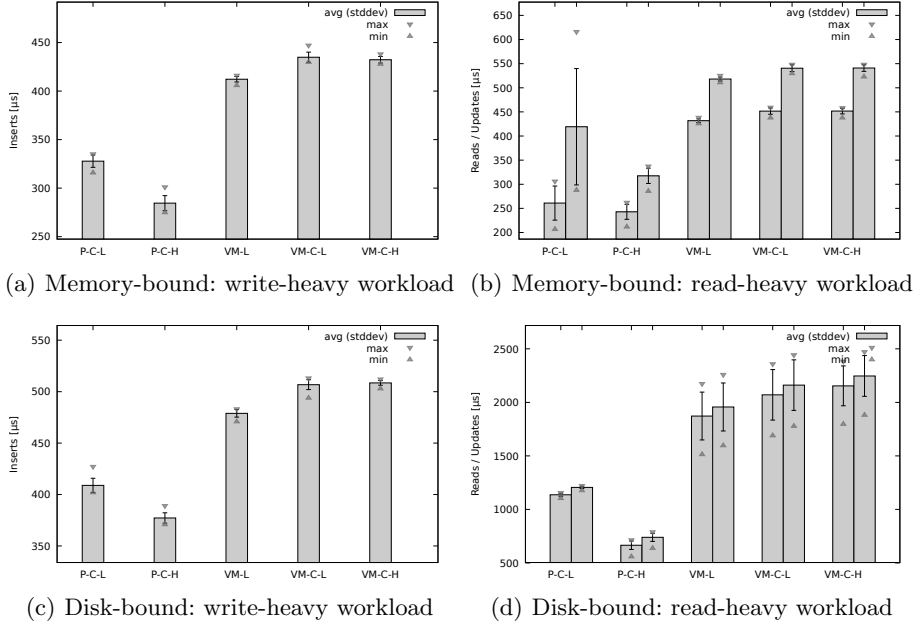


Fig. 4. Latency Results

performance deviations, also dependent on the memory- or disk-bound scenarios. Hence, the selection of the operational model in conjunction with the storage backend is a crucial decision for the DBMS operator, which has to be driven by the available operational models, the targeted performance and the demand of optional orchestration features.

Based on our methodology and the presented baseline results, we derive a set of open evaluation challenges, which have to be addressed to drive the selection process of the operational model for containerized DBMS: (i) The performance of the presented operational models needs to be evaluated based on public cloud offerings by considering additional hypervisors and containers with respect to memory- and disk-bound DBMS workloads. (ii) The presented storage backends require a dedicated evaluation with respect to different local and remote container storage drivers. This also comprises local and remote block storage of the host resource. (iii) As the DBMS performance deviation of the operational models VM-L, VM-C-L, and VM-C-H are in a tolerable margin, the advantages of VM-C-L and VM-C-H have to be analysed with respect to orchestration and the co-location with suitable applications. (iv) The presented methodology needs to be extended for additional DBMS to evaluate their containerization feasibility and container orchestration features with respect to the scalability and elasticity of distributed DBMS [11]. Further, container orchestration features for high availability and migration of containerized DBMS need to be evaluated in the context of the presented operational models and storage backends.

6 Related Work

With the increasing usage of containers besides VMs in cloud offerings, different comparative analysis of their performance overhead and resource isolation capabilities have been conducted. Moreover, the containerization of DBMS moved into the focus, especially in combination with container orchestrators.

6.1 Performance Overhead and Resource Isolation

The performance overhead of VMs in contrast to Docker containers running on physical hardware is evaluated by [5]. SysBench¹⁵ is used to compare the throughput of MySQL running on VMs against containerized MySQL. The results show that VMs cause a higher performance overhead as Docker containers for disk-intensive workloads. Further, the usage of the Docker AUFS storage driver causes a higher performance overhead as the usage of Docker volumes.

A related performance comparison of KVM VMs, Docker and LXC containers and a lightweight VM approach based on OSv¹⁶ is provided by [7]. The evaluation is based on different resource-specific micro-benchmarks and the results accord to [5] for the lower performance of VMs for disk-intensive workloads.

An analysis and evaluation of the Docker storage drivers with respect to filesystem performance is presented by [13]. The results demonstrate that the choice of the storage driver can influence the filesystem performance significantly where the Btrfs storage driver achieves the best performance but less stability as the other storage drivers.

The comparative analysis of the resource isolation capabilities of VMs and containers is provided by [16, 12, 8]. While [12] apply resource-specific micro-benchmarks, [16] and [8] use DBMS and respective DBMS workloads to evaluate the resource capabilities. All of these evaluation indicate a stronger resource isolation of VMs, especially for disk-bound workloads.

While existing performance evaluations focus either on micro-benchmarks or apply DBMS only for the evaluation of the resource isolation, our evaluation provides an evaluation across multiple operational model and storage backends. In addition, the operation of container on top of VMs is emphasized by [12] but so far no performance evaluation has considered this operational model.

6.2 Containerization of DBMS

The containerization of DBMS in combination with container orchestrators provides multiple adaptation actions to automate the operation of NoSQL DBMSs [2]. Hereby, the orchestration features of Kubernetes are enhanced with distributed DBMS-specific adaptation rules for proactive and low-cost adaptations to avoid the transfer of data between nodes. While container orchestrators typically manage the disk storage internally, modifying the persistent storage within container

¹⁵ <https://github.com/akopytov/sysbench>

¹⁶ <http://osv.io/>

orchestrators is difficult and hinders their adoption for DBMS [6]. Therefore, [6] present a persistent storage abstraction layer for container orchestrators, which eases the usage of containerized DBMS across different container orchestrators. Yet, the usage of container orchestrators and their internal handling of persistent storage can introduce additional performance overhead for DBMS. Hence, [14] analyse the performance overhead of using remote storage for containerized DBMS within Kubernetes.

While the usage of containerized DBMS with container orchestrators eases the automation of DBMS operation, there is potential performance overhead added by the different handling of the persistent storage of the container orchestrator. While, [14] provide a first step into analysing this overhead for remote storage, we provide a baseline evaluation for container local and host storage backends. In addition, we apply a memory- and disk-bound workload to identify the suitability of container for the respective DBMS workloads.

7 Conclusion and Future Work

The evolution of container leads to a variety of new operational models for distributed applications in the cloud. While containers work fine for stateless applications, stateful applications such as database management systems (DBMS) are receiving increasing attention recently. As DBMS add the persistence aspect to the operational model, storage backends for containers are evolving. Yet, the performance impact of these new operational and storage backends remains unclear for containerized DBMS. Hence, we analyse current operational and storage backends in the context of containerized DBMS. and derive a baseline evaluation methodology for a comparative evaluation of operational models and storage backends. Hereby, we define a memory- and a disk-bound scenario, which is applied on three operational models (container on physical hardware, virtual machines (VMs) and container on VMs) in combination with two storage backends (container filesystem and host filesystem), resulting in 20 evaluation configurations. The evaluation is executed in a private OpenStack with a containerized MongoDB. The results show a significant performance overhead of container running on VMs in contrast to container running on physical hardware. Yet, running container on VMs with different storage backends only causes a tolerable performance impact in contrast to running the DBMS directly on the VM. Further, the usage of the container internal filesystem causes a significant performance overhead compared to using the host filesystem.

Based on these baseline results, we conduct that container are suitable to operate DBMS but additional evaluations are required to get a clear understanding of potential performance bottlenecks. Therefore, we derive a set of open evaluation challenges, which will be addressed in future work: *(i)* evaluating additional operational models implementations; *(ii)* evaluating additional storage backends; *(iii)* consolidation of containerized DBMS and processing applications on top of VMs and *(iv)* the feasibility of DBMS containerization and orchestration for different DBMS. These challenges will be addressed within [9].

Acknowledgements The research leading to these results has received funding from the EC’s Framework Programme HORIZON 2020 under grant agreement number 731664 (MELODIC) and 732667 (RECAP).

References

1. Baur, D., Seybold, D., Griesinger, F., Tsitsipas, A., Hauser, C.B., Domaschka, J.: Cloud orchestration features: Are tools fit for purpose? In: UCC. pp. 95–101. IEEE (2015)
2. Bekas, E., Magoutis, K.: Cross-layer management of a containerized nosql data store. In: Integrated Network and Service Management (IM), 2017 IFIP/IEEE Symposium on. pp. 1213–1221. IEEE (2017)
3. Burns, B., Grant, B., Oppenheimer, D., Brewer, E., Wilkes, J.: Borg, omega, and kubernetes. *Queue* 14(1), 10 (2016)
4. Cooper, B.F., Silberstein, A., Tam, E., Ramakrishnan, R., Sears, R.: Benchmarking cloud serving systems with ycsb. In: ACM symposium on Cloud computing. pp. 143–154. ACM (2010)
5. Felten, W., Ferreira, A., Rajamony, R., Rubio, J.: An updated performance comparison of virtual machines and linux containers. In: ISPASS. pp. 171–172. IEEE (2015)
6. Mohamed, M., Warke, A., Hildebrand, D., Engel, R., Ludwig, H., Mandagere, N.: Ubiquity: Extensible persistence as a service for heterogeneous container-based frameworks. In: OTM Confederated International Conferences” On the Move to Meaningful Internet Systems”. pp. 716–731. Springer (2017)
7. Morabito, R., Kjällman, J., Komu, M.: Hypervisors vs. lightweight virtualization: a performance comparison. In: IC2E. pp. 386–393. IEEE (2015)
8. Rehman, K.T., Folkerts, E.: Performance of containerized database management systems. In: DBTEST. p. 6. ACM (2018)
9. Seybold, D.: Towards a framework for orchestrated distributed database evaluation in the cloud. In: Proceedings of the 18th Doctoral Symposium of the 18th International Middleware Conference. pp. 13–14. ACM (2017)
10. Seybold, D., Domaschka, J.: Is distributed database evaluation cloud-ready? In: ADBIS. pp. 100–108. Springer (2017)
11. Seybold, D., Wagner, N., Erb, B., Domaschka, J.: Is elasticity of scalable databases a myth? In: IEEE Big Data. pp. 2827–2836. IEEE (2016)
12. Sharma, P., Chaufournier, L., Shenoy, P., Tay, Y.: Containers and virtual machines at scale: A comparative study. In: Proceedings of the 17th International Middleware Conference. p. 1. ACM (2016)
13. Tarasov, V., Rupperecht, L., Skourtis, D., Warke, A., Hildebrand, D., Mohamed, M., Mandagere, N., Li, W., Rangaswami, R., Zhao, M.: In search of the ideal storage configuration for docker containers. In: FAS* W. pp. 199–206. IEEE (2017)
14. Truyen, E., Reniers, V., Van, D., Landuyt, B.L., Joosen, W., Bruzek, M.: Evaluation of container orchestration systems with respect to auto-recovery of databases (2017)
15. Verma, A., Pedrosa, L., Korupolu, M., Oppenheimer, D., Tune, E., Wilkes, J.: Large-scale cluster management at google with borg. In: Proceedings of the Tenth European Conference on Computer Systems. p. 18. ACM (2015)
16. Xavier, M.G., De Oliveira, I.C., Rossi, F.D., Dos Passos, R.D., Matteussi, K.J., De Rose, C.A.: A performance isolation analysis of disk-intensive workloads on container-based clouds. In: PDP. pp. 253–260. IEEE (2015)