Jens Knoop, Uwe Zdun (Hrsg.): Software Engineering 2016,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2016 27

Automated Workload Characterization for I/O Performance Analysis in Virtualized Environments

Axel Busch[1], Qais Noorshams[2], Samuel Kounev[3], Anne Koziolek[4], Ralf Reussner[5], Erich Amrehn[6]

Modern applications, such as mail servers, file servers, or video servers show highly I/O-intensive workload patterns. Their huge data volumes require powerful storage infrastructures. These applicarions are increasingly deployed in virtualized environments due to cost efficiency aspects. Nevertheless, consolidating several applications on one shared infrastructure introduces complex performance implications due to mutual interferences. To consolidate several applications while respecting certain Service Level Agreements necessitates a predicion of these implications up front. Such a prediction, however, requires tailored performance models that in turn require a significant amount of expertise to create the models [Kr12, Kr11]. Moreover, their accuracy depends on the quality of the input parameters which are often unclear how they could be determined [CH11]. We adress this discrepancy in our work. We develop an automated workload characterization approach to extract workload models [Ko09] that are representations of the main aspects of I/O-intensive applications in virtualized environments. We have tailored our approach to enable a non-invasive and lightweight monitoring, yet with a level of abstraction such that the parameters are practically obtainable. To evaluate our approach, we perform a comprehensive evaluation demonstrating its workload modeling performance for common business workloads using two case studies. The case studies demonstrate typical real-worl scenarios, such as consolidation of several workloads on one machine, and workload migration between two systems.

Our approach analysis the *low-level* read and write request that are generated by a *high-level* workload, such as a file server workload. The requests' properties are described by a formalized set of metrics (determined from one of our previous works [NKR13]). These metrics are particularly build for modeling I/O-intensive workloads in virtualized environments. Once extracted, the values are mapped to our reference benchmark, the *Flexible File System Benchmark* (FFSB). FFSB allows to emulate the original low-level workload on the target system.

The metrics set is comprised of six metrics: The average *file size* determining the space of the files physically allocated on the disk, limiting sequential requests. The *file set size* con-

[1] Karlsruhe Institute of Technology, busch@kit.edu
[2] Karlsruhe Institute of Technology, noorshams@kit.edu
[3] University of Wuerzburg, samuel.kounev@uni-wuerzburg.de
[4] Karlsruhe Institute of Technology, koziolek@kit.edu
[5] Karlsruhe Institute of Technology, reussner@kit.edu
[6] IBM Research&Development, amrehn@de.ibm.com

siders the total allocated space that influences the locality of requests, and other strategies, such as data placement and caching strategies. Further, we include the *workload intensity* that determines the running of parallel jobs accessing the disks. The *request mix* determines the ration between read and write requests, while the *average request size* models the average size of each read and write request that is accessed sequentially. Finally, the *disk access pattern* is represented by a heuristic algorithm extracting the ratio of sequential and parallel accesses.

Our approach uses the aforementioned metrics to extract the workload characteristics. For an automated execution, we formalized and implemented our metrics set in the *Storage Performance Analyzer* (SPA). SPA is a tool allowing to extract the workload characteristics automatically using a certain set of metrics.

For our evaluation, we use two state-of-the-art high performant virtualization environments, namely an IBM SYSTEM Z, equipped with a DS8700 storage system, and a SUN FIRE X4440 server system. We use two real-world workloads, namely a mail server and a file server workload. We generated our workloads using *Filebench*, a storage system benchmark that is widely used in the performance modeling community [Kr12, Ah07]. The extracted results are then used in two different scenarios, namely a migration and additionally a consolidation scenario. In the first, we used FFSB with the metric values of the file server workload to estimate its performance on the Sun Fire system. In the second, we emulate both workloads on the Sun Fire system at the same time. Again, it should be mentioned that both workloads are characterized on the IBM system. For the migration scenario we could show a prediction error of 21.59 % for read, and 20.98 % error for the write requests. In case of the consolidation scenario, we demonstrate 12.95 % error for read and 24.52 % for write requests. Both scenarios show the applicability of our approach that benefits in a fast and low-overhead estimation of an I/O-intensive workload that does not rely on complex performance prediction models. The demonstrated accuracy should be sufficient for initial estimations of the workload behaviour.

Literaturverzeichnis

[Ah07] Ahmad, Irfan: Easy and Efficient Disk I/O Workload Characterization in VMware ESX Server. IEEE Computer Society, 2007.

[CH11] Chiang, Ron C.; Huang, H. Howie: TRACON: interference-aware scheduling for data-intensive applications in virtualized environments. SC'11, New York, NY, USA, 2011.

[Ko09] Kounev, Samuel: Wiley Encyclopedia of Computer Science and Engineering - chapter Software Performance Evaluation. Wiley-Interscience and John Wiley & Sons Inc., 2009.

[Kr11] Kraft, Stephan; Casale, Giuliano; Krishnamurthy, Diwakar; Greer, Des; Kilpatrick, Peter: IO Performance Prediction in Consolidated Virtualized Environments. ICPE, 2011.

[Kr12] Kraft, Stefan: Performance Models of Storage Contention in Cloud Environments. Journal of Software and Systems Modeling (SoSyM), 2012.

[NKR13] Noorshams, Qais; Kounev, Samuel; Reussner, Ralf: Experimental Evaluation of the Performance-Influencing Factors of Virtualized Storage Systems. Springer Berlin Heidelberg, 2013.