

SPEC - Driving Better Benchmarks

Walter Bays
Oracle Corporation
walter.bays@oracle.com

Klaus-Dieter Lange
Hewlett-Packard Company
klaus.lange@hp.com

ABSTRACT

The driving philosophy for the Standard Performance Evaluation Corporation (SPEC) is to ensure that the marketplace has a fair and useful set of metrics to differentiate systems by providing standardized benchmark suites. This paper gives an overview of SPEC and its continuous drive for better benchmarks.

Categories and Subject Descriptors

H.3.4 [Systems and Software]: Performance evaluation (efficiency and effectiveness)

General Terms

Design, Experimentation, Measurement, Performance, Reliability, Standardization

Keywords

SPEC, SERT, Rating Tool, Benchmark, Energy Efficiency, Power, Server, Storage, Datacenter, EPA

1. IN A NUTSHELL

SPEC [1] was formed from the instigation and sponsorship of Electronic Engineering Times (E.E. Times), and by the cooperative development work of Hewlett-Packard Corp., Sun Microsystems Inc., Apollo Computer Inc., and MIPS Computer Systems Inc. Their effort to develop a benchmark standardizing activity was converted into a non-profit Corporation of the state of California, on November 14, 1988. Its mission is to ensure that the marketplace has a fair and useful set of metrics to differentiate the newest generation of a range of platforms from symmetric multiprocessing (SMP) and Non-Uniform Memory Architecture (NUMA) to clustered multiprocessing server systems.

SPEC has grown to become one of the most successful performance standardization bodies. SPEC's community has developed more than 30 industry-standard benchmarks for system performance evaluation in a variety of application areas and provided thousands of benchmark licenses globally. SPEC publishes several hundred different performance results each quarter spanning a variety of system performance disciplines.

SPEC reviews all submitted results and allows publication of benchmark results only if the results are found to be in compliance with run rules. Members closely monitor the public usage of SPEC results to ensure accordance with SPEC's General Availability and Fair Use rules.

The SPEC membership is open to any interested company or entity and currently includes computer hardware and software companies, educational institutions, and government agencies..

2. ORGANIZATION OVERVIEW

The Corporation is comprised of a Board of Directors (BoD), officers, president, and a staff (SPEC Headquarters) to carry out the business of SPEC. The BoD has established several SPEC Board Committees to handle specific tasks and supports several benchmark development groups under SPEC's umbrella. Each of the groups can support subcommittees, working groups, or project groups on their own.

SPEC has established four of those groups: the Open Systems Group (OSG) [2], the High Performance Group (HPG) [3], the Graphics and Workstation Performance Group (GPWG) [4] and most recently, the newly created Research Group (RG) [5]. These groups cover the major areas of desktop, workstation, handheld devices, and server benchmarking and performance evaluation. Individual committees are responsible for benchmarks characterizing CPU, Cloud Computing, Graphic, JAVA, File System, Session Initiation Protocol (SIP), Virtualization, and Energy Efficiency.

2.1 Open System Group

The OSG is organized into several committees and working groups focusing on the development of component-, and systems-level benchmarks for desktop, workstations, handheld devices and servers running open operating system environments (e.g., SPEC CPU2006, SPECjEnterprise2010, SPECjbb2005, SPECjms2007, SPECjvm2008, SPECsip_Infrastructure2011, SPEC SFS2008, SPECpower_ssj2008, SPECvirt_sc2010).

2.2 High Performance Group

The HPG develops benchmarks that represent large, real applications, in scientific and technical computing supporting industry standard parallel application programming interfaces (APIs), OpenMP and MPI. HPG benchmarks are designed to run on several data set sizes (from a few minutes to days of execution time).

2.3 Graphics and Workstation Performance Group

The GPWG entertains two Project Groups for developing consistent, repeatable graphics and workstation performance benchmarks that reflect user experiences with popular applications.

2.3.1 SPECapc

The Application Performance Characterization (SPECapc) group was formed in 1997 to provide a broad-ranging set of standardized benchmarks for graphics and workstation

applications. The group's current benchmarks span popular CAD/CAM, digital content creation, and visualization applications.

2.3.1 SPECgpc

The Graphics Performance Characterization (SPECgpc) group, began in 1993, establishes performance benchmarks for graphics systems running under OpenGL and other application programming interfaces (APIs). The group's SPECviewperf benchmark is the most popular standardized software for evaluating performance based on popular graphics applications.

2.4 Research Group

The RG is the newest group established to serve as a platform for collaborative research efforts in the area of quantitative system evaluation and analysis, fostering the interaction between industry and academia in the field.

The scope of the group includes computer benchmarking, performance evaluation, and experimental system analysis considering both classical performance metrics such as response time, throughput, scalability and efficiency, as well as other non-functional system properties included under the term dependability, e.g., availability, reliability, and security.

The conducted research efforts span the design of metrics for system evaluation as well as the development of methodologies, techniques and tools for measurement, load testing, profiling, workload characterization, dependability and efficiency evaluation of computing systems.

3. BENCHMARK DEVELOPMENT

The benchmark development is driven by the interest of the members who provide engineering resources to design and implement a new benchmark or workload. In each SPEC group there are formal and informal processes by which consensus is built and tasks are distributed. In general one or more members request establishing a working group to investigate the new area of interest. This working group will create a specific proposal and after approval, a subcommittee or project group tackles implementation (design, code, docs, and rules). When the benchmark is complete it begins general membership review (GMR), the release materials are polished and members run benchmarks for the first round of submissions, which are generally concurrent with the final release. The BoD approves product pricing and any related releases. Once released, the benchmark will be maintained and enter normal submission, review, and publication cycle.

4. RESULT PUBLICATION

SPEC publishes a large set of benchmark results on its web site. Measurements are most often performed and submitted to SPEC by hardware and software vendors testing their own system, nevertheless non-vendors publish results as well. Prior to publication, the results undergo peer review. During this typically two weeks review, the full disclosure report (FDR) is examined by members of the benchmark committee including competitors of the submitting vendors. Questions regarding any aspect of run rule compliance may be raised and answered before the result is accepted for publication. Additionally, those FDRs document all the configuration details and tuning parameters required for an independent party to duplicate the result on the same system.

5. CONCLUSIONS

Two major advantages of SPEC development methods are low cost and open review. Cost is lowered by shared development, common test methodology, and applicability to a range of platforms from single processor systems to clusters. Confidence in results is increased by public availability of the benchmark code, drivers, and detailed run and reporting rules, peer review, and ability of independent third parties to reproduce test results. Further advantages are that benchmarks are retired and replaced every three to five years to keep pace with technology and new benchmarks can be created whenever there is sufficient interest (influential members can have a strong guidance on its direction and space)

6. ACKNOWLEDGMENTS

The authors want to acknowledge all the participants from member organizations and the SPEC HQ who have contributed to the overall success of SPEC.

The name SPEC together with its tool, benchmark and service names are registered trademarks of the Standard Performance Evaluation Corporation (SPEC).

7. REFERENCES

- [1] Standard Performance Evaluation Corporation:
<http://www.spec.org>
- [2] SPEC OSG: <http://www.spec.org/osg/>
- [3] SPEC HPG: <http://www.spec.org/hpg/>
- [4] SPEC GWPG: <http://www.spec.org/gwpg/>
- [5] SPEC Research: <http://research.spec.org/>