Volume 2, Issue 2
April 2017

SPEC Research[SM] Group Newsletter

CONTENTS

CONTACT

Standard Performance Evaluation Corporation (SPEC)
7001 Heritage Village Plaza, Suite 225
Gainesville, VA 20155, USA

SPEC Research Group
Chair: Samuel Kounev (rgchair@spec.org)
Web: http://research.spec.org

ICPE 2018 WILL BE HELD IN BERLIN, GERMANY

Katinka Wolter and William Knottenbelt, the General Chairs of the next ACM/SPEC International Conference on Performance Engineering (ICPE 2018), invite interesting high-quality submissions. The conference will take place in April 2018 in Berlin, Germany.

Read more on page 4

FIVE SPEC RESEARCH WORKING GROUPS REPORT ON THEIR PROGRESS

The SPEC Research Working Groups IDS, Cloud, Big Data, DevOps Performance and the new group Power report on their progress, articles, benchmarks, and technical reports published in the year 2016. The Working Groups are always open for new members, feel invited to join us!

Read more on pages 5-8

SPEC DIXIT DISSERTATION AWARD 2016

The winning dissertation was authored by Scott Beamer III of University of California, Berkeley, under the supervision of Profs. Krste Asanovic and David Patterson. The selection committee appreciates the characterization and treatment of fundamental performance bottlenecks in graph processing and the contribution of the Graph Algorithm Platform (GAP) benchmark suite.

Read more on page 9

SPEC RESEARCH GROUP OFFICERS

Chair:
Samuel Kounev, University of Würzburg, Germany

Vice-Chair:
André van Hoorn, University of Stuttgart, Germany

Secretary:
Klaus-Dieter Lange, HPE, USA

Steering Committee:
J. Nelson Amaral, University of Alberta, Canada
Alexandru Iosup, TU Delft, The Netherlands
Klaus-Dieter Lange, HPE, USA
Samuel Kounev, University of Würzburg, Germany
Qais Noorshams, IBM, Germany
Meikel Poess, Oracle Corporation, USA
Anna Queralt, Barcelona Supercomputing Center, USA
Tilmann Rabl, Bankmark UG, Germany
Petr Tůma, Charles University of Prague, Czech Republic
André van Hoorn, University of Stuttgart, Germany

Publicity Officer:
Christoph Heger, NovaTec Consulting GmbH, Germany

Release Manager:
Vojtěch Horký, Charles Univers. Prague, Czech Republic

Newsletter Editors:
Nikolas Herbst, André Bauer, Piotr Rygielski, University of Würzburg, Germany

Web: http://research.spec.org

SPEC RESEARCH WORKING GROUPS

Cloud Working Group

Chair:
Alexandru Iosup, TU Delft, Netherlands

Vice-Chair and Release Manager:
Nikolas Herbst, University of Würzburg, Germany

Web: https://research.spec.org/en/working-groups/rg-cloud.html

IDS Benchmarking Working Group

Chair:
Marco Vieira, University of Coimbra, Portugal

Vice-Chair:
Alberto Avritzer, Sonatype, USA

Secretary and Release Manager:
Aleksandar Milenkoski, ERNW, Germany

Web: https://research.spec.org/working-groups/rg-ids-benchmarking.html

Big Data Working Group

Chair:
Tilmann Rabl, bankmark, Germany

Vice-Chair:
Todor Ivanov, Goethe University Frankfurt, Germany

Secretary:
Meikel Poess, Oracle Corporation, USA

Release Manager:
John Poelman, IBM, USA

Web: https://research.spec.org/en/working-groups/rg-big-data.html

DevOps Performance Working Group

Chair:
André van Hoorn, University of Stuttgart, Germany

Vice-Chair:
Felix Willnecker, fortiss GmbH, Germany

Secretary:
Robert Heinrich, Karlsruhe Institute of Technology, Germany

Release Manager:
Alexander Wert, NovaTec Consulting GmbH, Germany

Web: https://research.spec.org/en/working-groups/rg-devops-performance.html

Power Working Group

Chair:
Jóakim von Kistowski, University of Würzburg, Germany

Vice-Chair:
Klaus-Dieter Lange, HPE, USA

Secretary:
Mike Tricker, Microsoft, USA

Web: https://research.spec.org/working-groups/rg-power.html

WELCOME TO THE SPEC RESEARCH GROUP NEWSLETTER

We are delighted to present to you the next issue of the SPEC Research Group Newsletter. This regular publication provides information on latest developments, news, and announcements relevant to the benchmarking and quantitative system evaluation communities. Our newsletter is part of our mission to foster the exchange of knowledge and experiences between industry and academia in the field of quantitative system evaluation and analysis. The past year has been very intense and exciting for the SPEC RG. To the major activities and milestones reached, we include:

- New SPEC RG Power Working Group,
- Technical report on the Future of Cloud Metrics
- 7^{th} ACM/SPEC ICPE 2016 in Delft
- 13^{th} IEEE International Conference on Autonomic Computing ICAC 2016 in Würzburg
- 2^{nd} International Workshop on Quality-aware DevOps QUDOS 2016 in Saarbrücken
- SPEC RG face-to-face meeting 2016
- Survey on performance-aware DevOps
- New tools accepted:
  DynamicSpotter, inspectIT, LIKWID

We have been actively working on preparation, planning and organization of the ICPE 2017. We hope that the vivid exchange of ideas during the upcoming ICPE 2017 will be a great motivation for the next year of scientific and engineering work.

We hope that you will enjoy reading the newsletter. We welcome and encourage your contributions for articles and suggestions for future coverage.

Piotr Rygielski, Nikolas Herbst, André Bauer
(Newsletter Editors, University of Würzburg).

ANNOUNCEMENTS

https://www.spec.org/news/

CHAUFFEUR WDK 2.0 RELEASED

April 18, 2017

The Chauffeur WDK 2.0 simplifies workload development for researchers analyzing system performance and power consumption. Targeting both industry and academia the WDK is based on the SERT 2.0 infrastructure, providing runtime and platform configuration support to simplify worklet development. This major release includes significant enhancements to the hardware detection, customization options of generating HTML reports, and developer documentation. It has now reduced memory requirements for the Director when signing results files and reduced the size of the result output for large systems or clusters.

Chauffeur WDK 2.0 added worklet-specific normalization of results and an updated list of supported operating systems including Ubuntu (14.04 LTS and 16.04 LTS) as well as current versions of Windows Server, RHEL, SLES, AIX, and Solaris. The WDK also includes the latest PTDaemon integration for power analyzers and temperature sensors, along with data collection, validation and reporting.

The SPECpower committee that develops the Chauffeur WDK 2.0 includes representatives from AMD, Dell, HPE, IBM, Intel, Microsoft, and the University of Würzburg. Veteran SPEC benchmark developers Hansfried Block, Greg Darnell, and Karin Wulf are supporting contributors. ChauffeurWDK 2.0 has been tested in collaboration with the SPEC RG Power working group, which is presenting an Eclipse integration of the ChauffeurWDK, called Autopilot, at ICPE 2017.

RELEASE OF THE SERT SUITE 2.0

March 28, 2017

SPEC announces the release of the Server Efficiency Rating Tool 2.0, adding a single-value metric, reduces runtime, improves automation and testing, and increases device and platform support. Designed to be simple to configure and use via a comprehensive graphical user interface, the SERT suite uses a set of synthetic worklets to test discrete system components such as processors, memory and storage, providing detailed power consumption data at different load levels. The SERT metric, created with the support of the RG Power Working Group, rates the server efficiency of single- and multi-node servers across a broad span of configurations.

The SPECpower committee that develops the SERT suite includes representatives from AMD, Dell, HPE, IBM, Intel, Microsoft, and the University of Würzburg. Veteran SPEC benchmark developers Hansfried Block, Greg Darnell, and Karin Wulf are supporting contributors. The U.S. EPA provided regular high-level policy and directional guidance to the SPECpower committee, and other organizations around the world contribute input and feedback.

SPEC CLOUD IAAS 2016 V1.1 UPDATE RELEASED

January 03, 2017

SPEC Cloud IaaS 2016 V1.1 is now available. This update includes bug fixes and usability improvements, libcloud support for easy adapter development, and OpenStack driver updates. Licensees of version 1.0 are entitled to a complimentary upgrade. Starting January 23, 2017, all result submissions must be made using version 1.1 and version 1.0 will be retired. The benchmark is designed to stress provisioning as well as runtime aspects of a cloud using I/O and CPU intensive cloud computing workloads. SPEC selected the social media NoSQL database transaction and K-Means clustering using map/reduce as two significant and representative workload types within cloud computing.

RECEIVING THE CMG'S 2016 MICHELSON AWARD

November 24, 2016

The Computer Measurement Group (CMG) recognized Dr. André Bondi of Software Performance and Scalability Consulting LLC with the A. A. Michelson Award. Dr. Bondi was honored for his work on performance issues in several problem domains, including rail, data analytics, telecommunications, conveyor systems, web-based financial systems, building surveillance, telecommunications operations support systems, and network management.

NEW TOOL LIKWID ACCEPTED

June 02, 2016

A new tool was accepted for SPEC RG's repository of peer-reviewed tools for quantitative system evaluation and analysis: LIKWID. LIKWID is a set of command line tools supporting software developers, benchmarkers and application users to get the best performance on a given system. LIKWID is available for the Linux operations system, works with any standard Linux kernel and only depends on the GNU compiler and Perl to build. The tools support: system information and control, performance and energy profiling and micro-benchmarking.

ICPE 2017: STATISTICS

With all its different tracks, the 8th ACM/SPEC International Conference on Performance Engineering (ICPE) 2017, taking place on April 22-26 in L'Aquila, Italy, attracted a number of high–quality submissions selected in intensive review processes by different track committees.

In the research track, we had 65 submissions with 22 full and 2 short papers being accepted for presentation at the conference. In the Visions and Work-In-Progress track, 14 out of 25 contributions were selected. In the industry track, 5 papers out of 18 got accepted. In addition to those scientific papers, the program also includes three keynotes, a poster and demo track, tutorials, and several workshops.

Vittorio Cortellessa (Università dell'Aquila) and
Walter Binder (Università della Svizzera italiana)

ICPE 2018 IN BERLIN - PRELIMINARY ANNOUNCEMENT

The ACM/SPEC International Conference on Performance Engineering (ICPE) provides a forum for the integration of theory and practice in the field of performance engineering. It brings together researchers and industry practitioners to share ideas, discuss challenges, and present results of both work-in-progress and state-of-the-art research on performance engineering of software and systems.

ICPE 2018 will be held in Berlin (Germany), from April 9 to April 13. Berlin is the capital of Germany. It can be easily reached using some intercontinental flights or many connections to all major airports in Europe. The city has an intermediately long, not always happy history and the conference location will be right in the heart of the old city center. The conference will take place at the Academy of Sciences on Gendarmenmarket, very close to Brandenburg Gate, the Opera, the main building of Humboldt-University, the Museum Island and many other sites. The weather in April is often mild and sunny, so this is the perfect time to come to Berlin.

The contact person for ICPE 2018 is Katinka Wolter, who will be General Co-Chair along with William Knottenbelt from Imperial College London. The PC-Co-Chairs will be André van Hoorn from University of Stuttgart in Germany and Manoj Nambiar from Tata Consultancy Services, Mumbai, India. The industrial track chair will be Heiko Koziolek, ABB Ladenburg, Germany.

Katinka Wolter (Freie Universität Berlin)

REPORT: POWER WORKING GROUP

The Power Working Group is a new working group within SPEC RG that fosters interaction between industry and academia by contributing research that enhances and promotes methods and tools for energy efficiency evaluation. Energy efficiency of computing devices is an important concern for industry, academia, and regulatory institutions. With the proliferation of device types, new hardware components, and different workloads power and energy efficiency measurement faces new challenges each day.

Accurate and representative measurement and characterization of power consumption of computing devices requires representative workloads to be executed in a controlled and measurable fashion. Workloads must exercise a tested system's components in a manner that enables accurate power measurements. Consequently, workload placement, execution, and workload characteristics itself all require careful thought as part of the overall measurement methodology. Finally, metrics and models must be derived from measurement results, enabling system comparison and analysis.

The Power Working Group encourages research in these areas. It enables collaboration and facilitates the exchange of ideas, challenges, and approaches between members both from industry and academia. Additionally, the Power Working Group increases the visibility of research results by highlighting their usefulness both to members as well as outside observers. In the past, the Power WG has operated from within the SPEC OSG Power subcommittee and performed studies on server energy efficiency benchmarks and benchmarking, including studies on CPU workloads [1], workload distribution [2], and power variability [3].

Since its inception at the end of 2016, RG Power has been working in collaboration with OSG power towards developing a new unified energy efficiency metric for regulators based the SPEC SERT rating tool. The result of this work is a single score metric that is used by the SERT 2.0 in the new EPA sever energy efficiency standard. The RG power description of the new energy efficiency metric is included with the documentation of the SERT on the website [4] and as it is delivered to customers. It is the first RG produced white paper to be included by default in a standard SPEC kit. Some of the current members of the Power Working Group are Dell, HPE, IBM, Intel, Microsoft, and the University of Würzburg.

[1] J. v. Kistowski, H. Block, J. Beckett, K. D. Lange, J. A. Arnold, and S. Kounev: Analysis of the Influences on Server Power Consumption and Energy Efficiency for CPU-Intensive Workloads. ACM/SPEC ICPE 2015.

[2] J. von Kistowski, J. Beckett, K. D. Lange, H. Block, J. A. Arnold and S. Kounev: Energy Efficiency of Hierarchical Server Load Distribution Strategies. IEEE MASCOTS 2015, Atlanta, GA, 2015, pp. 75-84.

[3] J. v. Kistowski, H. Block, J. Beckett, C. Spradling, K. D. Lange, and S. Kounev: Variations in CPU Power Consumption. ACM/SPEC ICPE 2016. ACM, 147-158.

[4] SPEC RG Power. The SERT Metric and the Impact of Server Configuration. https://www.spec.org/sert2/SERT-metric-20170314.pdf

Jóakim von Kistowski (University of Würzburg)
https://research.spec.org/working-groups/power-working-group.html

REPORT: IDS WORKING GROUP

The SPEC RG IDS Benchmarking Working Group successfully concluded its agenda for 2016 and faces 2017 with a renewed commitment.

The SPEC RG IDS Benchmarking Working Group published a paper on IDS evaluation metrics at IEEE ISSRE 2016 (the 27th International Symposium on Software Reliability Engineering) [1]. This paper demonstrates the impact of elasticity on IDS attack detection accuracy. In addition, it proposes a novel metric and measurement methodology for accurately quantifying the accuracy of IDSes deployed in virtualized environments featuring elasticity.

In addition, the Group's Secretary, Aleksandar Milenkoski, successfully defended his doctoral dissertation entitled "Evaluation of Intrusion Detection Systems in Virtualized Environments". This dissertation summarizes the major contributions of the SPEC RG IDS Benchmarking Working Group.

[1] A. Milenkoski, K. R. Jayaram, N. Antunes, M. Vieira, and S. Kounev: Quantifying the Attack Detection Accuracy of Intrusion Detection Systems in Virtualized Environments. IEEE International Symposium on Software Reliability Engineering (ISSRE 2016), Ottawa, Canada, 2016, IEEE.

Aleksandar Milenkoski (ERNW Heidelberg)
https://research.spec.org/working-groups/rg-ids-benchmarking.html

REPORT: CLOUD WORKING GROUP

In 2016, the SPEC RG Cloud Group has pursued through a broad palette of activities its long-term mission of furthering cloud benchmarking, quantitative evaluation, and experimental analysis, in directions relevant for both academia and industry. Through our mission, we focus on novel cloud properties such as elasticity, performance isolation, dependability, and other non-functional system properties, in addition to classical performance-related metrics such as response time, throughput, scalability, and efficiency.

The scope of the group is 'to develop new methodological elements for gaining deeper understanding not only of cloud performance, but also of cloud operation and behavior, through diverse quantitative evaluation tools, including benchmarks, metrics, and workload generators'.

Among the keywords most discussed in the group, over 2016, are cloud metrics, elasticity, auto-scaling, service definition and configuration, workload characterization and modeling, and infrastructure, platform, and data as a service.

In 2016, through bi-weekly online meetings facilitated by WebEx and SPEC, we have advanced work on 5 main topics and a number of other sub-topics:

1) Cloud Usage Patterns (CUPs) [1]: The goal of this activity is to define a formalism for expressing cloud usage patterns and scenarios, which can be used by both general users and cloud experts. The proposed textual and visual formalism focuses on conciseness; this goes in contrast to machine-ready specifications. We have used CUPs to express over ten patterns commonly seen in academic and industrial practice. An extended article, with new core concepts, is currently under submission.
2) Cloud Metrics Survey and Design [2]: This ongoing activity focuses on surveying existing cloud metrics and on the design of key missing metrics that allow the quantitative assessment and characterization of typical cloud usage scenarios. Among the key new metrics we have introduced in 2016, the joint authors focus on various forms of elasticity and risk-quantifying metrics.
3) Benchmarking Autoscaling Techniques [3]: This ongoing joint activity focuses on a quantitative analysis and comparison of auto-scaling techniques in virtualized environments, for various application domains. So far, we have created a performance evaluation process and conducted real-world experiments with state-of-the-art autoscalers, focusing on workloads of scientific and engineering workflows, and on web site workloads.
4) Artifacts for Performance Measurement and Benchmarking: During 2016, the group has been involved in the cooperative development of many open-source software and open-access data artifacts, including LIMBO, BUNGEE and other elasticity-measuring tools, and Graphalytics. These tools have been crucial in R&D activities focusing on resource management and scheduling in clouds and in particular data centers, with important results in novel autoscaling techniques, risk assessment and management in data centers, model-driven architecture design, etc. For more details about these artifacts, see the Software and Data Artifacts article in this newsletter.
5) Reach-out and dissemination: our group has been active in disseminating its knowledge and artifacts to industry and academia. For example, Nikolas Herbst et al. give a tutorial on evaluating autoscaling techniques at ICPE 2017, and group artifacts have been well represented in higher education courses: at TUD, IN4392 Cloud Computing course, in the Netherlands, the ASCI A24 doctoral course on Introduction to Programming HPC Systems, etc. We have also fostered communication and dissemination, for example, by co-organizing in Delft the flagship conference of SPEC, the ACM/SPEC International Conference on Performance Engineering (ICPE 2016).

The group is happy to report a number of special achievements: dr. ir. Alexandru Iosup has been appointed Full (Tenured) Professor at the Vrije Universiteit Amsterdam, the Netherlands, has received the annual prize ICT Researcher of the Year in the Netherlands, and has been selected and appointed Member of the Young Royal Netherlands Academy of Arts and Sciences. Dr. Aleksandar Milenkoski received his Ph.D. for his thesis Evaluation of Intrusion Detection Systems in Virtualized Environments, from the University of Würzburg, Germany, and has moved to a corresponding role at ERNW, Heidelberg. The group has also led to the completion of numerous M.Sc. and B.Sc. projects.

Current participants in the RG Cloud Group include the Delft University of Technology (Delft), Lund University (Sweden), University of South Florida (FL, USA), Umea University (Sweden), and the University of Würzburg (Germany), on the academic side, and the IBM T.J. Watson Research Center (USA), MITRE (USA), Oracle (USA), Salesforce.com (USA), SAP (Germany), Tata TCS (India), on the industry side and intermittently. Besides the frequent participants, we have welcomed various invited speakers and guest participants. To conclude, 2016 was a full and successful year for the RG Cloud Group. We are looking forward to an even more successful 2017. For this, we are actively seeking new participants and activities.

[1] A. Milenkoski, A. Iosup, S. Kounev, K. Sachs, P. Rygielski, J. Ding, W. Cirne, and F. Rosenberg: Cloud usage patterns: A formalism for description of cloud usage scenarios. CoRR, abs/1410.1159, 2014.

[2] N. Herbst, R. Krebs, G. Oikonomou, G. Kousiouris, A. Evangelinou, A. Iosup, and S. Kounev: Ready for Rain? A View from SPEC Research on the Future of Cloud Metrics. CoRR. abs/1604.03470, 2016

[3] A. Ilyushkin, A. Ali-Eldin, N. Herbst, A. V. Papadopoulos, B. Ghit, D. Epema, A. Iosup: An Experimental Performance Evaluation of Autoscaling Policies for Complex Workflows, ACM/SPEC ICPE 2017

Alexandru Iosup (TU Delft)
http://research.spec.org/working-groups/rg-cloud.html

REPORT: BIG DATA WORKING GROUP

The research field of big data is evolving at a rapid pace utilizing a diverse spectrum of technical platforms and serving a wide range of applications. The combinations of large volumes of data, heterogeneous data formats, and the rapidly improving performance of both hardware and big data systems, makes it difficult to generalize architectural aspects that best suit all application requirements. This makes the investigation and standardization of such systems very challenging.

The big data working group in SPEC Research facilitates research and engages industry leaders for defining and developing performance methodologies of big data applications. The group fosters the close collaboration with industry because over the last two years big data has made its way from a mostly academic discipline to the radar screen of established data processing companies. Teradata, IBM, Oracle (among others) have started to incorporate the ideas of big data into their “main-stream” products. As a result the term “big data” has become a major force of innovation across enterprises of all sizes. Especially, the integration of Structured Query Language (SQL) into big data has propelled the adaption of big data in the industry.

As these big data systems are evolving, there is an inherent need to evaluate and quantify their performance with the ultimate goal of comparing these systems. Comparisons are desirable in different dimensions, such as software stack, hardware, use case, and tuning parameters. That is, one might want to compare a particular software stack on different hardware systems, a particular hardware setting on different software stacks, or one software stack on a particular hardware with different tunings.

With the rapid increase in big data solutions, both academia and industry alike are developing new benchmarks at a rapid pace. Driven by the ”velocity of change” many performance benchmark developers ”cut corners” by customizing their benchmarks too closely to the architectural characteristic of the system they want to benchmark, instead of abstracting its core performance attributes. These benchmarks become ”island solutions” that only fit the systems they targeted in the first place. This approach works well if the goal is to compare the performance of a particular software stack on a particular hardware setting. However, this does not work well to compare the performance of different software stacks on the same hardware platforms or vice versa.

The big data working group is, however, looking beyond the mainstream use of big data in the industry. Companies such as Intel and IBM are heavily investing in the development of deep learning, which resemble algorithms to emulate the human brain’s decision-making process. The hope of deep learning is to be able to solve complex problems, be able to answer questions that are not anticipated and ultimately aid humans from day-to-day repetitive tasks. We are working on ways to include deep learning aspects into big data benchmarks.

The big data working holds weekly meetings, with internal calls and open, public calls alternating. For the open calls, presentations on big data benchmarking, big data systems, performance tuning, and related research are invited. These calls are well attended and received. Topics in 2016 were:

- Sweta Singh, IBM, Benchmarking Spark Machine Learning using BigBench
- Jeremy Arnold, IBM, What is a good benchmark?
- Josep Berral, BSC, Brief Overview of Machine Learning towards Benchmarking
- Thomas P. Oberst: Big Data and Health Care
- Meikel Poess: TPC-DS Version 2
- Todor Ivanov: TPCx-HS Experiments
- Manoj Nambiar: Auto-tuning Hadoop Map Reduce
- Boris Zibitsker: Performance Assurance for Big Data Applications
- Paul Cao: The Experience On Running First Industry Big Data Benchmark TPCx-BB
- Álvaro Villalba: Multi-tenant Pub/Sub Processing for IoT Data Streams

Meikel Poess (Oracle)

https://research.spec.org/en/working-groups/rg-big-data.html

REPORT: DEVOPS PERFORMANCE WORKING GROUP

The DevOps Performance Working Group is concerned with questions on how to address performance concerns in DevOps. DevOps aims to increase the velocity and frequency of bringing software changes into stable production. This is achieved by a tighter integration of development (Dev) and operations (Ops), as well as a high degree of automation (e.g., via continuous delivery/deployment).

The DevOps Performance Working Group fosters and facilitates research in combining model-based and measurement-based APM (application performance management) and SPE (software performance engineering) activities for DevOps, e.g., by experience sharing, agreement on definitions, specification of metrics, and dissemination of novel methods, techniques, and tools for quantitative evaluation. We aim to identify cross-community collaboration, and to set the path for long-lasting collaborations towards performance-aware DevOps.

As the major joint activities in 2016, the DevOps Working Group has been working on the development of a reference platform [1] and an industry survey [3] on performance-aware DevOps. A talk on "Rethinking Performance Engineering in the DevOps World" was given at DevOpsDays in Kiel, Germany in May 2016. Both the work on the platform and the survey will be continued in 2017. Several sub-projects among subsets of the members have been conducted in 2016. Selected technical efforts include the OPEN.xtrace open-source format and tool support to interchange executions traces between (commercial and open-source) APM and SPE tools [2, 4], and the new collaborative research project Declare on "Declarative Performance Engineering" [5].

In July 2016, the group met for its second face-to-face meeting, co-located with the 13th IEEE International Conference on Autonomic Computing (ICAC 2016) in Würzburg, Germany. The agenda of the one-day meeting comprised discussions on the blueprint and the survey, three technical talks, a demo, and discussions on future working group activities. 13 group members participated in the meeting.

Together with two EU projects, the working group organized the second edition of the International Workshop on Performance-Aware DevOps (QUDOS 2016) [6]. QUDOS 2016, technically co-sponsored by the SPEC RG, was co-located with the renowned International Symposium on Software Testing and Analysis 2016 (ISSTA '16) in Saarbrücken, Germany in July 2016. The workshop comprised 15 talks (including an industrial keynote), 12 scientific papers, and around 25 registered participants. The third QUDOS edition [7] will be co-located with the 8th ACM/SPEC International Conference on Performance Engineering (ICPE 2017) in L'Aquila, Italy.

In September 2016, the group co-organized a GI-Dagstuhl seminar entitled "Software Performance Engineering in the DevOps World" [8]. The seminar brought together 26 international researchers in the areas of software engineering, performance engineering, cloud computing, and big data to present their current research projects, to exchange experience and expertise, to discuss research challenges, and to develop ideas for future collaborations.

The group meets on a biweekly basis to discuss current activities and new project ideas, including invited presentations. In 2016, 19 meetings were held, including talks about declarative performance engineering, migration to and adoption of microservice architectures, configuration optimization, prioritization of load tests, and joint projects of the group. A number of new members and member organizations joined the group.

For more information about the DevOps Performance Working Group (including mission, activities, meetings, presentations, projects), visit our web page. If you are interested in following the discussions or contributing actively, please get in touch with the working group chairs.

[1] T. F. Düllmann, R. Heinrich, A. van Hoorn, T. Pitakrat, J. Walter, and F. Willnecker: CASPA: A platform for comparability of architecture-based software performance engineering approaches. IEEE International Conference on Software Architecture (ICSA 2017).

[2] SPEC RG DevOps GitHub projects https://github.com/spec-rgdevops

[3] Survey on performance-aware DevOps http://invite.devops-performance.org/survey/?q=SPEC

[4] D. Okanović, A. van Hoorn, C. Heger, A. Wert, and S. Siegl: Towards Performance Tooling Interoperability: An Open Format for Representing Execution Traces. EPEW 2016.

[5] J. Walter, A. van Hoorn, H. Koziolek, D. Okanović and S. Kounev: Asking "What?", Automating the "How?": The Vision of Declarative Performance Engineering. ACM/SPEC ICPE '16.

[6] 2nd International Workshop on Quality-aware DevOps (QUDOS 2015). http://qudos2015.fortiss.org/

[7] 3rd International Workshop on Quality-aware DevOps (QUDOS 2016). http://qudos2016.fortiss.org/

[8] GI-Dagstuhl Seminar 16394, Software Performance Engineering in the DevOps World. http://www.dagstuhl.de/16394

André van Hoorn (University of Stuttgart), Felix Willnecker (fortiss GmbH), Robert Heinrich (Karlsruhe Institute of Technology)

https://research.spec.org/en/working-groups/rg-devops-performance.html

SPEC KAIVALYA DIXIT DISTINGUISHED DISSERTATION AWARD 2016

December 01, 2016

The SPEC Kaivalya Dixit Distinguished Dissertation Award is an annual award that aims to recognize outstanding doctoral dissertations within the scope of the SPEC Research Group in terms of scientific originality, scientific significance, practical relevance, impact, and presentation. This year, the award committee considered 13 excellent dissertation submissions from universities around the world - Umea University, Technische Universitat München, Queen's University (Canada), Imperial College London, Sapienza Universita di Roma, Bogazici University, Universidade de Lisboa, Charles University in Prague, University College Dublin, University of Waterloo, Ecole Polytechnique Federale de Lausanne, College of William and Mary and University of California, Berkeley.

The winning dissertation was authored by Scott Beamer III of University of California, Berkeley, under the supervision of Profs. Krste Asanovic and David Patterson. The selection committee appreciates the characterization and treatment of fundamental performance bottlenecks in graph processing and the contribution of the Graph Algorithm Platform (GAP) benchmark suite. The results of this dissertation are already in use and are likely to continue to produce contributions to the area of graph processing in the future.

The award is to be presented at the International Conference on Performance Engineering (ICPE) in Italy in April 2017.

Given the high quality of dissertations nominated for this award, the committee decided to publicly recognize the dissertation titled "Understanding and Efficiently Servicing HTTP Streaming Video Workloads" authored by Jim Summers under the supervision of professor Tim Brecht at the University of Waterloo as the runner-up to the award. The committee appreciates the contribution of a methodology, a benchmark, and a performance study for streaming video applications contributed by this dissertation.

The award selection committee for 2016 was chaired by Prof. J. Nelson Amaral of the University of Alberta and consisted of eleven members from academia and industry.

The SPEC Kaivalya Dixit Distinguished Dissertation Award was established in 2011 to recognize outstanding dissertations within the scope of the SPEC Research Group. Contributions of interest 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. Dissertations defended between October 2016 and September 2017 will be eligible to be nominated for the 2017 award.

https://research.spec.org/news/single-view/article/winner-of-spec-kaivalya-dixit-distinguished-dissertation-award-2016.html

KAIVALYA DIXIT DIST. DISS AWARD 2016: PHD THESIS ABSTRACT

Graph processing is experiencing a surge of renewed interest as applications in social networks and their analysis have grown in importance. Additionally, graph algorithms have found new applications in speech recognition and the sciences. In order to deliver the full potential of these emerging applications, graph processing must become substantially more efficient, as graph processing's communication-intensive nature often results in low arithmetic intensity that underutilizes available hardware platforms.

To improve graph algorithm performance, this dissertation characterizes graph processing workloads on shared memory multiprocessors in order to understand graph algorithm performance. By querying performance counters to measure utilizations on real hardware, we find that contrary to prevailing wisdom, caches provide great benefit for graph processing and the systems are rarely memory bandwidth bound. Leveraging the insights of our workload characterization, we introduce the Graph Algorithm Iron Law (GAIL), a simple performance model that allows for reasoning about tradeoffs across layers by considering algorithmic efficiency, cache locality, and memory bandwidth utilization. We also provide the Graph Algorithm Platform (GAP) Benchmark Suite to help the community improve graph processing evaluations through standardization.

In addition to understanding graph algorithm performance, we make contributions to improve graph algorithm performance. We present our direction-optimizing breadth-first search algorithm that is advantageous for low-diameter graphs, which are becoming increasingly relevant as social network analysis becomes more prevalent. Finally, we introduce propagation blocking, a technique to reduce memory communication on cache-based systems by blocking graph computations in order to improve spatial locality.

Scott Beamer III
http://www.scottbeamer.net/pubs/beamer-thesis.pdf

SOFTWARE AND DATA ARTIFACTS - CLOUD WORKING GROUP

1) Benchmarking software:

BUNGEE elasticity benchmarking framework: This ongoing activity focuses on the development and validation of a Java-based framework focusing on cloud elasticity, especially for IaaS cloud platforms and auto-scaling environments. The tool provides load-testing functionality, and automates the analysis of the quality of the elastic behavior of the system under test through several elasticity metrics and aggregated scores. The current set of metrics is described in detail in the respective articles to the selected abstracts below. Load-varying workloads can flexibly be defined manually or based on traces using the LIMBO framework.

LDBC Graphalytics is a comprehensive benchmarking suite comprised of diverse algorithms and datasets. It has been used so far to test single-node platforms such as Oracle PGX, Intel GraphMat, and Neo4j, distributed platforms such as Oracle PGX.D, Giraph, and Spark/GraphX, and GPU-based platforms such as Totem, MapGraph, and Medusa. The results have been made public, and are available online and in published peer-reviewed material. For more details, see the article on LDBC Graphalytics in this newsletter.

2) Data artifacts: The SPEC RG Cloud Group publishes various data artifacts of interest to our community. Examples of artifacts released as open-access data in 2016 are:

Datacenter trace #1, courtesy Solvinity, NL, processed and released by TU Delft, NL: This dataset contains the performance metrics of 1,750 VMs from a distributed datacenter from Bitbrains, which is a service provider that specializes in managed hosting and business computation for enterprises. Customers include many major banks (ING), credit card operators (ICS), insurers (Aegon), etc. Bitbrains hosts applications used in the solvency domain; examples of application vendors are Towers Watson and Algorithmics. These applications are typically used for financial reporting, which is used predominantly at the end of financial quarters.

Datacenter trace #2, courtesy of Materna, DE, released jointly with TU Delft, NL: The three datasets contain the performance metrics of 520, 527 and 547 VMs from a distributed datacenter of Materna. Materna is a full service provider in the premium segment and has been successfully implementing ITC projects for their customers for more than 35 years. Their client list reads like the “Who’s Who” of German companies and public sector organisations.

Alexandru Iosup (TU Delft), Nikolas Herbst (Uni Würzburg)

LDBC GRAPHALYTICS: BENCHMARKING PLATFORMS FOR LARGE-SCALE GRAPH PROCESSING

Graphs are abstract data models that can represent diverse human knowledge and processes: they model social networks, human knowledge, and other vital information for business, governance, and academic practice. Although both industry and academia are developing and tuning many graph-processing algorithms and platforms, the performance of graph-processing platforms has never been explored or compared in-depth. Moreover, graph processing exposes new bottlenecks in traditional HPC, distributed, multi-/many-core, and multi-threaded systems (as exemplified by the major differences in Top500 and Graph500 rankings).

Prof. dr. Alexandru Iosup is leading the Graphalytics team focusing on benchmarking graph-processing platforms through a comprehensive benchmarking suite comprised of diverse algorithms and datasets. The team spans several universities in Europe and industry partners Oracle, Intel, IBM, and Huawei; more partners are advising through the Linked Data Benchmark Council (LDBC). Key scientific and practical results from this collaboration are the design of the standard Graphalytics benchmark and the development and release of the open-source tools for it [1,2], enabling its users to explore and to explain the performance dependency Platform-Algorithm-Dataset [3]. In 2016, the team has also made advances in fine-grained tracing and iterative profiling of graph-processing platforms (the Granula module of Graphalytics) and is preparing a tool for bottleneck identification and analysis (Grade10). We are currently extending Graphalytics with elasticity features, so far tested through our distributed and elastic graph-processing platform JoyGraph [4].

Open-sourced since 2015, Graphalytics is now mature software. We have presented the methods, tools, and core results also as tutorials, from preliminary features to mature tools and optimization processes. All the introductory and training material is available online.

During 2017, LDBC Graphalytics will enable a global competition of graph-processing platforms. To this end, we invite the SPEC Newsletter audience to try Graphalytics and to contact our team, at A.Iosup@vu.nl.

This body of work has benefited from a generous gift from SPEC-member Oracle and from Netherlands NWO and STW sponsorship, and from advice and in-kind contribution from LDBC and SPEC members such as Oracle, Intel, IBM, Huawei, and Facebook.

[1] More information about Graphalytics: http://graphalytics.org

[2] Code on GitHub: https://github.com/ldbc/ldbc_graphalytics

[3] A. Iosup, T. Hegeman, W. L. Ngai, S. Heldens, A. Prat-Pérez, T. Manhardt, H. Chafi, M. Capota, N. Sundaram, M. J. Anderson, I. G. Tanase, Y. Xia, L. Nai, and A. Boncz: LDBC Graphalytics: A Benchmark for Large-Scale Graph Analysis on Parallel and Distributed Platforms. PVLDB 9(13): 1317-1328 (2016)

[4] Sietse Au, JoyGraph: A Distributed and Elastic Graph Processing System, M.Sc. Thesis TU Delft supervised by Alexandru Iosup, 2017.

Alexandru Iosup (TU Delft)

NEWS AND ADVANCEMENTS IN INSPECTIT

January 20, 2017

In the SPEC RG Newsletter published in March 2016 (Volume 2, Issue 1) the open source APM tool inspectIT has been presented which became a part of the SPEC RG Tool Repository in May of the same year. At that time, the capabilities of inspectIT covered the gathering and storing of runtime data from instrumented Java applications and the subsequent analysis of it.

Since then, a lot of development effort has gone into the tool which is reflected in the features and improvements that have been integrated. Thus, for example, in order to simplify working with multiple inspectIT installations, it is now possible to export and import instrumentation configurations. Furthermore, REST interfaces have been integrated which enable directly accessing collected data stored in the Central Measurement Repository (CMR) without needing the inspectIT UI.

Another new feature is the possibility to use inspectIT as a monitoring solution. It provides functionality to continuously monitor a system and visualize the gathered data using dashboards. Furthermore, it is possible to issue alerts if the monitored system violates predefined thresholds. In order to filter and categorize the collected data the definitions of applications and business transaction have been implemented. We are also proud to announce that inspectIT is the first open source APM software which is able to change the used instrumentation on-the-fly without having to restart the instrumented application.

In the last year, the license of inspectIT changed from “GNU AGPL v3” to “Apache License version 2.0” and there have also been many experimental features in the area of inspectIT. For example, agents have been developed which enable, firstly, the collection of data from Android mobile applications. Secondly, data from applications based on the .NET Framework can be gathered using a sampling-based C# agent. Besides the development of new agents, a generalized process to generate performance models from monitoring data which is based on approaches like iObserve or Wessbas has been prototypical integrated.

Currently, there are major features which will shortly be completed and greatly increase the possibilities inspectIT can be used for. On the one hand, a JavaScript browser agent is under development which allows end-user monitoring. By doing this, detailed data about the end-user experience can be gathered. In combination with the ability to combine and relate traces of multiple platforms, which is the second feature under development, the behavior of a complete system can be comprehend (distributed tracing).

In spite of the progress achieved, there are still many open features and tasks on the future roadmap of inspectIT. Two things which are particularly worth mentioning here are the integration and support of the tracing-system OpenTracing and the implementation of an anomaly detection feature including a fully automated detection of a system behavior and, subsequently, calculation of baselines and thresholds.

More information about inspectIT and a downloadable version can be obtained at the inspectIT website

Marius Oehler (NovaTec Consulting GmbH)
http://www.inspectit.rocks or https://github.com/inspectIT/inspectIT

SELECTED ABSTRACTS

An Experimental Performance Evaluation of Autoscaling Algorithms for Complex Workflows:

Simplifying the task of resource management and scheduling for customers, while still delivering complex Quality-of-Service (QoS), is key to cloud computing. Many autoscaling policies have been proposed in the past decade to decide on behalf of cloud customers when and how to use the elastic features of clouds. However, in prior work many of these policies are not compared to each other, and instead are often compared only to static provisioning or to a predefined QoS target. This reduces the ability of cloud customers and of cloud operators to choose and deploy a suitable autoscaling policy.

In our work, we conduct an experimental performance evaluation of autoscaling policies, using as application model workflows, a commonly used formalism for automating resource management for applications with well-defined yet complex structure. We present a detailed comparative study

of general, state-of-the-art, generic autoscaling policies, along with two new workflow-specific auto-scalers. To understand which policy is the best, we also conduct various forms of pairwise and group comparisons, and report both individual and aggregated metrics

A. Ilyushkin, A. Ali-Eldin, N. Herbst, A. V. Papadopoulos, B. Ghit, D. Epema, A. Iosup: An Experimental Performance Evaluation of Autoscaling Policies for Complex Workflows. ACM/SPEC ICPE 2017, ACM, PDF.

Ready for Rain? A View from SPEC Research on the Future of Cloud Metrics:

In the past decade, cloud computing has emerged from a pursuit for a service-driven information and communication technology (ICT), into a significant fraction of the ICT market. Responding to the growth of the market, many alternative cloud services and their underlying systems are currently vying for the attention of cloud users and providers. Thus, benchmarking them is needed, to enable cloud users to make an informed choice, and to enable system DevOps to tune, design, and evaluate their systems. This requires focusing on old and new system properties, possibly leading to the re-design of classic benchmarking metrics, such as expressing performance as throughput and latency (response time), and the design of new, cloud-specific metrics. Addressing this requirement, in this work we focus on four system properties:

(i) Elasticity
(ii) Performance Isolation
(iii) Availability, and
(iv) Operational Risk.

Focusing on key metrics, for each of these properties we review the state-of-the-art, then select or propose new metrics together with measurement approaches. We see the presented metrics as a foundation towards upcoming, industry-standard, cloud benchmarks.

N. Herbst, R. Krebs, G. Oikonomou, G. Kousiouris, A. Evangelinou, A. Iosup, S. Kounev. Ready for Rain? A View from SPEC Research on the Future of Cloud Metrics. Technical Report SPEC-RG-2016-01, SPEC Research Group - Cloud Working Group, CoRR, abs/1604.03470, 2016.

CASPA: A Platform for Comparability of Architecture-Based Software Performance Engineering Approaches:

Setting up an experimental evaluation for architecture-based Software Performance Engineering (SPE) approaches requires enormous efforts. This includes the selection and installation of representative applications, usage profiles, supporting tools, infrastructures, etc. Quantitative comparisons with related approaches are hardly possible due to limited repeatability of previous experiments by other researchers.

This paper presents CASPA, a ready-to-use and extensible evaluation platform that already includes example applications and state- of-the-art SPE components, such as monitoring and model extraction. The platform explicitly provides interfaces to replace applications and components by custom(-ized) components. The platform builds on state-of-the-art technologies such as container-based virtualization.

T. F. Düllmann, R. Heinrich, A. van Hoorn, T. Pitakrat, J. Walter, F. Willnecker. CASPA: A platform for comparability of architecture-based software performance engineering approaches. IEEE International Conference on Software Architecture (ICSA 2017).

Towards Performance Tooling Interoperability: An Open Format for Representing Execution Traces:

Execution traces capture information on a software system's runtime behavior, including data on system-internal software control flows, performance, as well as request parameters and values. In research and industrial practice, execution traces serve as an important basis for model-based and measurement-based performance evaluation, e.g., for application performance monitoring (APM), extraction of descriptive and prescriptive models, as well as problem detection and diagnosis. A number of commercial and open-source APM tools that allow the capturing of execution traces within distributed software systems is available. However, each of the tools uses its own (proprietary) format, which means that each approach building on execution trace data is tool-specific.

In this paper, we propose the Open Execution Trace Exchange (OPEN.xtrace) format to enable data interoperability and exchange between APM tools and software performance engineering (SPE) approaches. Particularly, this enables SPE researchers to develop their approaches in a tool-agnostic and comparable manner. OPEN.xtrace is a community effort as part of the overall goal to increase interoperability of SPE/APM techniques and tools. In addition to describing the OPEN.xtrace format and its tooling support, we evaluate OPEN.xtrace by comparing its modeling capabilities with the information that is available in leading APM tools.

D. Okanović, A. van Hoorn, C. Heger, A. Wert, and S. Siegl: Towards performance tooling interoperability: An open format for representing execution traces. EPEW 2016, LNCS, Springer, 2016.