

**spec**<sup>®</sup>

**SPEC 2016亚洲峰会**  
**SPEC 2016 ASIA SUMMIT**

# **SPEC Research Group and Big Data**

**Xiao Wei Zhang, Liang Lu**

**(IBM / SPEC RG Big Data WG)**

# Mission and Objectives



“The mission of the SPEC RG Big Data Working Group is to facilitate research in performance methodologies for benchmarking Big Data systems by engaging industry and academia in these efforts.”

## Objectives:

- Defining clear goals for the aspects of a Big Data system that are important to measure
- Developing sound rules and metrics to measure the performance of Big Data systems
- Fostering collaboration among related benchmarking efforts

## Big Data Benchmarking Community (BDBC)

- Formed in 2012
- Original founders:
  - Chaitan Baru (San Diego Supercomputer Center)
  - Tilmann Rabl (University of Toronto)
  - Milind Bhandarkar (Greenplum)
  - Raghu Nambiar (Cisco)
  - Meikel Poess (Oracle)
- Website: <http://clds.ucsd.edu/bdbc/>
- Introduction at the Extremely Large Database Conference (XLDB), September, 2012
- Workshops

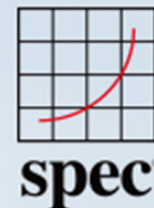
## Workshops

- 7 “Workshops on Big Data Benchmarking” (WBDB for short) held since 2012
- NSF (US National Science Foundation) and industry sponsored

## Past BDBC workshops

- 1st workshop, May 2012, San Jose, USA. Hosted by Brocade.
- 2nd workshop, Dec 2012, Pune, India. Hosted by Persistent Systems/Infosys.
- 3rd workshop, July 2013, Xi’an, China. Hosted by Xi’an University.
- 4th workshop, October 2013, San Jose, USA. Hosted by Brocade.
- 5th workshop, August 2014, Potsdam, Germany. Hosted by HPI.
- 6th workshop, June 16-17 2015, Toronto, Canada.
- 7th workshop, December 14-15, 2015, Delhi, India.

# Publications



Papers published in SIGMOD, TPCTC, Big Data Journal by the BDBC and SPEC Research Big Data Working Group

Examples of specific publications:

- Specifying Big Data Benchmarks. Volume 8163 of LNCS. Springer Berlin Heidelberg, 2014.
- Advancing Big Data Benchmarks. Volume 8585 of LNCS. Springer Berlin Heidelberg, 2014.
- Big Data Benchmark Compendium, 7th TPC Technology Conference (TPCTC), 2015.
  - Big Data survey style paper including SPEC RG vision on Big Data

# History and Events



SPEC Research Big Data Working Group formed in October, 2014

Bi-weekly public calls with presentations on Big Data topics

- Presentations available here: <https://research.spec.org/working-groups/big-data-working-group/meetings-and-presentations.html>

Article introducing the group in SPEC RG News Letter (Jan 30, 2015)

Poster promoting the group at International Conference on Performance Engineering (ICPE), February 2, 2015

Lightening talk and poster at Extremely Large Database Conference (XLDB), May, 2015

# Current and Future Work



Create new SPEC Big Data benchmark(s)

- Fill gaps

Benchmark areas under discussion include:

- Internet of Things (IoT)
- Deep Learning/Machine Learning

Development approach: from scratch versus reuse

# Impact of Cloud



## Cloud brings new challenges to performance benchmarking

- Noisy neighbor
- Less control and access to the execution environment



# Cloud Benchmark from SPEC



SPEC released a new Cloud “Infrastructure as a Service” (IaaS) benchmark in May, 2016

Oriented for Hadoop

Three primary metrics measure:

- Scalability
- Elasticity
- Cluster provisioning time

For more information:

[www.spec.org/cloud\\_iaas2016/](http://www.spec.org/cloud_iaas2016/)

# How to get involved ...



## Bi-weekly open calls (anyone can join)

- Mondays at 8:30 AM US Pacific time
- +1-650-429-3300, access code: 632 691 106
- Upcoming talks:
  - Nov 7: “What is a good benchmark” by Jeremy Arnold
- Mailing list: [bigdata@spec.org](mailto:bigdata@spec.org)
- Contact Tilmann Rabl if interested in scheduling your talk



## Working group contacts:

- Chair: Tilmann Rabl ([tilmann.rabl@bankmark.de](mailto:tilmann.rabl@bankmark.de))
- Co-Chair: Todor Ivanov ([todor@dbis.cs.uni-frankfurt.de](mailto:todor@dbis.cs.uni-frankfurt.de))
- Secretary: Meikel Poess ([meikel.poess@oracle.com](mailto:meikel.poess@oracle.com))
- Release manager: John Poelman ([poelman@us.ibm.com](mailto:poelman@us.ibm.com))

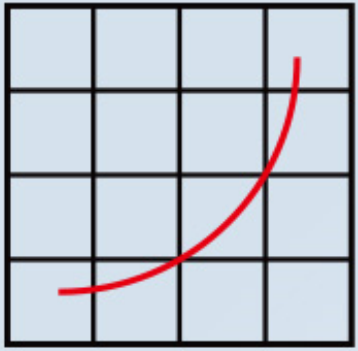
# Next Steps



Please join us tomorrow for SPEC Round Table discussions

Sessions especially relevant for Big Data:

- Cloud Computing Trends
- Big Data Trends

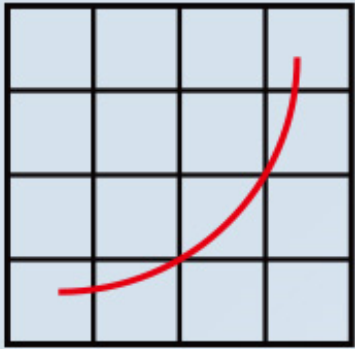


**spec**<sup>®</sup>

**SPEC 2016亚洲峰会**  
**SPEC 2016 ASIA SUMMIT**

**Q&A**





**spec**<sup>®</sup>

**SPEC 2016亚洲峰会**  
**SPEC 2016 ASIA SUMMIT**

# Thank you!

[info@spec.org](mailto:info@spec.org)

[www.spec.org](http://www.spec.org)