



Introducing the SPEC RG Big Data Working Group

May 19, 2015

XLDB 2015



Mission and Objectives



Mission Statement: “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

Origins of the SPEC RG Big Data Working Group

Big Data Benchmarking Community (BDBC) formed in 2012

- Original founders: Chaitan Baru (SDSC), Tilmann Rabl (University of Toronto), Milind Bhandarkar (Greenplum), Raghu Nambiar (Cisco), Meikel Poess (Oracle)
- Regular conference calls for talks and announcements
- Open to anyone interested, free of charge
- Website: <http://clds.ucsd.edu/bdbc/>
- Introduction at XLDB 2012

Workshops

- 5 “Workshops on Big Data Benchmarking” (WBDB for short) held between 2012-2014

Publications

- Published papers in SIGMOD, TPCTC, Big Data Journal
- 2 Springer publications (+1 in progress)

Formed in October, 2014

Weekly calls including several presentations:

- Daniel McClary: Oracle Big Data SQL: Architectural Deep Dive
- Min Li: SparkBench
- Thomas Oberst: Big Data Use Cases in Finance
- Martin Arlitt: IoTABench: an Internet of Things Benchmark
- Tilmann Rabl: BigBench Overview
- Nicolas Poggi: ALOJA - Cost Effectiveness Study of Hadoop Deployments
- Anna Queralt: Data Sharing in the Big Data Era
- John Poelman: IBM Big SQL 3.0: An Introduction
- André Pansani: Future SOC Lab – Overview

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

Poster promoting the group at ICPE (February 2, 2015)

Lightening talk and poster at XLDB 2015

Help organize two “Workshops on Big Data Benchmarking”

- 6th workshop: June 16-17, 2015, Toronto, Canada
- 7th workshop: December 14-15, 2015, Delhi, India

Publications/ Conference Participation

- Big Data survey style paper including SPEC RG vision on Big Data

Refining ideas introduced in BigBench for a SPEC Big Data benchmark

How to get involved...



Bi-weekly open calls (anyone can join)

- +1-650-429-3300, access code: 638 028 818
- Next public talk is Wednesday, May 27 at 10:00 am PDT
 - “The Graphalytics Benchmark for Graph-Processing Platforms” by Mihai Capota (Delft University of Technology)
- Mailing list: bigdata@spec.org
- Contact Tilmann Rabl if interested in scheduling your talk

Working group contacts:

- Chair: Tilmann Rabl (tilmann.rabl@bankmark.de)
- Co-Chair: Chaitan Baru (baru@sdsc.edu)
- Secretary: Meikel Poess (meikel.poess@oracle.com)
- Release manager: John Poelman (poelman@us.ibm.com)



Backup Slides

Workshops on Big Data Benchmarking (WBDB)



Launched in 2012

NSF and industry sponsored

Past 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.

Upcoming workshops

- 6th workshop, June 16-17 2015, Toronto, Canada.
- 7th workshop, December 14-15, 2015 Delhi, India.

Revised Selected Papers in Springer's LNCS

- Specifying Big Data Benchmarks. Volume 8163 of LNCS. Springer Berlin Heidelberg, 2014.
- Advancing Big Data Benchmarks. Volume 8585 of LNCS. Springer Berlin Heidelberg, 2014.