

# Introducing the SPEC Research Group on Big Data

## Mission

The mission of the Research Group on Big Data is to facilitate research in performance methodologies for benchmarking Big Data systems by engaging industry and academia in these efforts. Big Data has become a major force of innovation across enterprises of all sizes. New architectures claiming to support Big Data are being announced almost on a weekly basis with increasingly more features for managing big datasets. Yet, there is currently no good definition for what constitutes a Big Data system, nor any means for fair comparisons among such systems.

## Topics of Interest

- Differentiating Big Data applications from traditional large database applications, e.g., OLTP, OLAP
- 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
- Developing tools to help instantiate sample Big Data systems
- Fostering collaboration among related benchmarking efforts
- Filling gaps in the Big Data benchmarking state of the art

## Big Data Benchmarking

- Variety of Applications: Clickstream analysis, system monitoring, spatial data analysis, time series analysis, text mining, graph analysis and mining, social network analysis
- Variety of Systems: MapReduce, key-value stores, scientific databases, parallel databases
- The Big Data Vs: Volume, Variety, Velocity, Veracity, Value
- Opportunity for: Micro-benchmarks, multi-user benchmarks, domain-centric benchmarks

## Developments

- BigBench – Big Data analytics benchmark
  - First version is published, work under way for an industry standard version
- Deep Analytics Pipeline – generic Big Data processing model
  - Incubation stage, waiting for critical mass
- WBDB Workshops in Canada and India for 2015
- BigData Top100 – Top500 equivalent for Big Data

## Workshops on Big Data Benchmarking

1. [WBDB.us 2012, May 8-9, San Jose, USA](#)  
First workshop. 70 participants, invited talks, lightning talks
2. [WBDB.in 2012, December 17-18, Pune, India](#)  
Second workshop. [clds.ucsd.edu/wbdb2012.in](http://clds.ucsd.edu/wbdb2012.in)
3. [WBDB.cn 2013, July 17-18, Xi'an, China](#)  
Third workshop. [clds.ucsd.edu/wbdb2013.cn](http://clds.ucsd.edu/wbdb2013.cn)
4. [WBDB.us 2013, October 9-10, San Jose, USA](#)  
Fourth workshop. [clds.ucsd.edu/wbdb2013.us](http://clds.ucsd.edu/wbdb2013.us)
5. [WBDB.de 2014, August 5-6, Potsdam, Germany](#)  
Fifth workshop. [clds.ucsd.edu/wbdb2014.de](http://clds.ucsd.edu/wbdb2014.de)
6. [WBDB.ca 2015, June 16-17, Toronto, Canada](#)
7. [WBDB.in 2015, September 10-11, Delhi, India](#)

## Weekly Conference Calls

**Internal Meetings:** Biweekly internal conference calls , discussion of SPEC related organization and research. Open to all SPEC Members.

**Open Meetings:** Biweekly public conference calls , a forum to share ideas and developments. Open to anyone interested

**Time:** Wednesdays, 10 am – 11 am PST

Contact [tilmann.rabl@bankmark.de](mailto:tilmann.rabl@bankmark.de) for scheduling your talk

Past presentations at:

[research.spec.org/working-groups/big-data-working-group.html](http://research.spec.org/working-groups/big-data-working-group.html)

### RG Big Data Organization:

Chair: Tilmann Rabl, [tilmann.rabl@bankmark.de](mailto:tilmann.rabl@bankmark.de)

Co-Chair: Chaitan Baru, [baru@sdsc.edu](mailto:baru@sdsc.edu)

Secretary: Meikel Poess, [meikel.poess@oracle.com](mailto:meikel.poess@oracle.com)

### Mailing List:

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

