

THIRD WORKSHOP ON BIG DATA BENCHMARKING

clds.sdsc.edu/wbdb2013.cn

Center for Large-Scale Data Systems (CLDS)
San Diego Supercomputer Center, UC San Diego

CALL FOR PAPERS

The *Third Workshop on Big Data Benchmarking* (3rd WBDB) will be held **July 16-17, 2013 in Xi'an, China**—following upon the first and second workshops held in May 2012 in San Jose, CA and in December 2012 in Pune, India, respectively.

The objective of the WBDB workshops is to make progress towards developing industry standard benchmarks for evaluating hardware and software systems for big data applications.

The *BigData Top100 List*

The *BigData Top100 List* concept emerged from discussions at previous WBDB workshops and related meetings. The list would rank big data systems according to their performance on selected big data analytics workloads, enabling comparisons among different big data solutions.

A successful benchmark would be *simple* to implement and execute; *cost effective*, so that the benefits of executing the benchmark justify its expense; *timely*, with benchmark versions keeping pace with rapid changes in the marketplace; and *verifiable* so that results of the benchmark can be validated via independent means.

Workshop Themes

The 3rd WBDB will emphasize two benchmark proposals that are currently being considered: one based on a *Deep Analytics Pipeline* for event processing¹ and a second based on extending the TPC-DS benchmark with semistructured and unstructured data and new queries targeted at those data, called *BigBench*². The priority is to address the following issues in the context of these benchmark proposals:

- **Data generation:** Models and procedures for generating synthetic data with requisite properties.
- **Workload:** Representative big data business problems and corresponding specific implementations for each step and/or query in the workload.
- **Benchmark execution:** Rules and regulations for running the benchmark; data scale factors; benchmark versioning; benchmark metrics.
- **Metrics for efficiency:** Measuring the efficiency of the solution, e.g. based on costs of acquisition, ownership, energy and/or other factors.

Papers on early implementations of the *Deep Analytics Pipeline* or *BigBench*, or describing lessons learned in benchmarking big data applications are solicited. Discussions of enhancements to these benchmarks are also encouraged, for example, including more data genres (e.g. graphs) in the workload; considering a range of machine learning and other algorithms, etc. Papers proposing other benchmarking alternatives will also be considered.

Workshop Format

- The workshop will include invited talks, regular presentations, “lightning” talks, and discussion sessions.
- A registration fee of **US\$150** will be charged to cover workshop expenses.
- Attendance will be capped to ensure effective discussions and active participation.
- In selecting papers, preference will be given to papers that directly address the themes of the workshop and to diversity of representation across organizations and institutions.
- Extended versions of selected papers will be published in the Springer Verlag *Lecture Notes in Computer Science* series.
- Workshop sponsors will be provided a speaking slot. **Sponsorship opportunities are available. Please contact Chaitan Baru, baru@sdsc.edu if you are interested.**

Paper Submission / Important Dates

Papers should be formatted using the **Springer LNCS Proceedings** format.

Submission instructions will be provided at the workshop website: clds.sdsc.edu/wbdb2013.cn

Important Dates:

May 1, 2013: Submission of short version of paper (**4-8 pages in length**). All papers must be submitted using the CMT system.

May 15, 2013:

Notification to authors about paper acceptance with designation of “full presentation” or “lightning talk”.

July 16-17, 2013:

WBDB2013.cn meeting, Xi'an, China.

August 16, 2013:

Submission of full papers (**16 pages**) for post workshop proceedings.

Organizing Committee:

Chaitan Baru – CLDS, UCSD

Milind Bhandarkar – Greenplum

Eyal Gutkind – Mellanox

Jian Li – IBM

Tong Liu – Mellanox (Local arrangements)

Raghunath Nambiar – Cisco

Meikel Poess – Oracle

Tilmann Rabl – MSR, U Toronto

Xiaohui Zhou – Xi'an University (Local host)

¹ Presentation on Deep Analytics Pipeline see: <http://cc.readytalk.com/play?id=1hws7t>. Enter your name.

² BigBench: Towards an Industry Standard Benchmark for Big Data Analytics, to appear in SIGMOD Conference 2013