Introducing the Big Data Benchmarking Community

**Objective**
- Laying grounds for the development of industry standards for measuring the effectiveness of hardware and software technologies dealing with Big Data
- Exploiting synergies between benchmarking efforts
- Offering a forum for presenting and debating platforms, workloads, data sets and metrics relevant to Big Data

**clds.sdsc.edu/bdbc**
- Actian
- AMD
- Argonne National Labs
- BMMSof t
- Brocade
- CA Labs
- Cisco
- Cloudera
- CMU
- Convey Computer
- CW/IPermonet
- DataStax
- Dell
- EPFL
- Facebook
- GaTech
- Google
- Greenplum
- Hortonworks
- Hewlett-Packard
- IBM
- Indiana University
- InfoSizing
- Intel
- Johns Hopkins U.
- LinkedIn
- MapR/Mahout
- Mellanox
- Microsoft
- NetApp
- Netflix
- NIST
- NSF
- OpenSFS
- Oracle
- Oregon State U.
- Paypal
- PNNL
- Red Hat
- San Diego Supercomputer Center
- SAS
- Seagate
- Shell
- ...  
- SLAC
- SNIA
- Teradata
- Twitter
- Univ. of Minnesota
- UC Berkeley
- UC Irvine
- UC San Diego
- Univ. of Passau
- Univ of Toronto
- Univ. of Washington
- VMware
- WhamCloud
- Yahoo!

**Participants**
- BDBC Steering Committee:
  - Chaitan Baru, CLDS, UCSD, baru@sdsc.edu
  - Milind Bhandarkar, Greenplum, EMC, milind.bhandarkar@emc.com
  - Raghunath Nambiar, Cisco Systems Inc., rnambiar@cisco.com
  - Meikel Poess, Oracle Corporation, meikel.poess@oracle.com
  - Tilmann Rabl, MSRG, University of Toronto, tilmann.rabl@utoronto.ca

---

**Big Data Benchmarking**

- Various Applications:
  - Monitoring, spatial data, time series, text, graphs, ...
- Various Systems:
  - MapReduce, key-value stores, scientific databases, parallel databases, ...
- 3-4 Vs:
  - 1) Volume, 2) Variety, 3) Velocity, and 4) Veracity
- Space for
  - Rich Big Data micro benchmarks; multi-user, domain-centric, self-managing, challenging benchmarks

**Recent Developments**

- TPC-DS+: Extending TPC-DS with semi-structured and unstructured data and NoSQL queries
- Workshops on Architecture and Management of Big Data Systems
- Prototype Benchmark: Analogy to the TOP500 project

**Workshop on Big Data Benchmarking**

1. **WBDB.us 2012**
   - First workshop, May 8-9 in San Jose
   - 70 participants, invited talks, lightning talks, discussion sessions
2. **TPCTC 2012**
   - Invited session - WBDB 2012 Report, August 27, Istanbul
3. **WBDB.in 2012**
   - Second workshop, December 17-18, Pune, India
   - CFP clds.ucsd.edu/wbdb2012.in
   - Paper submission date: **Oct 1, 2012**
4. **WBDB.cn 2013**
   - Third workshop, Late June, Xi’an, China
   - CFP clds.ucsd.edu/wbdb2012.cn

---

**Community Calls**

Biweekly conference calls. Open to anyone interested.
Contact tilmann.rabl@utoronto.ca for scheduling your talk.
Information online at clds.ucsd.edu/bdbc/community

**Contact Information**

Call in number: +1-866-740-1260
Access code: 5345802
Regular date: Thursday, 2 pm – 3 pm PST
For upcoming presentations go to:
clds.ucsd.edu/bdbc/community

---

**BDCC Steering Committee:**
Chaitan Baru, CLDS, UCSD, baru@sdsc.edu
Milind Bhandarkar, Greenplum, EMC, milind.bhandarkar@emc.com
Raghunath Nambiar, Cisco Systems Inc., rnambiar@cisco.com
Meikel Poess, Oracle Corporation, meikel.poess@oracle.com
Tilmann Rabl, MSRG, University of Toronto, tilmann.rabl@utoronto.ca