

Introducing the Big Data Benchmarking Community

Objective

clds.sdsc.edu/bdbc

Participants

- 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

- | | | | |
|-------------------------|--------------------|----------------------------------|-----------------------|
| ○ Actian | ○ Facebook | ○ NetApp | ○ SLAC |
| ○ AMD | ○ GaTech | ○ Netflix | ○ SNIA |
| ○ Argonne National Labs | ○ Google | ○ NIST | ○ Teradata |
| ○ BMMsoft | ○ Greenplum | ○ NSF | ○ Twitter |
| ○ Brocade | ○ Hortonworks | ○ OpenSFS | ○ Univ. of Minnesota |
| ○ CA Labs | ○ Hewlett-Packard | ○ Oracle | ○ UC Berkeley |
| ○ Cisco | ○ IBM | ○ Ohio State U. | ○ UC Irvine |
| ○ Cloudera | ○ IndianaU/HTRF | ○ Paypal | ○ UC San Diego |
| ○ CMU | ○ InfoSizing | ○ PNNL | ○ Univ of Passau |
| ○ Convey Computer | ○ Intel | ○ Red Hat | ○ Univ of Toronto |
| ○ CWI/Monet | ○ Johns Hopkins U. | ○ San Diego Supercomputer Center | ○ Univ. of Washington |
| ○ DataStax | ○ LinkedIn | ○ SAS | ○ VMware |
| ○ Dell | ○ MapR/Mahout | ○ Seagate | ○ WhamCloud |
| ○ EPFL | ○ Mellanox | ○ Shell | ○ Yahoo ! |
| | ○ Microsoft | | ○ ... |

Big Data Benchmarking

Recent Developments

- 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

- 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

Community Calls

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

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

BDDB 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
Tilman Rabl, MSRSG, University of Toronto, tilmann.rabl@utoronto.ca

