

February 28, 2013 | By Jan Zverina

SDSC Coordinates Effort to Establish the BigData Top100 List

Benchmarking Initiative Outlined in Inaugural Issue of 'Big Data Journal'



The San Diego Supercomputer Center (SDSC) at the University of California, San Diego, today announced plans for a community-based effort to create the BigData Top100 List, the first global ranking of its kind for systems designed for big data applications.

The BigData Top100 List will rank systems according to their performance on an application-level workload specification, while also reporting on system efficiencies in terms of price/performance. As an application-level benchmark, the list will complement other rankings of high-performance computing (HPC) systems, such as the [Top500](#) and [Graph500](#).

Preliminary information about the new list is at www.bigdatatop100.org, which also includes information for those interested in joining this consortium and supporting its development.

The National Science Foundation (NSF) has defined [big_data](#) as “large, diverse, complex, longitudinal, and/or distributed data sets generated from instruments, sensors, Internet transactions, email, video, click streams, and/or all other digital sources available today and in the future.”

“The explosion in data and the value of repurposing and exploiting data assets have created what we now call the ‘big data’ phenomenon,” said Chaitan Baru, SDSC Distinguished Scientist and director of the Center for Large-scale Data Systems research (CLDS), an industry-supported center of excellence within SDSC focusing on both the technical and management aspects of big data and other data-enabled applications now becoming pervasive among academia, government, and industry.

Baru also was recently appointed SDSC's Associate Director, Data Initiatives. The position reflects the Center's focus on addressing both the management and technical aspects of big data and other data-enabled applications now becoming pervasive among academia, government, and industry.

The BigData Top100 List initiative was announced at the O'Reilly Strata Conference in Santa Clara, California this week in a joint presentation by Baru and Milind Bhandarkar, Chief Scientist, Greenplum, a division of EMC and an industry sponsor of the CLDS. Baru and Bhandarkar are members of the initial BigData Top100 List steering group, which also includes Dhruva Borthakur (Facebook), Eyal Gutkind (Mellanox), Jian Li (IBM), Raghunath Nambiar (Cisco), Ken Osterberg (Seagate), Scott Pearson (Brocade), Meikel Poess (Oracle), Tilmann Rabl (University of Toronto), Richard Treadway (NetApp), and Jerry Zhao (Google).

A comprehensive article outlining the introduction of the new list and describing the benchmarking initiative was published in the March 2013 inaugural issue of the quarterly journal Big Data.

"The creation of a new journal focused solely on big data underscores the importance of this trend across all applications domains, in science as well as business," said Bhandarkar. "The benchmarking effort is a pioneering activity in big data, and creating such a benchmark is a vital step toward fostering competition and innovation in the field."

In addition, an online competition for refining the benchmark dataset and benchmark workload will be announced shortly on kaggle.com, a leading platform for predictive modeling competitions.

"We were excited when SDSC approached us with the idea of benchmarking data systems," said Will Cukierski, who co-leads development of public competitions for Kaggle.

"Competitions reward objective merit, and merit is what ought to matter as this industry matures."

"Big data is now part of every sector and function of the global economy, and the tremendous growth in data has created the need for benchmarks to quantify system performance and price/performance on big data tasks and applications," added Baru. "The existence of such benchmarks enables healthy competition among technology and solution providers, resulting eventually in product improvements and evolution of new technologies."

MEDIA CONTACT

Warren R. Froelich, 858-822-3622, froelich@sdsc.edu

For Comment: Chaitan Baru, (+1 858 534-5035) or baru@sdsc.edu Raghunath Nambiar (+1 408 527 3052) or rnambiar@cisco.com Meikel Poess, (+1 650 633 8012) or meikel.poess@oracle.com Tilmann Rabl, tilmann.rabl@utoronto.ca Milind Bhandarkar, milind.bhandarkar@emc.com

UC San Diego's Studio Ten 300 offers radio and television connections for media interviews with our faculty, which can be coordinated via studio@ucsd.edu. To connect with a UC San Diego faculty expert on relevant issues and trending news stories, visit <https://ucsdnews.ucsd.edu/media-resources/faculty-experts>.