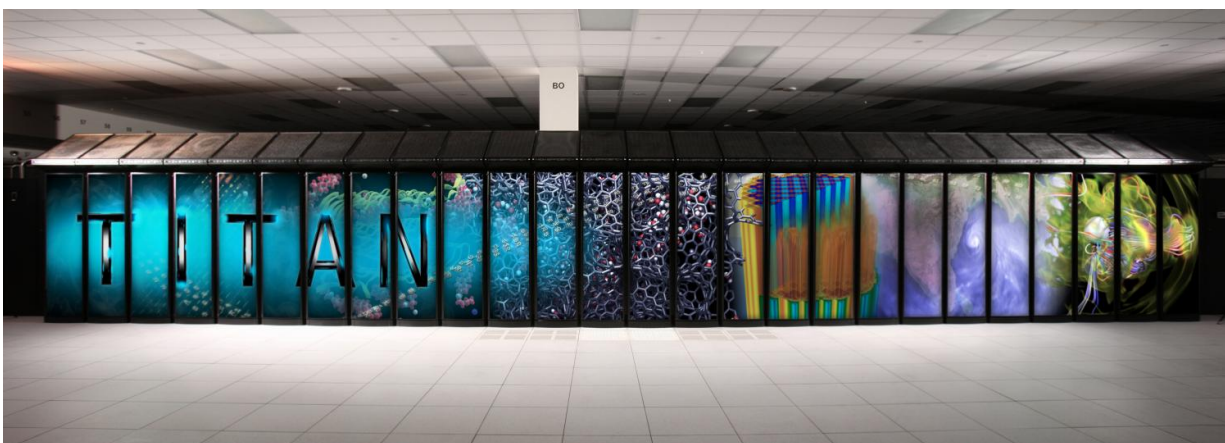# Big PanDA tackles big data for physics and other future extreme scale scientific applications

August 16 2016



A workload management system developed by a team including physicists from Brookhaven National Laboratory taps into unused processing time on the Titan supercomputer at the Oak Ridge Leadership Computing Facility to tackle complex physics problems. New funding will help the group extend this approach, giving scientists in other data-intensive fields access to valuable supercomputing resources. Credit: Oak Ridge Leadership Computing Facility

A billion times per second, particles zooming through the Large Hadron Collider (LHC) at CERN, the European Organization for Nuclear Research, smash into one another at nearly the speed of light, emitting subatomic debris that could help unravel the secrets of the universe.

Collecting the data from those collisions and making it accessible to more than 6000 scientists in 45 countries, each potentially wanting to slice and analyze it in their own unique ways, is a monumental challenge that pushes the limits of the Worldwide LHC Computing Grid (WLCG), the current infrastructure for handling the LHC's computing needs. With the move to higher collision energies at the LHC, the demand just keeps growing.

To help meet this unprecedented demand and supplement the WLCG, a group of scientists working at U.S. Department of Energy (DOE) national laboratories and collaborating universities has developed a way to fit some of the LHC simulations that demand high computing power into untapped pockets of available computing time on one of the nation's most powerful supercomputers-similar to the way tiny pebbles can fill the empty spaces between larger rocks in a jar. The group-from DOE's Brookhaven National Laboratory, Oak Ridge National Laboratory (ORNL), University of Texas at Arlington, Rutgers University, and University of Tennessee, Knoxville-just received $2.1 million in funding for 2016-2017 from DOE's Advanced Scientific Computing Research (ASCR) program to enhance this "workload management system," known as Big PanDA, so it can help handle the LHC data demands and be used as a general workload management service at DOE's Oak Ridge Leadership Computing Facility (OLCF, https://www.olcf.ornl.gov/), a DOE Office of Science User Facility at ORNL.

"The implementation of these ideas in an operational-scale demonstration project at OLCF could potentially increase the use of available resources at this Leadership Computing Facility by five to ten percent," said Brookhaven physicist Alexei Klimentov, a leader on the project. "Mobilizing these previously unusable supercomputing capabilities, valued at millions of dollars per year, could quickly and effectively enable cutting-edge science in many data-intensive fields."

Proof-of-concept tests using the Titan supercomputer at Oak Ridge National Laboratory have been highly successful. This Leadership Computing Facility typically handles large jobs that are fit together to maximize its use. But even when fully subscribed, some 10 percent of Titan's computing capacity might be sitting idle-too small to take on another substantial "leadership class" job, but just right for handling smaller chunks of number crunching. The Big PanDA (for Production and Distributed Analysis) system takes advantage of these unused pockets by breaking up complex data analysis jobs and simulations for the LHC's ATLAS and ALICE experiments and "feeding" them into the "spaces" between the leadership computing jobs. When enough capacity is available to run a new big job, the smaller chunks get kicked out and reinserted to fill in any remaining idle time.
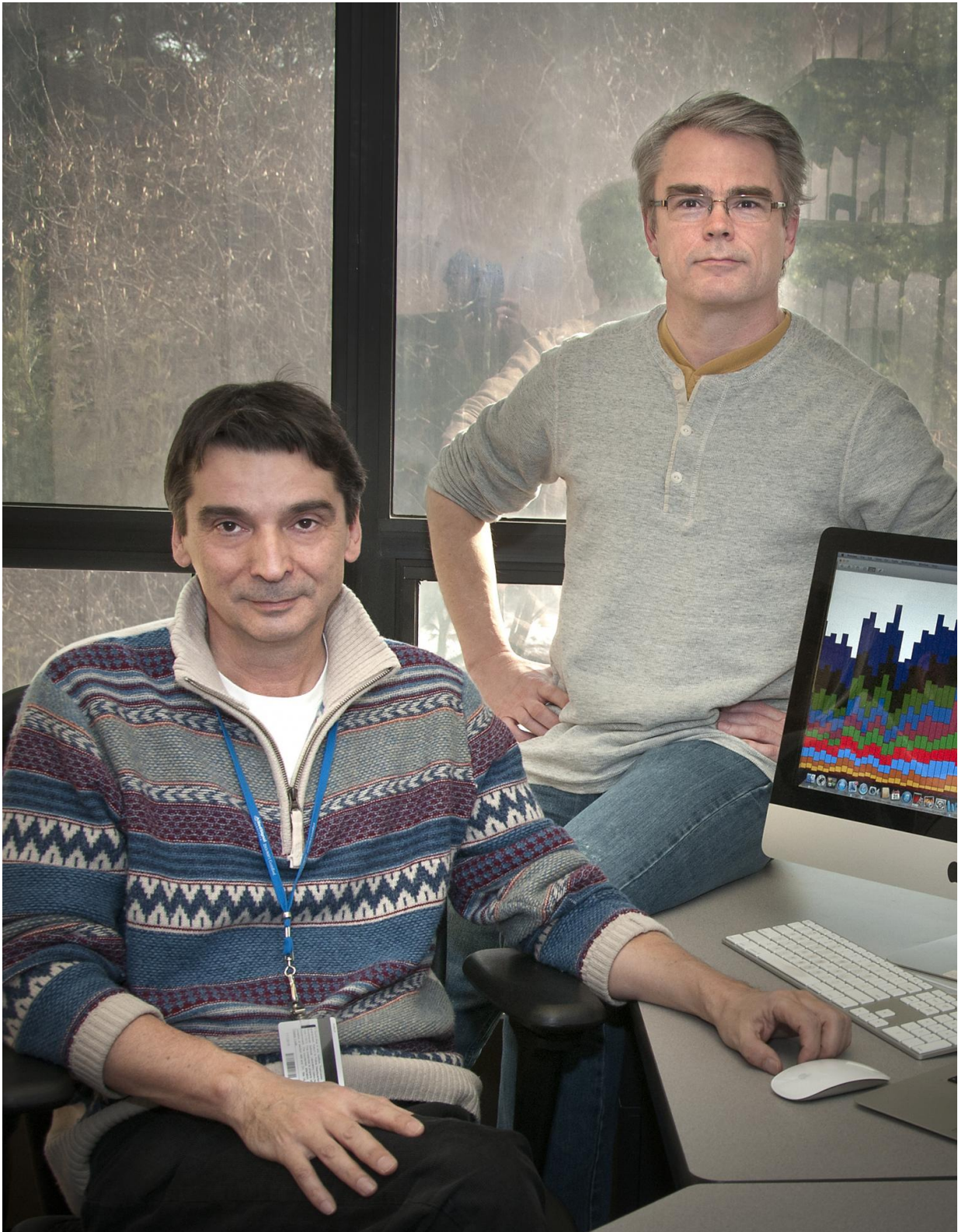
Brookhaven physicists Alexei Klimentov and Torre Wenaus helped design the

Big PanDA workload management system. Credit: Brookhaven National Laboratory

"Our team has managed to access opportunistic cycles available on Titan with no measurable negative effect on the supercomputer's ability to handle its usual workload," Klimentov said. He and his collaborators estimate that up to 30 million core hours or more per month may be harvested using the Big PanDA approach. From January through July of 2016, ATLAS detector simulation jobs ran for 32.7 million core hours on Titan, using only opportunistic, backfill resources. The results of the supercomputing calculations are shipped to and stored at the [RHIC & ATLAS Computing Facility](link), a Tier 1 center for the WLCG located at Brookhaven Lab, so they can be made available to ATLAS researchers across the U.S. and around the globe.

The goal now is to translate the success of the Big PanDA project into operational advances that will enhance how the OLCF handles all of its data-intensive computing jobs. This approach will provide an important model for future exascale computing, increasing the coherence between the technology base used for high-performance, scalable modeling and simulation and that used for data-analytic computing.

"This is a novel and unique approach to workload management that could run on all current and future leadership computing facilities," Klimentov said.

Specifically, the new funding will help the team develop a production scale operational demonstration of the PanDA workflow within the OLCF computational and data resources; integrate OLCF and other leadership facilities with the Grid and Clouds; and help high-energy and nuclear physicists at ATLAS and ALICE-experiments that expect to

collect 10 to 100 times more data during the next 3 to 5 years-achieve scientific breakthroughs at times of peak LHC demand.

As a unifying workload management system, Big PanDA will also help integrate Grid, leadership-class supercomputers, and Cloud computing into a heterogeneous computing architecture accessible to scientists all over the world as a step toward a global cyberinfrastructure.

"The integration of heterogeneous computing centers into a single federated distributed cyberinfrastructure will allow more efficient utilization of computing and disk resources for a wide range of scientific applications," said Klimentov, noting how the idea mirrors Aristotle's assertion that "the whole is greater than the sum of its parts."

Provided by Brookhaven National Laboratory

Citation: Big PanDA tackles big data for physics and other future extreme scale scientific applications (2016, August 16) retrieved 20 April 2024 from https://phys.org/news/2016-08-big-panda-tackles-physics-future.html