

Data-intensive science executed within leadership-scale computing facilities.

Jack C. Wells
Director of Science
National Center for Computational Sciences
Oak Ridge National Laboratory

High-performance computing centers, like the Oak Ridge Leadership Computing Facility (OLCF) have realized growth in data-intensive research projects. The goals of experimental and observational data-intensive (EOD) science, like the ATLAS and ALICE experiments at the Large-Hadron Collider (LHC), are joining the goals of simulation studies in their requirements for access to computing at the largest scales. The BigPanDA project, along with other data projects, has served as a driver for innovation at OLCF. These innovations have included the opportunistic backfill in Titan's scheduled compute nodes with the large, malleable workload available from distributed-computing projects; as sand may fill the gaps between rocks packed into a jar. The OLCF has also deployed multiple container strategies to automate deploying containers as a framework for providing user-required services and applications, and HPC container runtimes focused on use within in a batch submission system. Moreover, Titan's GPU-accelerated architecture has attracted a surge in machine-learning workloads. With the advent of the Summit supercomputer in 2018 with over 27,000 machine-learning-optimized GPUs, high-bandwidth data movement, and large node-local memory, the volume of data analysis and machine-learning workloads is expected to grow significantly into the future. Summit will deliver more than five to ten times the computational performance compared with Titan. Upon completion, Summit will allow researchers in all fields of science unprecedented access to solving some of the world's most pressing challenges.