

What is a Core Hour on Titan?



The OLCF is home to Titan, the world's most powerful supercomputer for open science with a theoretical peak performance exceeding 20 petaflops (quadrillion calculations per second). That kind of computational capability—almost unimaginable—is on par with each of the world's 7 billion people being able to carry out 3 million calculations per second. Image courtesy Oak Ridge National Laboratory

For 2013 the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program awarded 1.84 billion core hours on Titan, a hybrid-architecture high-performance computing (HPC) system that sports central processing units (CPUs) and graphics processing units (GPUs), which are radically different, in the same machine. So what exactly does it mean to allocate a core hour on Titan?

The Oak Ridge Leadership Computing Facility (OLCF) operates Titan based on nodes, or functional units in a message-passing network that can be assigned to work together on a specific task, such as calculating how the Earth's climate might change over a decade. Each node on Titan runs a single copy of the Cray Linux operating system. Likewise, a node's cores can work together to run a simulation. Because communication is faster within the components of a single node than between nodes, data is structured so calculations

stay within nodes as much as possible. For individual projects running on Titan, different nodes can be assigned to run concurrently.

Whereas nodes are the functional units of supercomputers, cores are used to explain how users get "charged" for time on Titan. For each hour a node runs calculations on Titan, the OLCF assesses users an accounting charge of 30 core hours. The number 30 is the sum of the 16 x86 CPU cores on each node's CPU microprocessor and the 14 streaming multiprocessors on each node's GPU processor. The Department of Energy's Advanced Scientific Computing Research (ASCR) program has for several years used a policy that describes allocations in terms of core hours. The number of cores has increased with ensuing generations of leadership-class HPC systems. Allocations for new and returning projects in 2013 will be made in terms of 30 "Titan core hours," and the OLCF will manage accounting and reporting to ASCR in these same units. Users are encouraged to take the greatest possible advantage of the GPUs because the accounting method charges 30 core hours for each node hour consumed regardless of how many CPU cores or GPU streaming multiprocessors are used.

A Cray XK7 system, Titan contains 18,688 nodes, each built from a 16-core AMD Opteron 6274 (traditional x86) processor and an NVIDIA Tesla K20X GPU accelerator. Titan has 299,008 CPU cores (18,688 nodes × 16 cores per CPU = 299,008) and 261,632 GPU cores (1 GPU per node × 18,688 nodes × the assigned value of 14 to represent the streaming multiprocessors of the GPU = 261,632).

Since its first allocations in 2004, INCITE has awarded time on HPC systems with diverse architectures, including the IBM machines Mira and Intrepid at the Argonne Leadership Computing Facility (ALCF) and the Cray Titan at the OLCF. The architectures have evolved through the years. For example, on Titan's predecessor, Jaguar, x86 CPU processors were upgraded in stages from single-core processors to dual-core, quad-core, hex-core, and eventually 16-core ones. Then GPUs were added. Regardless of evolving architectures, however, the ALCF and OLCF, which jointly manage the INCITE program, make awards through one process, so a system had to be devised to allocate time on resources with architectures that are not equivalent. Managers chose to allocate core hours.

The birth of hybrid systems

Before 2004 the price of a computer chip was dictated by its clock speed. HPC centers bought chips of 200, 400, and then 800 megahertz and later 1 or 2 gigahertz. But the heat generated in a chip increases exponentially with speed, so as clock speeds got faster, chips started getting too hot. The first wave in

the evolution of cores was their multiplication on chips when manufacturers concluded that a chip's clock speed could not keep increasing.

Because chips couldn't go faster, manufacturers instead made cores smaller, thanks to fabrication processes that kept shrinking transistors, and placed more cores on a processor. Instead of one processor on one chip, they put two on a chip and called it a dual-core chip. From there the number of cores per chip increased to 4 then 6 and now 16. That's how the OLCF scaled up Jaguar—first growing the number of cabinets to 200 and then increasing the number of cores per processor.

Titan is the first major step in a new direction as computational scientists seek further increased parallelism. Scaling Titan using the traditional paradigm of increasing numbers of cores on a processor would have led to 32 cores per node, as in the Blue Waters machine. That approach might at best have doubled performance from 2 to 4 petaflops, leaving it far short of Titan's 27 peak petaflops, which was enabled by bringing in the GPU.—by Dawn Levy

Job resource accounting in a nutshell

(abridged from http://www.olcf.ornl.gov/kb_articles/job-resource-accounting/ by Adam Carlyle):

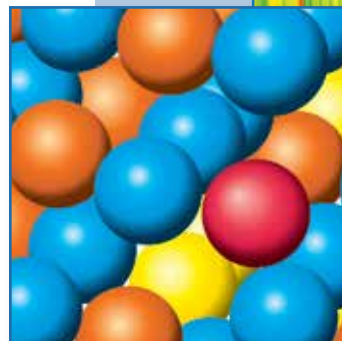
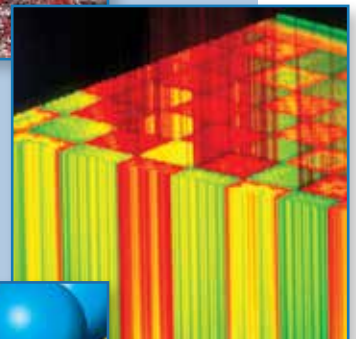
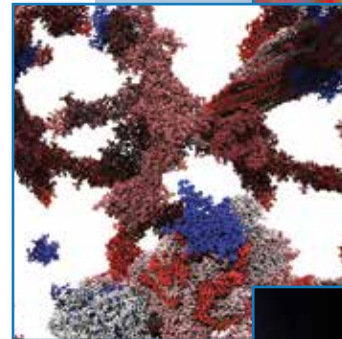
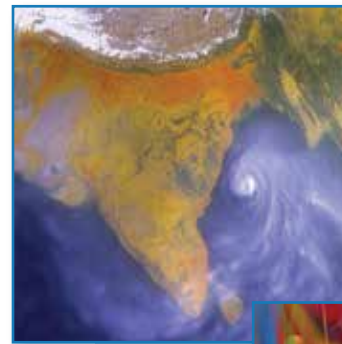
The hybrid nature of Titan's accelerated XK7 nodes mandated a new approach to its node-allocation and job-charge units. For the sake of resource accounting, each Titan XK7 node will be defined as possessing 30 total cores (e.g., 16 CPU cores + 14 GPU streaming multiprocessor equivalents). Jobs consume charge units in "Titan core hours," and each Titan node consumes 30 of such units per hour.

As in years past, jobs on the Titan system will be scheduled in full-node increments; a node's cores cannot be allocated to multiple jobs. Because the OLCF charges based on what a job makes unavailable to other users, a job is charged for an entire node even if it uses only one core on a node. To simplify the process, users are required to request an entire node through the Portable Batch System [OLCF's charging mechanism].

Notably, codes that do not take advantage of GPUs will have only 16 CPU cores available per node; however, allocation requests—and units charged—will be based on 30 cores per node. Whole nodes must be requested at the time of job submission, and associated allocations are reduced by 30 core hours per node, regardless of actual CPU or GPU core utilization.

Core Hours Calculator Example

4 Nodes × 30 × 1 Hours = 120 Core Hours



For more information, contact:

Oak Ridge Leadership Computing Facility
Oak Ridge National Laboratory
Phone: 865-241-7202
Fax: 865-241-2850
E-mail: help@olcf.ornl.gov
www.olcf.ornl.gov